Evidence synthesis methods to inform decisions about complex behaviour change interventions in public health: brief alcohol interventions as a case study

Fiona Ruth Beyer

Thesis submitted for the degree of Doctor of Philosophy

Population Health Sciences Institute, Newcastle University

November 2022

Abstract

Background

Decisions about health interventions should be based on best available evidence. Evidence synthesis comprises a suite of methods for robustly collating and integrating existing research to inform these decisions. I explored the use and limitations of established evidence synthesis methods for informing policy decisions about the use of complex interventions, using brief alcohol interventions as a case study. Hazardous and harmful alcohol consumption costs the NHS £3.5 billion annually, and brief alcohol interventions aim to reduce it and prevent alcohol-related harm in people who are neither aware of the risks nor seeking treatment. Practitioner delivered interventions involve in-person conversations that provide feedback and options for reducing consumption; digitally delivered interventions have become more common as personal devices have become more sophisticated and convenient.

Methods

I followed established systematic review and pairwise and network meta-analysis methods to estimate the effectiveness of brief alcohol interventions, and used subgroup and meta-regression methods to explore heterogeneity. I critically appraised the methods and limitations of meta-analysis of randomised controlled trials (RCTs) for evaluating such complex interventions, and explored elements of complexity missed by these analyses.

Results

Both practitioner and digitally delivered interventions were effective compared to no or minimal treatment control groups for reducing weekly but not episodic drinking. Practitioner delivered interventions reduced consumption more than digitally delivered interventions for the first six months post-intervention. However, the meta-analyses of RCTs could neither incorporate all relevant data nor fully explore contextual factors for these complex interventions.

Conclusions

Despite a wealth of systematic reviews and RCTs in this field, conventional evidence synthesis methods were unable to fully evaluate these complex interventions. Further RCTs alone are unlikely to enhance this evidence base; rather, observational and real-world data should be utilised to enrich our understanding of how best to use complex interventions.

Acknowledgements

I am grateful for the support and collaboration of many colleagues in the conduct of this work. I would like to heartily thank my supervisors Professor Dawn Craig, Professor Eileen Kaner, and Dr Niina Kolehmainen for all their patient and good-humoured encouragement, support, and teaching over these last years. Professor Amy O'Donnell, Professor Vera Araujo Soares, and Dr Ruth McGovern have provided useful feedback and encouragement as my assessors, and I am grateful for their time and enthusiasm.

I would like to thank my co-authors for being such a great team and my PPI colleagues, in particular Lois Neal and Sandy Harvey, for their interest and willingness to keep asking questions and challenging my assumptions and interpretations. Sandy was tireless to the end and a huge inspiration for much more than this PhD.

I am grateful to Professor Elaine McColl and Ms Janice Fuller for keeping me on the straight and narrow and providing helpful advice throughout.

I would like to acknowledge the NIHR School for Public Health Research who funded part of this work.

Finally and most importantly, I would like to express wholehearted and sincere gratitude to my husband Trevelyan and daughters Erin and Isla for keeping the faith, keeping me going, and keeping me grounded in the most important things in life.

Contents

List of figures

List of tables

List of abbreviations

AUDIT	Alcohol Use Disorders Identification Test
BCT	Behaviour Change Technique
CINeMA	Confidence In Network Meta-Analysis
CCA	Cost Consequence Analysis
CI	Confidence Interval
CUA	Cost Utility Analysis
DALY	Disability Adjusted Life Year
FDA	Food and Drug Administration
FRAMES	Feedback, Responsibility, Advice, Menu, Empathy, Self-Efficacy
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
JBI	Joanna Briggs Institute
MI	Motivational Interviewing
MRC	Medical Research Council
NCD	Non-communicable disease
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network Meta-Analysis
ORBITAL	Outcome Reporting in Brief Intervention Trials: ALcohol
PICO(S/T)	Population, Intervention, Comparator, Outcome (Study design, Time)
PRESS	Peer Review of Electronic Search Strategies
PRISMA	Preferred Reporting Items for Meta-Analysis

PRISMA-P	Preferred Reporting Items for Meta-Analysis – Protocol extension
PRISMA-S	Preferred Reporting Items for Meta-Analysis – Searching extension
PPI	Patient and Public Involvement
RCT	Randomised Controlled Trial
RoB	Risk of Bias
SBI	Screening and Brief Intervention
SD	Standard deviation
SMS	Short Message Service (text message)
SPICE	Setting, Population, Intervention, Comparator, Evaluation
SWiM	Synthesis Without Meta-analysis
WHO	World Health Organization

Introduction

This programme of work explores the use and limitations of established evidence synthesis methods for informing decisions about the effectiveness of interventions, using brief alcohol interventions as a case study of a complex behaviour change intervention. Given that 82% of people aged 16 and over in the UK report themselves as 'current drinkers', and at least a fifth of these drink at hazardous levels that could risk their health, alcohol interventions could potentially have a large impact on public health. Brief alcohol interventions were first introduced some three decades ago and have been researched ever since, so they provide a rich field for evidence synthesis research.

This doctoral statement charts my journey through the process of evaluating the effectiveness of brief alcohol interventions and the difficulties I encountered. In the following pages I present a critique of both my own papers in their use of systematic review and meta-analysis methods, and the method in general in the context of complex behaviour change interventions. Although brief interventions are used around the world and the evidence base is international, I focus here on the UK context for the purposes of considering the use of evidence synthesis results for policy making.

Outline of this programme of work

Table 1 presents my core portfolio of papers and contributions to each. Part I of this thesis describes the evaluation of the effectiveness of brief alcohol interventions using conventional meta-analysis of randomised controlled trials (RCTs). Following a background chapter and a methods chapter that describes evidence synthesis methodology in detail, in chapter 3 I critique the methods and assumptions of two Cochrane reviews that evaluated the effectiveness of practitioner and digitally delivered interventions respectively (Kaner *et al.*, 2017; Kaner *et al.*, 2018), and specific methodological issues arising for these complex interventions. Due to the lack of available data, it was impossible to determine whether practitioner or digitally delivered interventions were the most effective in reducing heavy drinking. Having published a critical overview of this evidence exploring its complexities (Beyer, Lynch and Kaner, 2018), I incorporated the trials from both Cochrane reviews plus more recently published ones into a network meta-analysis (Beyer *et al.*, 2022), which allowed the use of data from trials that did not compare the interventions of interest directly and provided a more precise estimate of comparative

effectiveness. In chapter 3 I continue the critique of methods, assumptions, and issues for network meta-analysis of complex interventions. I also worked with team members to explore the differences in the digital interventions by coding and analysing them according to the Behaviour Change Techniques (BCT) (Garnett *et al.*, 2018b), and to explore the use of theory in the development of digitally delivered interventions (Garnett *et al.*, 2018a).

Although the above work followed best practice for estimating effectiveness of interventions, for these complex behaviour change interventions they were one-dimensional and incomplete. Whereas such an assessment of these interventions as a 'black box' was necessary to answer the overall question of whether they were effective compared to each other or to no treatment, it gave little sense of the impact of contextual factors, such as participant characteristics, co-morbidities, or social or environmental factors, on the effectiveness of the interventions. These analyses did not afford the data or tools to fully understand features of the characteristics, environment, and social links of the participants that may affect the impact of a complex intervention.

Part 2 explores these contextual factors – why they are important to consider in complex interventions and how to make better use of available data to evaluate them. I worked on two systematic reviews of the effectiveness of interventions that specifically targeted such contextual factors – interventions for heavy drinkers with linked unhealthy eating behaviours and co-morbid depression respectively (Scott *et al.*, 2018; O'Donnell *et al.*, 2022). After some background about context in chapter 4, in chapter 5 I discuss and critique the methods used in these two reviews - narrative synthesis and meta-analysis respectively. Complex interventions for a multi-faceted human behaviour like heavy drinking are likely to be affected by so many potential contextual factors that dealing with them intervention by intervention like this is likely to be inefficient – and it is likely that some individuals may be dealing with multiple contextual factors that would be better addressed together. Finally, then, I take a broader look at hazardous alcohol consumption in context and consider how other evidence than RCTs can contribute to good decision making in health care.

Papers	My contributions
1. a. Kaner EFS, **Beyer FR**, Muirhead C, Campbell F, Pienaar E, Bertholet N, et al. Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database of Systematic Reviews*. 2018;2 https://doi.org/10.1002/14651858.CD004148.pub4	Corresponding author Ran search updates Screened results and extracted data Carried out risk of bias assessments Contributed to design of analyses Project management
b. **Beyer FR**, Campbell F, Bertholet N, Daeppen JB, Saunders JB, Pienaar ED, et al. The Cochrane 2018 review on brief interventions in primary care for hazardous and harmful alcohol consumption: a distillation for clinicians and policy makers. *Alcohol and Alcoholism*. 2019;54(4):417-27 https://doi.org/10.1093/alcalc/agz035	Corresponding author Created first draft Incorporated co-authors' contributions and comments
2. Kaner EFS, **Beyer FR**, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. *Cochrane Database of Systematic Reviews*. 2017;9 https://doi.org/10.1002/14651858.CD011479.pub2	Corresponding author Intellectual contribution to protocol Designed and ran searches Screened results and extracted data Carried out risk of bias assessments

Papers	My contributions
	Contributed to design of analyses Project management
3. Garnett CV, Crane D, Brown J, Kaner EFS, **Beyer FR**, Muirhead C, et al. Behavior Change Techniques used in digital behavior change interventions to reduce excessive alcohol consumption: a meta-regression. *Annals of Behavioral Medicine*. 2018;52(6):530-43 https://doi.org/10.1093/abm/kax029	Contribution to design of project Screened results Commented on drafts
4. Garnett C, Crane D, Brown J, Kaner E, **Beyer FR**, Muirhead C, et al. Reported theory use by digital interventions for hazardous and harmful alcohol consumption, and association with effectiveness: meta-regression. *Journal of Medical Internet Research*. 2018;20(2) https://doi.org/10.2196/jmir.8807	Contributed to design of project Screened results Commented on drafts
5. **Beyer FR**, Lynch E, Kaner E. Brief interventions in primary care: an evidence overview of practitioner and digital intervention programmes. *Current Addiction Reports*. 2018;5(2):265-73 https://doi.org/10.1007/s40429-018-0198-7	Corresponding author Created first draft Incorporated co-authors' contributions and comments
6. **Beyer FR**, Kenny RPW, Johnson E, Caldwell DM, Garnett C, Rice S, et al. Practitioner and digitally delivered interventions for reducing hazardous and harmful alcohol consumption in people not seeking alcohol treatment: a systematic review and network meta-analysis. *Addiction*. 2022 https://doi.org/10.1111/add.15999	Corresponding author Designed and led the project Designed and ran searches Screened results and extracted data Carried out risk of bias assessments Carried out analysis

Papers	My contributions
	Created first draft Incorporated co-authors' contributions and comments
7. Scott S, **Beyer FR**, Parkinson K, Muir C, Graye A, Kaner E, et al. Non-pharmacological interventions to reduce unhealthy eating and risky drinking in young adults aged 18–25 years: a systematic review and meta-analysis. *Nutrients*. 2018;10(10):1538 https://doi.org/10.3390/nu10101538	Advised on search design Advised on analysis Contributed to the writing of methods, results and discussion Commented on drafts
8. O'Donnell A, Schmidt CS, **Beyer FR**, Schrietter MS, Anderson P, Jané-Llopis E, et al. Effectiveness of digital interventions for people with comorbid heavy drinking and depression: A systematic review and narrative synthesis. *Journal of Affective Disorders*. 2022;298:10-23 https://doi.org/10.1016/j.jad.2021.11.039	Intellectual contribution to protocol Designed and ran searches Carried out risk of bias assessments Created harvest plots Contributed to the writing of methods, results and discussion Commented on drafts

Table 1: Portfolio of papers and my contributions

Contribution of this programme of work

This programme of work contributes to the fields of both alcohol and evaluation research. The Cochrane reviews were the most comprehensive systematic reviews of alcohol interventions to date (Kaner *et al.*, 2017; Kaner *et al.*, 2018), and contributed evidence to several national guidelines (NICE, 2010; Beeston *et al.*, 2016; Hoffmann and Kiefer, 2020; Kaiser Permanente, 2020). The network meta-analysis was the first published analysis using all the available trial evidence to directly compare practitioner and digitally delivered interventions.

Meta-analysis provides a robust method to combine the results of many RCTs into one weighted summary estimate of intervention effectiveness (Deeks, Higgins and Altman, 2022). However, RCTs provide insufficient evidence on their own to fully evaluate complex interventions, and the richness of other data can enhance the evidence base for decision makers. The field of alcohol interventions has been less likely than other fields to make full use of other evidence than RCTs, such as observation, qualitative or real-world evidence. Rather than omitting data that doesn't fit robust methods, after the initial evaluation we should adapt the methods to make the most use of existing observational and real-world evidence to understand fully how interventions work.

Aims and objectives

The aim of this programme of work was to explore and critically reflect upon the use of evidence synthesis methods for informing decisions about the use of complex behaviour change interventions, using brief alcohol interventions as a case study.

The objectives were as follows:

1. To apply and critically appraise alternative evidence synthesis methods to evaluate the effectiveness of alcohol interventions as an example of a complex behaviour change intervention;
2. To explore methodological issues and reflect upon the usefulness of these methods for decision making for complex behaviour change interventions;
3. To explore contextual factors and further data that can contribute to decision making.

PART I: evaluating effectiveness of complex interventions

Part I of this work presents and critiques my papers evaluating the effectiveness of brief interventions for reducing hazardous and harmful alcohol consumption.

Chapter 1. Background (Part I)

This chapter introduces decision making for health care and the importance of evidence synthesis methods for informing these decisions. It then describes why brief alcohol interventions were chosen as a case study, in relation to hazardous and harmful alcohol consumption as a public health problem, and why they are characterised as complex interventions. Finally, it describes Behaviour Change Techniques as an example of a method of exploring the impact of individual components of these complex interventions.

1.1 Health care policy making and evidence

The National Institute for Health and Care Excellence (NICE) exists to improve outcomes for anyone in the UK using public health and social care services, and its core purpose is 'to improve health and wellbeing by putting science and evidence at the heart of health and care decision making' (NICE, 2021). The four overarching strategic pillars in the latest NICE strategy (2021-6), shown in Figure 1, all articulate the importance of carrying out and communicating high quality research (NICE, 2021).

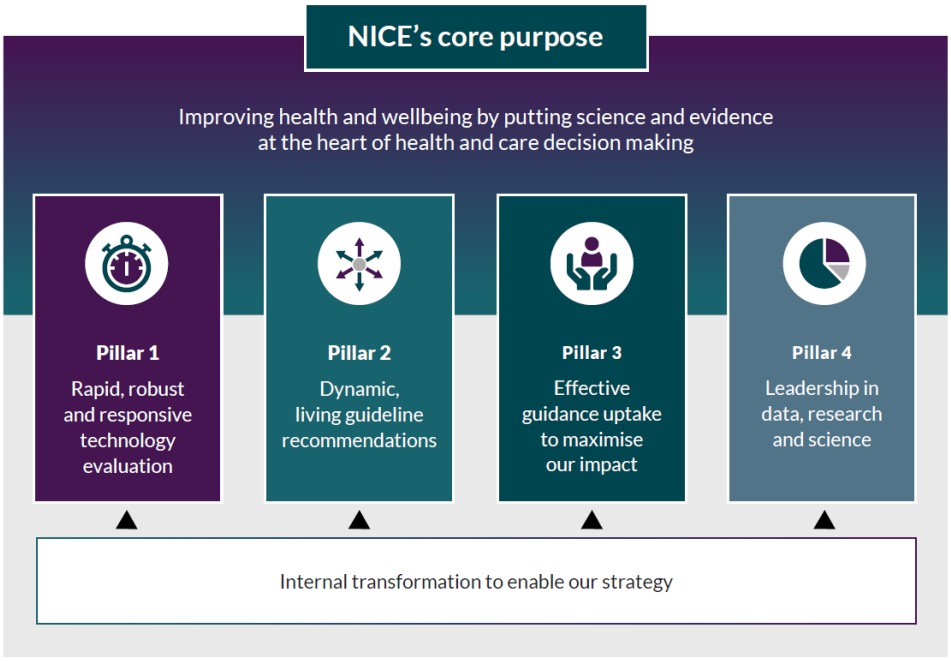


Figure 1: NICE's four strategic pillars demonstrating the focus on high quality research (NICE, 2021)

It follows that decisions about health care and interventions to be funded by NICE should be based on the best available evidence. Indeed, NICE's methods manuals for both health technology evaluation and guideline development emphasise the importance of a robust, systematic, comprehensive and methodologically sound evidence base to inform their decisions and guidance (NICE, 2012; NICE, 2022a). Rigorous and unbiased synthesis of knowledge has been described as by far the most useful academic contribution to policy-making (Whitty, 2015).

The best available evidence to evaluate the effectiveness of interventions designed to improve health is the randomised controlled trial (RCT), where participants are split at random into a group that receives the intervention and a group (or groups) that receives an alternative or no intervention (Meinert, 1986; Piantadosi, 2020). Randomisation (or random allocation) of participants should mean that everything is equal between the groups except for the intervention, so that any difference between the groups at follow-up can be causally attributed to the intervention. The strength of randomisation is that it controls both for factors that investigators and clinicians are likely to be aware of, such as severity of illness, but also 'unknown unknowns'. Successful randomisation removes the chance that a clinician or investigator may consciously or unconsciously introduce systematic differences between the groups by making biased decisions

about allocation – for example, allocating people who are considered more in need of the intervention (Altman, 1991). This would influence the analysis because more severely ill people receiving the intervention could make it look less effective than if the groups were more equal in terms of severity.

The most robust way to evaluate the effectiveness of an intervention is to gather, appraise and synthesise the evidence from RCTs using a systematic review and (ideally) meta-analysis. Systematic review methods provide a clear, transparent and reproducible method of collecting all the relevant evidence and critically appraising it (Centre for Reviews and Dissemination, 2009; Higgins *et al.*, 2022b). Meta-analysis is the most appropriate method for synthesising RCTs and comprises a statistical technique for integrating the results of individual trials to provide a single pooled estimate of effect (Egger, Davey-Smith and Phillips, 1997). I will discuss these methods in more detail in Chapter 2.

There is a tension between academia on the one hand, where researchers are encouraged to articulate well-focused research questions and reduce the amount of heterogeneity and 'messiness' to obtain a precise answer to inform clinical decisions (Sutcliffe *et al.*, 2015), and policy making on the other hand, where questions tend to be broader and factors like public opinion, lobbying by interested stakeholders, and competition for limited budgets may have at least as much influence as evidence from research studies. Unsurprisingly, collaboration and good relationships between policy makers and researchers make it more likely that evidence will be used, as do clearly presented, authoritative, and discernibly relevant research findings (Oliver *et al.*, 2014). When working with policy makers, researchers must resist the temptation to deliver more and more precise answers to less useful questions, and rather focus on the messier questions that are more useful for decision makers but may necessitate a more approximate and caveated answer (Skivington *et al.*, 2021). Brief alcohol interventions provide a good example of this, where trials are still being registered to test particular variations of brief interventions, sometimes for smaller and smaller groups of people, whereas there is already abundant evidence from multiple trials and meta-analyses that overall they reduce heavy drinking (O'Donnell *et al.*, 2014).

1.2 Complex behaviour change interventions

Non-communicable diseases (NCDs) are estimated to cause nearly 90% of deaths in the UK (World Health Organization, 2014). Because behavioural risk factors make the greatest contribution to death and disability from NCDs, behaviour change interventions which encourage individuals and communities to adopt more healthy lifestyles have the potential, when they work well, to have a huge impact on public health. Public Health England (PHE, since re-named Office for Health Improvement and Disparities) cited behavioural science along with digital technologies for interventions as a key tool in its vision (Public Health England, 2019a), so ensuring that behaviour change interventions work as well as possible for the maximum number of people is critical.

Behaviour change interventions are 'complex interventions' according to the definition provided by the Medical Research Council in its guidance for intervention development (Skivington *et al.*, 2021). They contain multiple interacting components which can be tailored at the point of use (examples include advice, a conversation about the triggers for unhealthy behaviour, or setting specific behavioural goals), and multiple behaviours are required by those delivering and receiving the intervention. It is also the case that an identical behaviour change intervention delivered to different people can plausibly have completely different results; for example, between someone who feels ready for and confident about changing their behaviour and someone who perceives many barriers to change in their life circumstances or simply doesn't feel any need to change. Various factors come into play when designing complex behaviour change interventions (Colom *et al.*, 2014). For example, social relationships and norms as well as individual beliefs and values can have a strong influence on behaviour; and change should be considered as a process over time rather than an event that takes place at the point of intervention. This makes it difficult to predict how people will respond, and certain that different people will respond in different ways.

The complexity of behaviour change interventions can make their implementation very complicated – integrating a new intervention into routine care involves its own behaviour change on the part of health professionals. In a primary care setting time is often a barrier (Barry *et al.*, 2004), particularly where an intervention requires extended questioning or explanation; or the health professional may be concerned about damaging their relationship with the patient by addressing the potentially sensitive issue of drinking, as many primary care professionals lack confidence and knowledge (Chan *et al.*, 2021).

1.3 Hazardous alcohol consumption

Drinking alcohol is popular in the UK: 82% of people aged 16 and over in England define themselves as 'current drinkers' (including occasional drinkers) (Office for National Statistics, 2020). Alcohol units vary between countries; in the UK one unit contains eight grams or ten millilitres of alcohol (NHS, 2021). A pint of lower strength beer or a standard (175ml) glass of wine each contain around two UK units.

Most people in the UK drink at low risk levels, defined as less than 14 units per week spread over more than three days, with two or three abstinent days in the week (*UK Chief Medical Officers' low risk drinking guidelines*, 2016). However, 20-33% of UK drinkers drink at a level that risks causing harm to their health, and many don't realise it. Different patterns of drinking can have varying impacts on health. Hazardous drinking is defined as a pattern that increases the risk of physical or psychological harm, whilst harmful consumption is characterised by quantifiable detrimental effects to the drinker and/or those around them (World Health Organization, 2019). For example, heavy episodic or 'binge drinking', defined as more than eight units in a single occasion for men or more than six for women (*The Government's Alcohol Strategy*, 2012), increases the risk of accidents, injuries and liver disease. Drinking every day of the week without the recommended alcohol-free days may seem less risky if the person doesn't seem drunk, but has an adverse impact on many organs in the body and is a risk factor for cancer. Hazardous and harmful levels of drinking do not only affect the drinker: their children and families, colleagues and friends can experience harm, as can strangers encountered whilst drunk (Public Health England, 2019b). These kinds of hazardous and harmful levels of drinking can be addressed with brief alcohol interventions, which aim to prevent the harm and are the case study for my PhD (discussed further in the next section). Brief alcohol interventions are not appropriate for people who are dependent on alcohol, demonstrated by impaired ability to control their use despite obvious negative consequences (World Health Organization, 2022). This group may need pharmacotherapy and/or structured psychosocial intervention (Berglund *et al.*, 2003).

Alcohol related harm is a major public health problem (NICE, 2010). Heavy drinking is a causal factor in more than 60 medical conditions including cancer, high blood pressure, liver disease and depression (Public Health England, 2016). Hazardous and harmful consumption is the biggest risk factor for mortality, morbidity and disability among 15-49 year-olds in the UK, and the fifth biggest risk factor across all ages (Public Health England, 2016). As well as its effect on individuals, heavy

drinking causes or exacerbates social harms to others as well as the drinker, for example with relationship problems and domestic violence and abuse (Boden, Fergusson and Horwood, 2012; Boden, Fergusson and Horwood, 2013), or absence from work and loss of earnings (Schou and Moan, 2016). The NHS incurs around £3.5 billion of costs relating to alcohol each year (Public Health England, 2016).

It follows that reducing the consumption of people who are drinking at hazardous or harmful levels has the potential to improve health outcomes and save a large amount of treatment and rehabilitation cost. At a population level, more alcohol-related problems are attributable to this hazardous and harmful group than the addicted or dependent group because the former is so much larger in number (World Health Organization, 2006). This is an example of the 'alcohol prevention paradox', where prevention of harm in the entire drinking population is more effective than targeting the smaller high-risk population of dependent drinkers (Kreitman, 1986).

An issue of particular public health concern has been dubbed the 'alcohol harm paradox', whereby wealthier households on average exhibit higher levels of risk drinking than low-income ones, but more disadvantaged households are more likely to experience harm or hospital admissions for alcohol-related conditions (Bellis *et al.*, 2016). The reasons for this are not fully understood, but the alcohol harm paradox is yet another indicator of inequity.

1.4 Alcohol interventions

The UK alcohol harm prevention guideline promotes both population and individual level interventions to reduce hazardous drinking and alcohol related harm (NICE, 2010). Recommended population level interventions include setting a minimum price per unit, which has been established and evaluated in Scotland (Beeston *et al.*, 2020; Anderson *et al.*, 2021), reducing availability in terms of number of outlets and their opening hours, and reducing advertising to children. At the individual level, the guideline promotes screening for hazardous or harmful drinking and for those that screen positive, a brief intervention.

Often known as Screening and Brief Interventions (SBI), brief alcohol interventions aim to identify those who exhibit but are not necessarily aware of hazardous alcohol consumption and help them to recognise and reduce their drinking to prevent alcohol-related harm. Consequently, optimising brief alcohol interventions to reduce hazardous or harmful consumption at the earliest opportunity in the maximum number of people has the potential to prevent much of this harm

and save a good deal of money. The screening element is important because many people do not appreciate that their drinking is a problem or understand the consequences for their health, and so it needs to take place in people who are not seeking help for their drinking. Brief alcohol interventions have been used in primary care settings for more than 30 years (Skinner and Holt, 1983) and are recommended in NICE guidelines and government policy (NICE, 2010; NICE, 2011). An early model was FRAMES (Miller and Sanchez, 1993), an in-person conversation which involves giving **F**eedback on the person's intake, impressing the **R**esponsibility for change onto them, offering **A**dvice, listing a **M**enu of options, having an **E**mpathic approach, and building **S**elf-efficacy. Many brief interventions have since incorporated elements of Motivational Interviewing, in which the health professional does not challenge any resistance to change from the drinker but rather 'rolls with' it, instead using the conversation to 'develop discrepancy' between the drinker's goals and their ability to achieve them with their current consumption levels (Miller, 1983; Rollnick and Miller, 1995). Motivational interviewing describes the style of the interaction between counsellor and client, whereas brief intervention refers to the content.

Brief interventions cannot operate effectively in isolation (McCambridge, 2021); they need to be one element of a broader policy response to hazardous and harmful consumption (Alcohol Change UK, 2018). Companion policies include reducing the affordability of alcohol, regulating marketing (particularly targeting children), reducing opening hours, and legislating against drink driving (Public Health England, 2016). This acknowledges that hazardous and harmful drinking is not a purely individual decision but that people are influenced in their decisions by their surroundings, and an 'alcogenic environment' containing many alcohol outlets and opportunities for drinking increases the risk of harmful consumption for some people (Huckle *et al.*, 2008; Hill, Foxcroft and Pilling, 2018).

The relatively long history of brief alcohol interventions means that they are the subject of a wealth of published research, which raises the question of why trials are still being registered but also makes them an ideal case study for exploring the use of evidence synthesis methods. A systematic review of reviews identified 24 systematic reviews published between 2002 and 2012 which reported on a total of 56 randomised trials of practitioner delivered brief alcohol interventions in primary care (O'Donnell *et al.*, 2014). Although the evidence is not equally comprehensive across different populations, it was consistently reported that these interventions are effective in reducing alcohol consumption in this setting. An interesting feature of several of

these trials was that all participants tended to decrease their alcohol consumption, regardless of whether they were in the intervention or control group, suggesting that something unintentional was being delivered within the control groups that influenced consumption. This has led to many reports that 'brief interventions are ineffective' because there is no statistically significant difference between the control and intervention arms at follow-up. However, in the context of all participants reducing their consumption this seems overly simplistic. The assumption behind an RCT is of no effect in the control arm because control participants receive nothing that will influence the target behaviour, and that the randomisation process will ensure that any confounders that may influence the outcome are evenly distributed between the groups and cancel each other out. In this context of no change in the control group, no difference between control and intervention can be correctly interpreted as no effect of the intervention. However, when everyone in the trial reduces their consumption and the intervention group reduce it more than the control group, this could suggest not only that the intervention is effective in reducing consumption, but that something else is going on within the trial environment that influences consumption. This led to my hypothesis that interventions that contain alcohol-related information comprise a minimal intervention in themselves, which I tested by splitting the control groups in the network meta-analysis (see Chapter 3 and Figure 3b).

However, there are other potential interpretations of the observation that all trial participants reduce their consumption. Regression to the mean is a statistical phenomenon that occurs where an unusually large or small measurement tends to be followed by a measurement closer to the mean value of the population (Barnett, van der Pols and Dobson, 2004; Heather, 2014). In the case of brief intervention trials where participants have been screened in according to their level of consumption, those people that would have had a low level of consumption are screened out of the trials and so their regression *up* to the mean is missing. A cohort study that measured levels of drinking in a sample of university students at baseline and six months later (with no intervention) noted that although the mean AUDIT score of the whole sample increased by 1 point, that of higher drinking students decreased, and the decrease was larger for those students drinking more heavily at baseline (McCambridge, Kypri and McElduff, 2014). A common screening threshold for brief intervention trials is AUDIT ≥ 8; students matching this threshold showed a mean increase of 0.4, and students drinking more heavily than in most trials (AUDIT ≥ 20) showed a mean decrease of 2, suggesting that regression to the mean is likely to be a feature of brief intervention trials.

A second interpretation is known as 'assessment reactivity' – the idea that the act of focusing on a person's drinking to complete a baseline assessment itself prompts a change in their consumption (McCambridge and Kypri, 2009). One of the trials included in my reviews tested this by including two control arms: one where participants were screened and received a leaflet about the health effects of alcohol and then had no further contact until follow-up, and another where they were screened and received the same leaflet, then completed a baseline assessment exercise, with no further contact until follow-up (Kypri *et al.*, 2007). In this case the only thing that was different between the arms was the assessment exercise in the second group. This trial reported a higher reduction in consumption for some of the alcohol outcomes in the assessment group than the no-assessment group, suggesting that assessment reactivity also plays a part in brief intervention trials.

There is, therefore, some evidence for both regression to the mean and assessment reactivity in alcohol intervention studies. However, there is no reason to expect that these phenomena would influence the intervention or control arms of a trial differently, so their presence challenges only the size not the existence of the intervention effect.

Now that people routinely carry and wear mobile devices, which can be cheaper and more convenient than an appointment with a health professional and deliver interventions in a more targeted way than a public health campaign, there has been an explosion in the availability of websites and apps aiming to help risky drinkers reduce their alcohol consumption. Digital alcohol interventions may have the potential to overcome long-standing difficulties in implementing practitioner delivered interventions (Keurhorst *et al.*, 2015), and they have been the subject of much recent evaluation focus due to wider and cheaper availability of mobile devices. Nineteen systematic reviews have been published in this area, the most recent of which included 94 computer-delivered interventions (Black, Mullan and Sharpe, 2016; Fowler, Holt and Joshi, 2016; Field *et al.*, 2019). Digital alcohol interventions are accessible via computer or mobile device; they also tend to give feedback on the person's alcohol consumption and provide them with methods by which to reduce their consumption. Their output ranges from a single screen of information which requires a few minutes to view, to multiple sessions involving education and 'homework'. Again, many of the reviews reported a modest but statistically significant reduction in weekly alcohol consumption in people using digital alcohol interventions versus those who didn't. Few

trials have directly compared the effectiveness of practitioner versus digitally delivered interventions, they are small and report mixed results.

1.5 Exploring intervention components: Behaviour Change Techniques

As well as evaluating the effectiveness of brief alcohol interventions overall, my work explores the influence of particular components (or active ingredients) of interventions on their effectiveness. To do this robustly the components need to be categorised using a standardised and consistent method, and individually tested to assess whether each of them is associated with effectiveness. Behaviour Change Techniques (BCTs) provide one standardised method of categorising intervention components (Michie *et al.*, 2015). They were developed using a Delphi process and comprise an 'observable, replicable and irreducible component of an intervention designed to alter or redirect causal processes that regulate behaviour'; that is, a BCT is proposed to be an 'active ingredient' (Michie *et al.*, 2013). The aim of providing a set of standardised and replicable BCTs is to allow for clear instruction and manualisation to aid fidelity of delivery of behaviour change interventions, and to facilitate accurate coding of interventions that are reported in trials to allow effective evaluation. In a group of heterogeneous interventions, BCTs provided a method of articulating the differences between them and assessing whether ingredients or components were more or less associated with the effectiveness of the intervention as a whole.

Chapter 2. Methodology: evidence synthesis

This chapter describes the conventional methods of systematic review and meta-analysis of RCTs that I have used to evaluate effectiveness of alcohol interventions, and how each step contributes to the robustness of the analyses. It goes on to describe other evidence synthesis methods that may usefully utilise other types of study data to improve the evaluation of complex interventions.

2.1 Evidence synthesis

Evidence synthesis is today's version of Isaac Newton's 'standing on the shoulders of giants'. It allows us to make the best use of previous work, build on existing studies and create something that is more than the sum of their parts, so avoiding research waste (Chalmers *et al.*, 2014). Evidence synthesis comprises a suite of tools that use explicit and rigorous methods to bring together information from existing research and articulate the sum of the resulting knowledge to inform decisions and policy (Royal Society and Academy of Medical Sciences, 2018; Evidence Synthesis International, 2021).

Gough describes a spectrum of synthesis methods (Figure 2) that articulate how the purpose of a synthesis dictates the method to be used (Gough, Oliver and Thomas, 2012). At one end of the spectrum is 'aggregative' synthesis. Situated in a positivist framework that claims a single objective reality that can be measured by scientific methods (Allsop, 2013), aggregative synthesis uses statistical methods such as meta-analysis to test hypotheses, seeking to provide a single 'true' numerical estimate about the effectiveness of an intervention. The driver for methodological robustness in this space is to minimise bias in the final estimate as much as possible and make it as close to the purported 'true' value as possible. To achieve this, randomised or non-confounded trial design, comprehensive search methods and clear and transparent pre-specified analyses are all key. For the purposes of evaluating the effectiveness of alcohol interventions the appropriate method is an aggregative synthesis to establish whether they are overall effective or not. Part I of my PhD thesis used these aggregative synthesis methods to explore effectiveness of brief alcohol interventions.

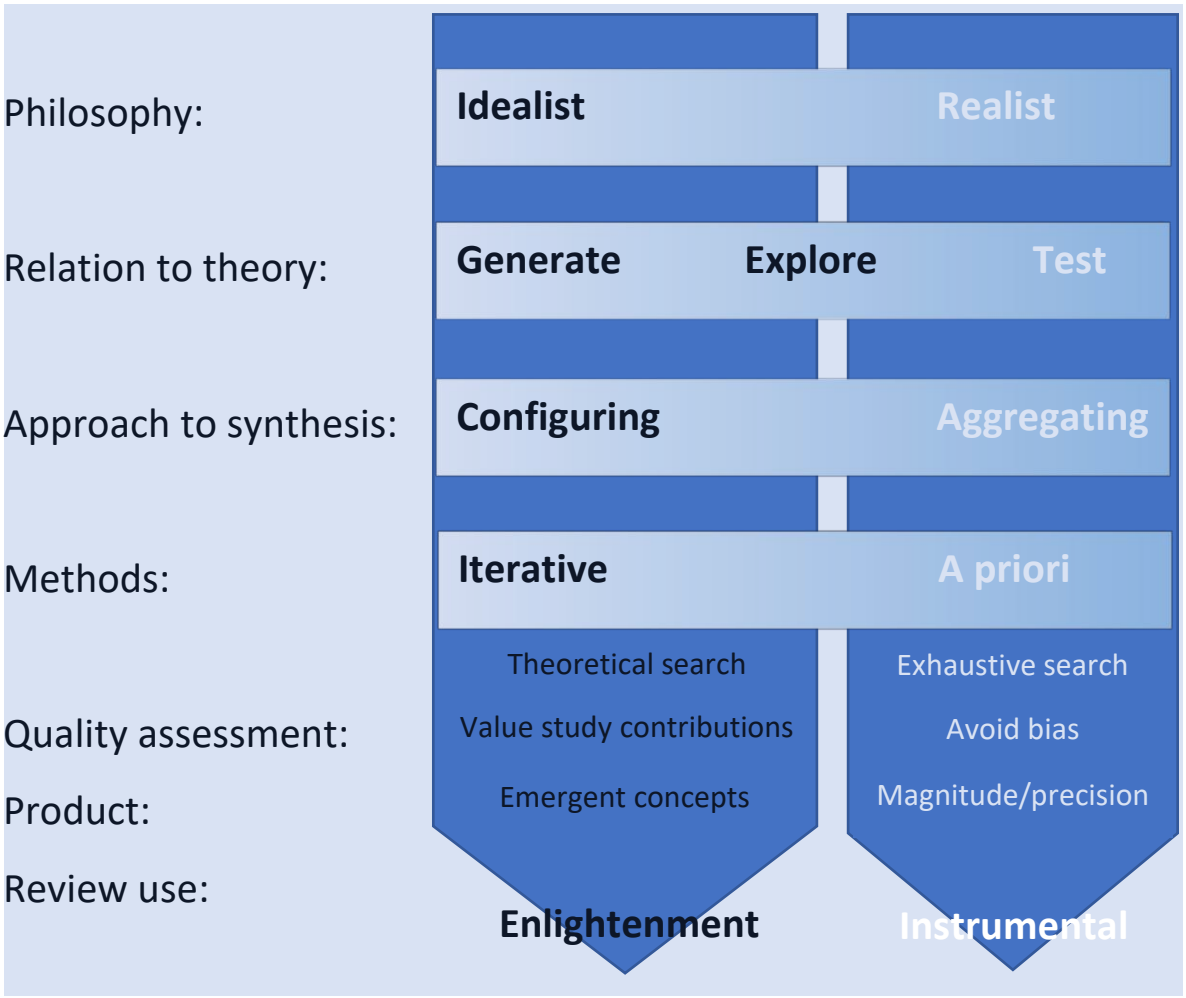


Figure 2: Gough's continua of approaches in aggregative and configurative reviews (Gough, Oliver and Thomas, 2012)

At the other end of the spectrum is 'configurative' synthesis, conceptualised by Gough et al as a mosaic. I will return to this end of the spectrum in Part II.

2.2 Systematic review methods

Systematic reviews address clear and explicit research questions by critically appraising and collating evidence from studies that fit pre-specified eligibility criteria (Higgins *et al.*, 2022b). They use transparent and repeatable methods for identifying, selecting, critically appraising, and synthesising the results of multiple primary studies. Organisations like Cochrane, the Campbell Collaboration, and the Joanna Briggs Institute (JBI) have built robust methods for effectiveness reviews (Aromataris and Munn, 2020; White *et al.*, 2020; Higgins *et al.*, 2022b). These robust processes help to minimise potential bias in the systematic review process and seek to explore and

make explicit any remaining bias in the included studies. Reporting guidance is provided with the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) suite of tools (Page *et al.*, 2021).

A systematic review process can be broken into four main phases: articulating the purpose, scope, and research question; gathering relevant evidence; critically appraising the studies that comprise the evidence; and synthesising data from the included studies. Where the aim is to provide a pooled estimate of effectiveness of an intervention, each of these steps follows robust procedures as discussed below to avoid introducing bias which may skew the results.

2.2.1 Articulating the purpose, scope, and research question

Articulating the purpose and scope is vital for any research project, but pre-specifying and publishing the scope and research question in a protocol is critical for a systematic review (McKenzie *et al.*, 2022). It increases the transparency of the process, allowing a user of the review to see whether proposed methods have been followed, and it avoids the review process being driven by available data, potentially introducing bias. Common tools for articulating the scope include **PICO**(**S/T**) which breaks the eligibility criteria for included studies into **P**opulation, **I**ntervention, **C**omparator and **O**utcome (and sometimes **S**tudy design or **T**ime) (Haynes, 2006) or **SPICE** (**S**etting, **P**erspective, **I**ntervention, **C**omparison, **E**valuation) (Booth, 2004). It is useful to include all stakeholder views at this early stage, including public, patient, carer, clinician, and funder, to ensure that no perspectives or relevant interventions or outcomes are missed.

A protocol should be registered to prove and articulate the advanced consideration of these areas and allow the reader to compare what was planned with what was reported, reducing the likelihood that eligibility criteria could be tweaked later, or pre-specified outcomes neglected. If changes have been driven by data that is known or found by the reviewer, then they may introduce bias in the systematic review. PRISMA-P provides reporting guidance for its content (Moher *et al.*, 2015).

2.2.2 Involving stakeholders: commissioners, clinicians, lay people

Effectively articulating and clarifying the purpose and scope of a systematic review requires multidisciplinary and multi-stakeholder consultation. The commissioner needs to be clear that the proposed protocol is going to answer the question of interest. Experts in the topic area, including professionals and lay 'experts by experience' provide a sense check and ensure that the eligibility

criteria are relevant. Patient and public involvement (PPI) in systematic reviews has only been used relatively recently (Shokraneh and Adams, 2018), but the perspective of those most affected by an evaluated intervention can have a big influence on the conduct of a review.

2.2.3 Gathering relevant evidence

A comprehensive search for eligible studies is crucial and the bedrock of a good systematic review (Lefebvre *et al.*, 2022). If studies are missing from a review, particularly if a group of studies is missing for the same reason, then the results could be skewed regardless of how flawless the subsequent analyses are. Search strategies should be designed by an experienced information specialist in collaboration with topic experts, and assessed by a second information specialist; the PRESS checklist provides a standardised framework against which to do this (McGowan *et al.*). All the functionality of bibliographic databases, such as thesaurus and field searching and Boolean logic, should be used to aspire to the highest sensitivity (percentage of relevant studies found) and specificity (percentage of irrelevant studies discarded), thus reducing the screening burden as much as possible. Grey literature, comprising material not necessarily published in journals, such as conference abstracts, theses, and charity or government reports, should be searched, and the reference lists of eligible and background studies should be checked (Lefebvre *et al.*, 2022). Restrictions on dates of publication and language of material should be as light as possible. These multi-pronged methods help to reduce publication and language bias. Publication bias occurs where studies with more 'interesting' results (whether that be in terms of effectiveness or potential controversy) are more likely to be written up, submitted, and accepted for publication, so that a synthesis of published literature may provide a skewed result and null results are less likely to be published (Vevea, Coburn and Sutton, 2019). Language bias occurs where searches are restricted to a single language and studies published in any other language are missed (Grégoire, Derderian and Le Lorier, 1995; Stern and Kleijnen, 2020). Even if there are not the resources to fully integrate studies in all languages into a systematic review, the searches should not be restricted by language, so that the reviewer can report the number of relevant studies published in different languages and therefore how much of a problem language bias is likely to be. Articulating the methods for gathering relevant studies is important for transparency, and reporting is guided by PRISMA-S (Rethlefsen *et al.*, 2021).

The results obtained from the search need to be carefully screened to select the eligible studies, ideally by two reviewers working independently, to avoid missing eligible studies (Lefebvre *et al.*,

2022). If resources do not allow two reviewers to carry out both screening steps in duplicate, then a percentage should be screened by both followed by a discussion to reconcile disagreements and clarify eligibility criteria (with other stakeholders if necessary) before one screener continues alone. This ensures that screening is taking place against criteria agreed by the whole team.

2.2.4 Critical appraisal of included studies

A thorough critical appraisal of the risk of bias of included studies is required to understand any methodological deficiencies that may influence the analyses in the review, and so it is crucial to carry out a standardised appraisal and use the results to inform the interpretation and conclusions of the review (Higgins *et al.*, 2022a). One of the most commonly used tools for critically appraising RCTs is the Cochrane Risk of Bias tool (Sterne *et al.*, 2019). Further detail is provided in section 3.1.1, where I discuss the risk of bias assessment in my Cochrane reviews (Kaner *et al.*, 2017; Kaner *et al.*, 2018).

2.3 Synthesis methods

The papers included in part I of this thesis used meta-analysis to synthesise the results of eligible RCTs, and so I will describe this method first. Other types of data are considered in part II, and a description of other synthesis methods follows in the rest of section 2.3.

2.3.1 Meta-analysis

Meta-analysis is a statistical technique for integrating the results of individual quantitative studies to provide a single pooled estimate of effect (Egger, Davey-Smith and Phillips, 1997; Normand, 1999). It is one of the most common and powerful synthesis methods for addressing questions about the effectiveness of interventions (Chaimani *et al.*, 2022; Deeks, Higgins and Altman, 2022). By statistically combining data to create a mean effect size that measures treatment effect, weighted by study precision, it increases the power to detect effects of interventions, provides a more precise estimate of a treatment effect than that of the individual studies, and allows exploration of heterogeneity between the studies. A fixed-effect model assumes that there is a single 'true' effect size that all studies would, in a perfect world, estimate; whereas the random-effects model assumes that the estimate from each study is different, although related to, all the other included studies (Borenstein *et al.*, 2009a; Borenstein *et al.*, 2009b).

Meta-analysis can compare two interventions to each other in a *pairwise* meta-analysis or compare multiple conditions at the same time in a *network* meta-analysis (Figure 3b).

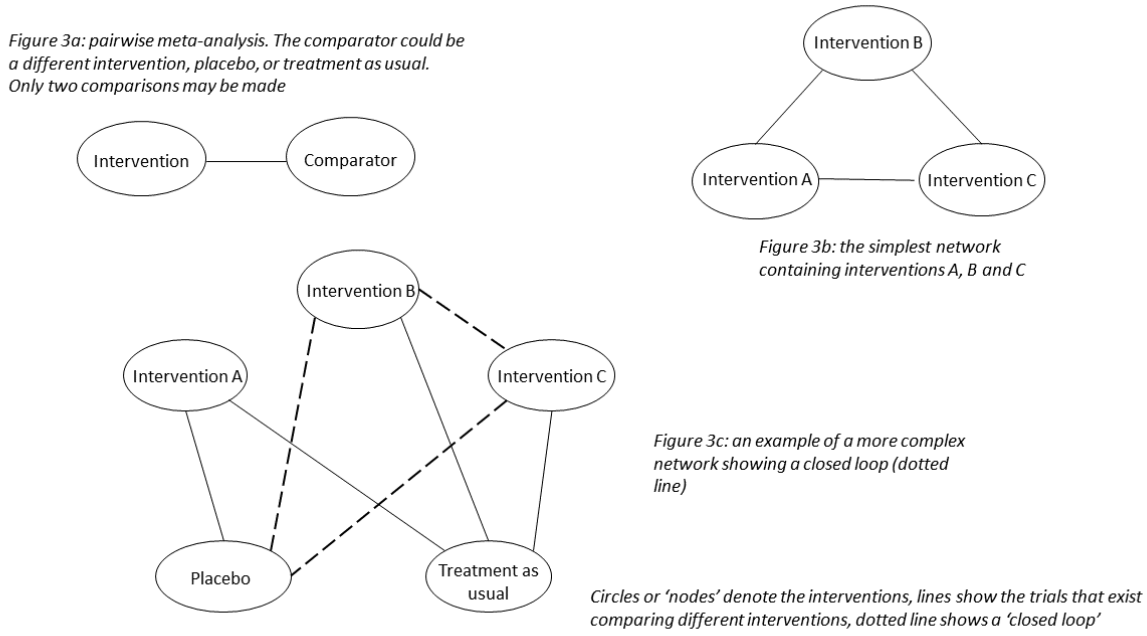


Figure 3: Examples of the geometries of pairwise and network meta-analysis

Pairwise meta-analysis

Figure 3a demonstrates a pairwise meta-analysis – only two conditions can be compared, whether two different interventions or an intervention versus another type of comparator.

Three assumptions guide a pairwise meta-analysis. Firstly, participants are assumed to be successfully randomised between arms so that participant characteristics are equal in every way other than the intervention itself, meaning that any changes in outcome can be attributed to the intervention alone. If one group was systematically different in a way that would affect their response to the intervention, then the pooled estimate would be biased. Secondly, objective outcomes are assumed to be collected blindly so that neither participant nor triallist is aware of the participants' allocations, which avoids unconscious assumptions about how participants 'ought' to fare playing into the outcomes of the trials. For example, if triallists could predict the allocation they may consciously or unconsciously be more likely to assign a participant with a higher perceived need to the intervention group, which could result in higher baseline consumption in the intervention group and a biased result. Finally, the analysis is assumed to be carried out according to the principle of intention to treat, whereby participants are analysed in the group to which they were randomised, regardless of any difference in actual treatment (Senn, 2021).

A pairwise meta-analysis is presented as a forest plot (Figure 4). Figure 4 is an example taken from my work (Beyer *et al.*, 2022) and shows the pooled estimate for the effectiveness of digitally delivered interventions compared to assessment only controls at one month post-intervention. Each line on the plot represents an included study. The point estimates represent a change in weekly consumption and are represented by a black dot. The 95% confidence interval for each estimate is represented by the horizontal line around this dot. The vertical line represents the line of no effect; studies to the left of this line show a benefit of the intervention because drinking less alcohol (lower grams per week) is the desired effect. This meta-analysis reports a high level of heterogeneity (explained further in section 2.42.4) at I^2 = 72.8%, and this can be seen intuitively by the fact that some of the studies do not overlap with each other at all (e.g. Collins 2014 and Pedersen 2017). The weighted pooled estimate appears at the bottom, with the relatively small 95% confidence interval demonstrating high precision made possible by the large number of participants from all the trials.

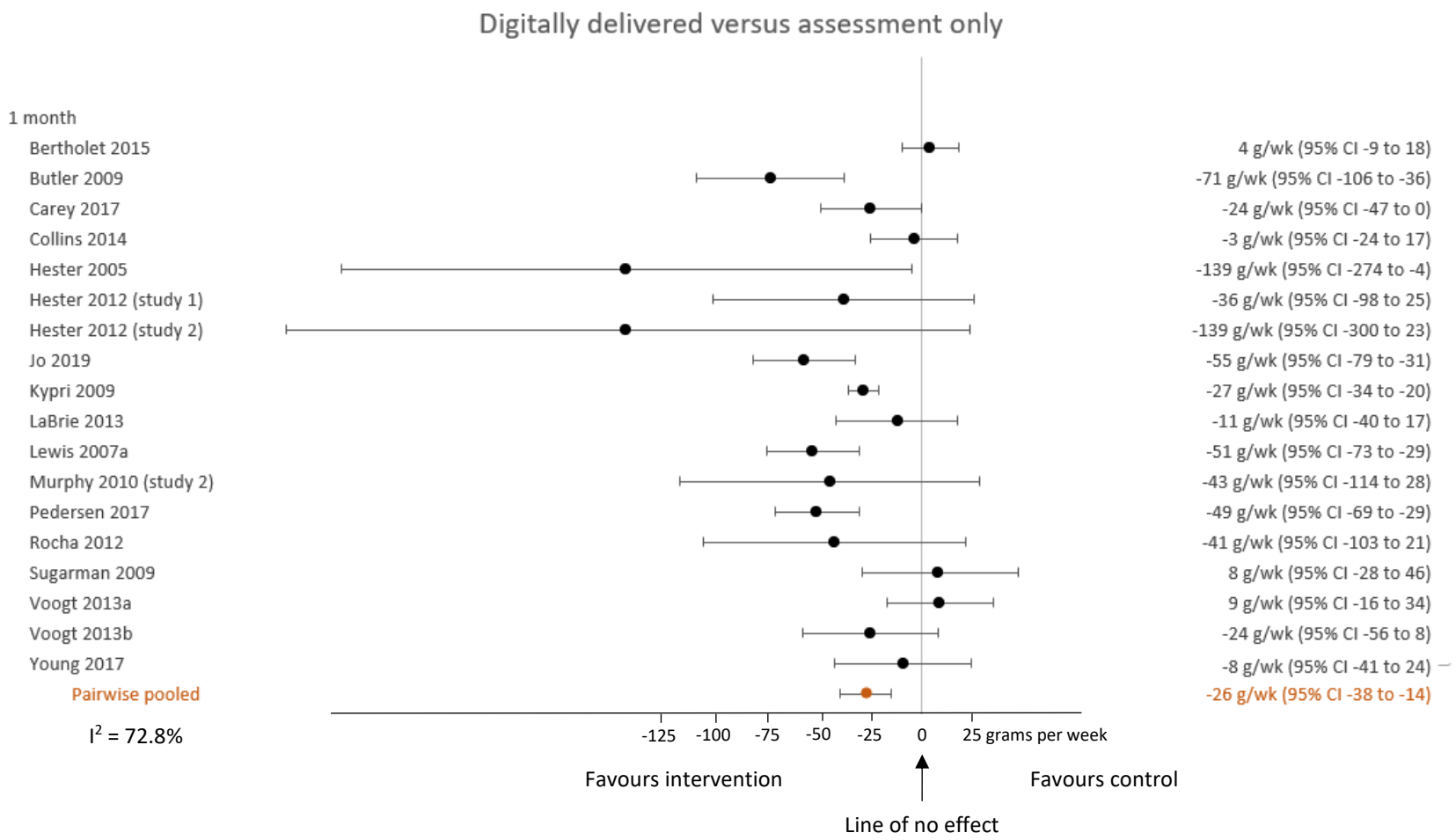


Figure 4: Forest plot, taken from Beyer 2022 (Beyer et al., 2022)

Network meta-analysis

A network meta-analysis can use the data from more than two arms of a trial to inform the analysis (Caldwell, Ades and Higgins, 2005; Efthimiou *et al.*, 2016). Figure 3b demonstrates the simplest version of a network meta-analysis comparing interventions A, B, and C simultaneously. In Figure 3c, there are three interventions of interest (interventions A, B, C) and two types of comparator (placebo, treatment as usual). For each intervention there exist both trials comparing intervention to placebo and trials comparing intervention to treatment as usual. There are also head-to-head trials comparing interventions B and C, but there are no trials that directly compare A and B. Using network meta-analysis we can use all of the indirect evidence (A vs placebo, A vs treatment as usual, B vs placebo, etc) to contribute to an indirect effect estimate for A versus B. Indirect comparisons are non-randomised and provide observational rather than randomised evidence, so may be more subject to biased treatment effect estimates due to confounding – although randomisation takes place *within* trials, it does not inform comparisons *across* the network (Chaimani *et al.*, 2022).

Whereas the assumptions described for pairwise meta-analysis relate to elements of individual trials, further assumptions for network meta-analyses relate to the relationships between trials in the network. Transitivity is the central assumption, and states that the estimate for a comparison between two interventions can be informed by trials that include a third (Chaimani *et al.*, 2022). For example, in Figure 3b, the transitivity assumption states that if no trials exist comparing interventions A and B directly, then the comparisons A-C and B-C can be used to indirectly obtain an estimate for the comparison A-B. This implies several things. Firstly, intervention C should be the same or very similar in the trials of A-C and of B-C that are providing the indirect evidence. If not, estimates for A and B that feed the comparison A-B will not be consistent. A second element of the transitivity assumption is that participants from all trials should be representative of the population of interest and theoretically be able to be randomised anywhere across the network. However, this would not be the case if, for example, interventions in the network took place at different points in the care pathway. Finally, effect modifiers should not be different in the A-C and B-C trials, because again these will affect the indirect estimates that feed the A-C comparison (Jansen and Naci, 2013). For example, if patients in the A-C trials are significantly older that patients in the B-C trials, and age is associated with severity of disease or other measure of poorer outcome, then there will be a bias in the indirect estimates.

Transitivity cannot be statistically tested, but the conceptual 'testing' described above should take place informed by knowledge of the condition and interventions to identify potential weaknesses in the analysis. The statistical manifestation of transitivity is consistency. This is expressed in equation 1, known as the consistency equation (Salanti *et al.*, 2008).

$$\mu_{AB} = \mu_{AC} - \mu_{BC} \qquad \textit{Equation 1}$$

μ denotes the true relative effect between treatments. If this equation does not hold for one of the reasons listed above then the network is said to be inconsistent or incoherent, leading to less reliable results. In a consistent network estimating the relative effect of A versus B, any direct evidence from trials of A-B should be consistent with indirect evidence from trials of A-C and B-C (this is also known as coherence). If there is a difference in the characteristics of the trials in A-C compared to B-C, such as in the effect modifiers, it could mean that the effect sizes reported from these trials are confounded by the effect modifiers, making it impossible to estimate the 'raw' effect size without the modifiers. A consistent network also assumes that any missing evidence is missing at random, so that it cannot bias the indirect evidence (Salanti, 2012).

Consistency can only be statistically tested when there are closed loops in the network, as shown in Figure 3c. Two approaches to testing consistency are available, both of which express a null hypothesis that there is no difference between direct and indirect evidence for a particular comparison, so that a statistically significant difference between direct and indirect estimates defines statistical inconsistency (Efthimiou *et al.*, 2016). A 'loop specific approach' investigates one loop in isolation from the rest of the network and reports the inconsistency factor as the difference between the direct and indirect estimates of a particular comparison. The variance in the inconsistency factor is the sum of the variances of the direct and indirect estimates. The p value comes from comparison of the Z-statistic with a normal distribution. The loop specific approach is useful for pinpointing specific loops that are inconsistent.

However, the loop specific approach quickly falls down in more complex networks where a treatment comparison can be part of more than one loop. For example, in Figure 3c, the comparison *Intervention B vs Intervention C* is a part of both the loop containing *Placebo* and the loop containing *Treatment as usual*. Here the loop specific test can only account for the indirect evidence from the one loop under investigation, missing data from other loops. Where there are

multi-arm studies, loop inconsistency may not be appropriate as there is by definition no inconsistency in a single multi-arm trial. Hypothesis tests for different loops cannot be independent, and if there are many loops then assessing them all will suffer from the multiple comparison problem. An alternative option, the side-splitting approach, compares the direct estimate from a particular treatment comparison to the indirect estimate from the rest of the network. Inconsistency between these two estimates is assessed using a Z-test as in the loop specific approach.

Global approaches to testing for inconsistency are not restricted to particular loops or comparisons but instead assess the network as a whole. The design by treatment interaction model is such a global approach, and this encompasses both the individual loop inconsistencies described previously and inconsistencies between the difference designs (Higgins *et al.*, 2012; White *et al.*, 2012). Design inconsistency can occur whether there are trials containing more than two arms and explores whether A-B comparisons in a network produce similar results whether they are from trials of A-B only, of A-B-C, of A-B-D, and so on.

Component analysis

Component analysis is a type of network meta-analysis that uses individual components of interventions as units of analysis (Welton *et al.*, 2009). Instead of allocating interventions to nodes (Figure 3c), each node represents a component of the interventions. The analysis progresses in the same way as described above, allowing the effectiveness of individual components to be evaluated and compared.

Reporting guidance

Reporting guidance for meta-analyses was first published in 2009, and was updated in 2020 (Page *et al.*, 2021). A PRISMA extension for network meta-analyses is also now available (Hutton *et al.*, 2015). This overlays onto the original PRISMA elements pertinent to network meta-analysis, such as a visual output and explanation of the characteristics of and gaps in the network.

2.3.2 Narrative synthesis

Narrative synthesis comprises methods of synthesising textual data, whether descriptively summarising quantitative (numeric) data or qualitative, interview-based data exploring verbal accounts of lived experiences, to create more than the sum of the individual studies. Further detail about synthesising qualitative data is provided in the next section. It is important to provide a

narrative synthesis of studies in a systematic review in addition to any meta-analysis. This gives the reader a sense of the overall picture before getting into the pooled analyses, and it avoids the risk of studies that cannot be incorporated into any meta-analysis being ignored and not contributing their data to the overall evidence synthesis. Many types of review use narrative synthesis. One of the most commonly used methodological guides for narrative synthesis of quantitative data proposes four stages: developing an a-priori theory of why the intervention works (a 'framework' on which to hang the included study data), developing a preliminary synthesis of findings of included studies, exploring relationships in the data, and assessing the robustness of the synthesis (Popay *et al.*, 2006). One benefit of this method is that it advises the reviewer to consider the theory and logic of how the intervention works *before* getting into the data, to help to avoid being data-led and therefore blindly carrying through to the review results any biases present in the studies. Another is that it encourages the reviewer to actively look for points on which the studies disagree with each other, which reduces the risk of studies with a less commonly expressed viewpoint or discordant data being ignored; this type of 'refutational synthesis' was first proposed as a method for synthesising qualitative studies (Noblit and Hare, 1988a). Finally, because the last stage is to assess the robustness of the synthesis, it is more likely that caveats and limitations will be explicitly described. One of the most common methods of visualising results from a narrative synthesis, aiming to be analogous to the forest plot in meta-analysis, is the harvest plot (Ogilvie *et al.*, 2008; Rohwer *et al.*, 2021). Harvest plots can be used to display the most pertinent features of the data in addition to effect sizes; I created the harvest plots shown in the O'Donnell paper (O'Donnell *et al.*, 2022). Synthesis Without Meta-analysis (SWiM) guidance can be used to report narrative synthesis (Campbell *et al.*, 2020).

2.3.3 Synthesis of qualitative data

Several methods exist for synthesising qualitative data, depending on the richness of the data and the purpose of the synthesis. Meta-ethnography, which has in fact been established for several decades but has more recently been picked up in the service of evidence synthesis, aims to translate the findings of studies using lines of argument between their findings to provide new interpretations, and tends to be used for synthesising rich qualitative data (Noblit and Hare, 1988b; Toye *et al.*, 2014). Another example is thematic synthesis where text is inductively coded and descriptive themes are created as a means of synthesis across studies, this can also be used to

integrate qualitative with quantitative data (Thomas and Harden, 2008; Ring, Jepson and Ritchie, 2011).

2.3.4 Mixed methods synthesis

Often it is most useful to synthesise qualitative and quantitative data together – incorporating the most relevant data regardless of its source (Hong *et al.*, 2017; Noyes *et al.*, 2019; Thompson Coon *et al.*, 2020). The approach to synthesis depends on the type of data available and the research question; for example it could start with a conventional meta-analysis of RCTs and then map qualitative data to the results (Thomas and Harden, 2008), or conversely start with qualitative data if the quantitative data is sparse or of poor quality (Bray *et al.*, 2020). Alternatively, a more integrated approach could synthesise quantitative and qualitative data from the start (Hong *et al.*, 2017).

Realist synthesis draws from combinations of quantitative or qualitative data to develop theory about how an intervention works, for whom and under what circumstances (Pawson *et al.*, 2005; Jagosh, 2019). It aims to provide a more detailed picture of the impact of an intervention than a synthesis of RCTs, which usually reports a single weighted and pooled estimate of effectiveness compared to a comparator. Realist synthesis works by identifying 'context-mechanism-outcome (CMO) configurations', generating different explanations for how the resources provided by an intervention can act in certain contexts to provide particular outcomes.

2.4 Heterogeneity

Heterogeneity is the variability in the trials included in an analysis (Deeks, Higgins and Altman, 2022). Heterogeneity is important because it demonstrates that the intervention may have a different impact across the trials, so exploring these differences and their impact on the effect estimates can provide extra information compared to a pooled meta-analysis estimate.

Clinical heterogeneity refers to the differences in interventions, participant characteristics and outcomes between trials. Interventions are heterogeneous across the trials in my papers, varying in terms of their mechanism of delivery, content, duration, and frequency. They can be delivered by more than one type of professional (who may use their judgement to 'tweak' an intervention), in more than one setting, and can be accepted in different ways (or not at all) by the recipient. The recipient's attitude is more likely to influence the outcome in this kind of complex behaviour change intervention than for a drug – the latter is more likely to 'work' to address the recipient's

illness whether they believe in it, feel keen or well-disposed to it, or not (Skivington *et al.*, 2021). Hazardous levels of drinking are common across the population in many countries, so the participant characteristics are mixed, and some trials are more targeted than others. For example, some trials aim to recruit anyone (usually adult) who is drinking at a certain level, whereas other trials focus on people with certain conditions or in certain settings. The outcomes are very diverse – even where trials measure 'consumption' they do it in many ways, including drinks or units per week (quantity), drinking days per week (frequency), drinks per drinking occasion (intensity), and specially developed screening tools. One commonly used screening tool is the Alcohol Use Disorders Identification Test (AUDIT) which was developed by the World Health Organization (Saunders *et al.*, 1993; World Health Organization, 2001).

Methodological heterogeneity describes differences in study design. This is less of an issue in the body of work described in part I because the eligibility criteria specified randomised controlled trials as the best design from which to estimate effectiveness. However, there was a range of methodological quality which was assessed using version 1 of the Cochrane Risk of Bias tool (further details in section 3.1.1) (Higgins, Altman and Sterne, 2011).

Finally, statistical heterogeneity is 'the statistical variability of results beyond what would be expected by chance alone' (Higgins and Li, 2022). Intuitively, clinical and methodological heterogeneity can be expected to influence statistical heterogeneity – if we can already see many differences in the characteristics of the trials then we might expect variability in the results. Statistical heterogeneity can be assessed visually by comparing the individual study results on the forest plot – if the point estimates all point the same way (i.e. all state that the intervention is or is not effective) and the confidence intervals overlap, this suggests lower levels of heterogeneity compared to if they do not. The chi-squared test can be used and is automatically presented in programmes such as the Cochrane RevMan tool: it tests the null hypothesis that there is no heterogeneity and provides a p value. If the p value is low then it suggests that heterogeneity is present, although it has low power to detect heterogeneity where there are few studies or a low sample size. Finally, the I^2 statistic can be used to assess the influence of the heterogeneity on the analysis – to estimate 'the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance)' (Deeks, Higgins and Altman, 2022).

Exploring heterogeneity in trials can begin to highlight some nuances in the effectiveness of the interventions and contribute towards decision making. Subgroup analysis is used to understand

whether there is a difference in the reported effect size for different groups of studies, for example those containing mostly older people versus mostly younger, or those targeting men versus women (Borenstein *et al.*, 2009d; Deeks, Higgins and Altman, 2022). It provides a separate pooled estimate for each group, allowing the estimates to be compared to each other and to the overall estimate to see whether those in the groups are experiencing different effects of the intervention (Richardson, Garner and Donegan, 2019). Meta-regression can be thought of as an extension to subgroup analysis for continuous data and is conceptually similar to simple regression (Borenstein *et al.*, 2009e; Deeks, Higgins and Altman, 2022). The change in effect estimate is predicted according to the values of one or more proposed study-level covariates (Thompson and Higgins, 2002). Meta-regression cannot infer causation because it does not maintain the randomisation of trial participants within trials; it is therefore an observational design and should be interpreted with care.

There is a tension between including as many trials as possible in the analysis to utilise all available data to obtain a 'big picture' assessment of whether practitioner or digitally delivered interventions are more effective, and ensuring that there is not 'too much' heterogeneity in the analysis. 'Too much' is a subjective judgement, and I felt it was useful to combine the maximum number of trials to inform decision making by understanding the comparative effectiveness of interventions and to explore the heterogeneity, rather than restricting the number of trials in the analysis. In the end, the number of trials in the analysis was restricted by heterogeneity in outcome measures.

2.4.1 Heterogeneity in trial populations

Hazardous drinking has changed over time. In the UK, a small increase in both overall percentage of current drinkers and average number of daily drinks between 1990 and 2016 belies a bigger increase in attributable deaths and DALYs, particularly for men (Table 2) (Griswold *et al.*, 2018). Given that most current drinkers consume alcohol at low risk levels (section 1.3), this suggests that those drinking hazardously are experiencing more harm – so potentially drinking at higher levels more recently.

Year	Gender	Current drinkers (%)	Abstainers (%)	Population average daily drinks	Alcohol attributable deaths (%)	Alcohol attributable DALYs (%)
1990	Female	82	18	2.8	3.8	4.5
	Male	86	14	2.6	3.2	2.2
2016	Female	83	17	3.0	4.0	5.3
	Male	88	12	3.0	5.5	5.3

Extracted from Global Burden of Disease study (appendix 2) (Griswold et al., 2018)

Table 2: Change in alcohol consumption in the UK, 1990 – 2016

Over time, the population in trials recruiting in the general population have become more varied, as shown by I^2 which was 52% in the main analysis in the 2007 version of the Cochrane review (Kaner *et al.*, 2007) and 73% in the more recent version (Kaner *et al.*, 2018). Trials have taken place in populations perceived to be more at risk (e.g. students (Mastroleo, 2018; Hennessy *et al.*, 2019; Plotnikoff *et al.*, 2019)), in different settings (Voogt *et al.*, 2013; Khadjesari *et al.*, 2014; Watson *et al.*, 2015; Guillemont *et al.*, 2017), and different countries (Joseph and Basu, 2017).

This variation in participants may reduce the generalisability of the effectiveness estimate, and it becomes important to explore whether different components of an intervention are effective for people of different ages, genders, or cultures. Where the population was relatively homogeneous in the earlier trials, we could be more confident that the effect estimate applied to that same relatively homogeneous population that was most at risk (in higher income countries at least). One of the issues raised in my PPI discussion about alcohol interventions was to query whether trials account for local cultural, religious, or political situations. For example, if a trial takes place where there is a large cultural or religious group which is less likely to drink, or where there is a recent or ongoing political or environmental crisis, the results may differ compared to trials in other settings. However, now that populations are sometimes more disparate within the same trial, it is more difficult to pinpoint whether there are different elements of an intervention that apply to certain subgroups.

2.4.2 Heterogeneity in interventions

The term 'brief alcohol intervention' covers different types and ingredients of intervention. What started as a 10-15 minute, 'FRAMES'-based intervention delivered face to face in a family doctor's surgery has developed until 'brief interventions' can incorporate elements of motivational interviewing or cognitive behavioural therapy, be delivered in different settings (e.g. hospitals, dentists, schools, or workplaces) and through different media (e.g. websites, SMS, or smartphone app). The original version of the Cochrane review about effectiveness of practitioner-delivered interventions contained less heterogeneity in terms of the interventions because many of them followed this 10-15 min FRAMES framework (median 18 minutes, interquartile range 10 to 40) (Kaner *et al.*, 2007). The most recent update contained many trials of interventions based on Motivational Interviewing with longer and multiple sessions (median 33 minutes, interquartile range 18 to 45) (Kaner *et al.*, 2018). Digital interventions also vary widely in content. They can be delivered through websites, stand-alone computer programmes, mobile tablets, or phones. The least intensive involves a single screen or print-out presenting a summary of the person's drinking habits as collected in the screening or baseline assessment. Although presenting back to a person the information that they have themselves just provided seems simplistic, it is often described as a shock to see a summary of the amount drunk over a week or month, how that consumption compares to others (normative feedback), or how much money is being spent or calories ingested (Lapham *et al.*, 2012; Savic *et al.*, 2016).

2.4.3 Heterogeneity in trial control groups

Another area of heterogeneity in both practitioner and digitally delivered trials was in the content of their control groups. Some trials provided nothing or 'usual care' to their control groups, whereas more recently ethical concerns about screening someone as a hazardous drinker and not intervening have meant that control conditions often provide very brief advice or a leaflet of alcohol-related resources (sometimes referred to as 'minimal interventions'). Large variation exists between trials in duration of procedures, so that a control group participant can take anything from two minutes to one hour to undergo baseline assessment. Another issue is a lack of detail in the description: 'usual care' could mean many things depending on the situation and setting - including advice to cut down alcohol consumption, which could confound the intervention. What this means in practice is that in some cases a 'control' participant in one trial has virtually the same trial experience as an 'intervention' participant in another, so when categorising the arms of

the included studies as ‘intervention’ or ‘control’ I used the reported descriptions of what was provided to the groups rather than the labels applied by authors to their trial arms.

2.4.4 Heterogeneity in trial outcomes

Alcohol consumption was used as the most commonly reported outcome for both Cochrane reviews. It is preferable in behaviour change interventions to separate the target behaviour for change (heavy drinking) from the primary outcome measure (for example harms occurring because of heavy drinking). However, in many of the trials, reporting of harms is absent or inconsistently reported via several different outcome measures, so difficult to synthesise effectively, although they were reported as secondary outcomes in the Cochrane reviews. Consequently, the more commonly reported measure of consumption has been used as a primary outcome in my reviews (as it often is in others).

Various measures of consumption exist, and multiple outcome measures within each. This makes evaluation difficult because without a single or standardised outcome measure, studies cannot be combined in a meta-analysis. The scale of the problem was demonstrated in a systematic review that reported 2641 separate outcomes, measured in approximately 1560 different ways, in 405 trials of alcohol interventions published from 2000 to 2017 by the Outcome Reporting in Brief Intervention Trials: Alcohol (ORBITAL) group (Shorter *et al.*, 2019b). A core outcome set for alcohol brief interventions has recently been reported and will hopefully improve this situation for future systematic reviewers (Shorter *et al.*, 2021).

We selected grams per week consumed as an intuitively understandable primary outcome measure, convertible from commonly reported measures such as ‘drinks per week’ or ‘units per week’, whilst mitigating for the fact that a ‘drink’ and a ‘unit’ could mean different amounts in different countries. An alternative would have been to combine more of the consumption outcomes into a standardised mean difference and report effect size, but this is difficult to interpret and to translate back into a volume of consumption. Other commonly reported measures that are not convertible to grams per week included number of drinking days per week (frequency), number of drinks per drinking occasion (intensity), and scores on validated tools such as AUDIT (Saunders *et al.*, 1993). The secondary outcome was heavy episodic or ‘binge’ drinking, because this describes a different pattern of consumption.

2.5 Confidence in the conclusions of evidence synthesis

Collecting a comprehensive list of studies that address a question, assessing the quality of the studies, and synthesising the studies to assess what they mean is valuable but lacks the final step of articulating how confident the reader can be in the results. A conclusion that an intervention is 'effective' may arise from a large group of high-quality studies that focus on the population of interest to the reader and tell the same story about the balance of benefit and harm, making it unlikely that new studies would change the conclusion. Alternatively, it may come from a smaller group of heterogeneous studies, including some that are at high risk of bias or include different types of patients, where new studies could change the resulting conclusions and recommendations. A common tool to evaluate the confidence in results in this way is the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework (Guyatt *et al.*, 2011; Brignardello-Petersen *et al.*, 2017). We used GRADE to assess confidence in the results of both Cochrane reviews. GRADE was originally conceived to rate the quality of evidence included in clinical guidelines and to grade the strength of recommendations. It provides an explicit, structured, and transparent pathway for considering key features and 'upgrading' or 'downgrading' the quality of evidence ratings. It considers the impact of the domains of risk of bias (formalised in individual study risk of bias assessments), inconsistency (whether the studies show a similar direction of effect), indirectness (how relevant the populations, interventions and outcomes are to the decision that needs to be made), imprecision (how wide the confidence intervals around the effect estimate are), and publication bias on confidence in the evidence.

GRADE has been further developed into a framework called Confidence in Network Meta-Analysis (CINeMA) which has been shown to improve transparency and avoid the selective use of evidence when making judgements about how robust network meta-analysis results are (Nikolakopoulou *et al.*, 2020). Like GRADE, it considers within-study bias, reporting bias, indirectness, imprecision, and heterogeneity, and adds incoherence (how well direct and indirect estimates from the network match).

2.6 Dissemination and knowledge translation

Dissemination and knowledge translation are vital to make the most use of research, demonstrate good use of tax-payers' pounds, and avoid research waste (Wilson *et al.*, 2010; Chapman *et al.*, 2020). Dissemination pushes research out in formats that are accessible to all relevant

stakeholders: this may include a short report, video, podcast, or graphic piece with a focus on plain English for lay readers, a policy briefing for policy makers, or a journal article or conference presentation for researchers. Knowledge translation actively involves those that will be influenced by, use, or recommend the results of the research, and encompasses exchange and application of knowledge (Straus, Tetroe and Graham, 2009). This means being involved throughout: for policy makers this entails working closely with researchers in defining the scope of research questions and interpretation of results, for intervention recipients it means co-design. Meaningful co-design finds a mutual vision between researchers and participants, sets clear roles and responsibilities for participants, ensures that communication and training or coaching is timely and helps participants to fully understand, and rewards participants for their time and contribution (Slattery, Saeri and Bragge, 2020).

Chapter 3. Effectiveness and cost effectiveness of alcohol interventions

Having described the content and methods of systematic review and meta-analysis and what makes them robust, I will now critique the methods of the papers I have published concerning the effectiveness of alcohol interventions and describe their results and implications. Section 3.1.1 critiques the methods of my reviews of practitioner and digitally delivered interventions, each compared to no or minimal intervention or treatment as usual. Section 3.1.2 explains how I explored heterogeneity in these trials. Section 3.2 describes the papers I wrote distilling this evidence for clinicians and policy makers and proposing an analysis directly comparing practitioner delivered to digitally delivered interventions. Section 3.3.1 critiques the methods of this network meta-analysis; although this work is chronologically the last of my papers and the culmination of my journey through this PhD in terms of learning and leadership, it fits into this section because it addresses effectiveness using meta-analysis and builds on the papers discussed in section 3.1. Section 3.4 describes knowledge translation activities, and section 3.5 puts the results of all these analyses into context. Finally, section 3.6 reflects on the methodological issues I encountered whilst carrying out these reviews and what they mean for evaluating effectiveness of complex behaviour change interventions.

3.1 Effectiveness of interventions compared to comparator

My first paper is a Cochrane review of the effectiveness of brief alcohol interventions for reducing hazardous and harmful consumption in people seeking treatment in primary care for conditions other than their drinking (Kaner *et al.*, 2018). This review was first published in 2007, when brief interventions were mostly delivered in primary care settings and digitally delivered interventions for hazardous alcohol consumption were not yet well known (Kaner *et al.*, 2007). In 2017-8 I led an update of this review which increased the number of included trials from 29 to 69, and demonstrated that these interventions were being used in emergency as well as primary care settings (Kaner *et al.*, 2018). This review reports the effectiveness of what I came to refer to as practitioner delivered interventions compared to no intervention, treatment as usual (for the non-alcohol-focused presenting condition), or minimal intervention (very brief advice or written resources). This work was also summarised as a journal article for clinicians and policy makers (Beyer *et al.*, 2019).

Paper 1a – Cochrane review

Kaner EFS, **Beyer FR**, Muirhead C, Campbell F, Pienaar E, Bertholet N, et al. Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database of Systematic Reviews*. 2018;2
https://doi.org/10.1002/14651858.CD004148.pub4

Paper 1b – summary for clinicians and policy makers

Beyer FR, Campbell F, Bertholet N, Daeppen JB, Saunders JB, Pienaar ED, et al. The Cochrane 2018 review on brief interventions in primary care for hazardous and harmful alcohol consumption: a distillation for clinicians and policy makers. *Alcohol and Alcoholism*. 2019;54(4):417-27
https://doi.org/10.1093/alcalc/agz035

The next paper is also a Cochrane review and is analogous to the first but concerns digitally delivered interventions compared to no or minimal intervention or treatment as usual (Kaner *et al.*, 2017). This was part of a piece of work that was also funded to explore the content of the digital interventions using behaviour change techniques, and the use of theory in the development of the interventions (discussed later as Garnett 2018a and Garnett 2018b respectively).

Paper 2

Kaner EFS, **Beyer FR***, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. *Cochrane Database of Systematic Reviews*. 2017;9
https://doi.org/10.1002/14651858.CD011479.pub2
** corresponding author*

3.1.1 Systematic review methods – Cochrane reviews

A peer reviewed protocol was registered for each review on the Cochrane Library (Kaner *et al.*, 2004; Kaner *et al.*, 2015), and the most recent was automatically published on Prospero (CRD42015022135). I became involved shortly after the protocol for practitioner delivered interventions was published in 2004 so I am not an author on it, and it was published before Prospero was launched in 2011. Clear rationales for all amendments from the protocols are provided in each systematic review ("Differences between protocol and review") to demonstrate that they were unlikely to influence the results.

The search design and screening processes were followed as described in section 2.2.3.

Critical appraisal of included studies

We tested the assumptions of pairwise meta-analysis using the Cochrane Risk of Bias tool (Higgins, Altman and Sterne, 2011). Although Risk of Bias 2.0 is now available, it was not when the studies were being appraised for these reviews (Sterne *et al.*, 2019). The first assumption of successful randomisation was addressed with the first domain of the tool, which considered whether allocation to groups was truly random and concealed from all involved in the trial. A red flag for unsuccessful randomisation exists if a statistically significant difference between groups is reported in their baseline characteristics. This domain was not a big issue in this body of trials: 10% and 4% of trials were at high risk of bias in practitioner and digitally delivered trials respectively (although 63% and 42% were judged unclear in this domain).

For the second assumption of objective outcomes collected blindly, the risk of bias tool considered whether those providing and receiving the intervention, as well as outcome assessors carrying out trial analyses, were blind to who was in which group. Blinding is impossible where participants interact with a person to discuss alcohol consumption, although some triallists tried to blind participants by 'disguising' the alcohol content in other health-related material. Often, trials of digital interventions were more likely to be judged at low risk of bias in this domain because they took place entirely online: participants joined the trial, were randomised, received the intervention, and reported their outcomes without encountering trial personnel at all, so there was no likelihood of the latter influencing the trial. We assigned a judgement of low risk of participant blinding if triallists had concealed either the fact that participants were in a trial, or the nature of the trial, by 'hiding' questions and information about alcohol consumption amongst

other health-related information. Finally, trials were assumed to report on all randomised participants according to the intention-to-treat principle.

We tested the robustness of these findings according to the different assumptions using sensitivity analysis, omitting studies judged at high or unclear risk of bias from the main analysis to see whether they change the overall pooled estimate. The results were robust to these analyses: in the practitioner delivered review there was very little change in the estimates, and for the digitally delivered review the sensitivity results were smaller but still statistically significant.

Ideally, all the included studies would have contributed data to the meta-analyses. However, only a subset of studies could be included in all the meta-analyses because there is such a variety of outcomes in the field of alcohol consumption (Shorter *et al.*, 2019a). Missing studies from the meta-analysis risked a biased pooled estimate if the studies included were systematically different in some way to those that weren't. However, all consumption outcomes were self-reported by participants either keeping a diary, using a timeline follow back method, or unaided memory, so all equally likely to over- or under-estimate consumption regardless of the actual effect measure. Self-reporting of outcomes has been shown not to be a big problem in alcohol research, at least for students (Kypri *et al.*, 2016).

In both Cochrane reviews I believe that it would have given a better account of all the evidence to provide a narrative synthesis that included results from *all* studies before the results of the meta-analyses were presented. Although the 'Characteristics of studies' and 'Risk of bias' sections of the results contained data from all studies, in each case the 'Effects of interventions' section goes straight to the meta-analyses, hence ignoring 51% and 28% of the included studies respectively. Both Cochrane reviews state in the discussion that the direction of the evidence is broadly consistent based on the studies included *in the meta-analysis*, really this statement should have encompassed *all included studies*. I don't believe that this omission has impacted the results of the Cochrane reviews in this case, because the reported effects of the interventions do point in the same direction across more than 90% of included studies, whether they appear in the meta-analysis or not. However, in a review of studies reporting more varied results, this omission could have introduced bias to the review results if studies that favoured the intervention were either more or less likely to be included in the meta-analysis.

Synthesis: pairwise meta-analyses

The weighted mean difference method was used to estimate pooled effect sizes for the primary outcomes and 95% confidence intervals (Khan, 2020). This was appropriate because the chosen primary outcome measure was mean difference in grams per week of alcohol consumed, which was the same across intervention and comparator arms in all studies and therefore could be used without conversion or standardisation. An alternative to this would be to use a standardised mean difference, or to standardise the different measures of volume to a uniform scale. However, because there are so many different reported consumption outcomes and because the resulting effect size would not give an intuitive understanding of the actual effect of interventions, grams/week was chosen. Hopefully future triallists will heed the core outcome set for alcohol brief intervention trials reported by the ORBITAL team (Shorter *et al.*, 2019b).

This raw outcome of grams per week was weighted by the inverse variance (or precision), so that more precise estimates from larger studies were given more weight in the analysis. The use of a random-effects rather than a fixed-effect model was appropriate (DerSimonian and Laird, 1986; Borenstein *et al.*, 2009c; Deeks, Higgins and Altman, 2022) because the anticipated levels of heterogeneity did not suggest a common effect size as assumed by a fixed-effect model. Despite the common name of 'brief intervention', they could be based on different styles (e.g. basic advice or motivational interviewing), contain different components, or be delivered by different health professionals. All studies that reported an outcome convertible to grams per week were included in the meta-analyses. This allowed the maximum number of trials to contribute to the pooled estimate, particularly small studies that themselves do not have the power to demonstrate an effect.

3.1.2 Exploring heterogeneity

It was important to explore potential heterogeneity (section 2.4) to understand if there were any differences in effectiveness of the interventions depending on their content or the characteristics of those using them. In this section I will discuss the subgroup and meta-regression analysis undertaken as part of the Cochrane reviews.

Subgroup analysis

To avoid bias and data-led analyses, studies were separated into groups according to pre-specified plausible rationales developed at the protocol stage, rather than identifying subgroups after data

extraction in the light of available data. In both Cochrane reviews, subgroup analyses were carried out according to gender and age. In the review of practitioner delivered interventions, studies were subgrouped by setting (primary care or emergency care) and type of intervention (advice-based or counselling-based). Subgroup analysis according to setting was not carried out for digitally delivered intervention analysis because the interventions were often mobile and could be accessed according to participant preference, rendering an analysis by setting meaningless. In the review of digitally delivered interventions, subgroup analyses were planned according to modality of intervention (e.g. website, smartphone app) but there were insufficient types of modalities – most published evaluations were of websites. It would have been useful to explore whether the interventions affected inequalities and we looked for data on income, ethnicity, employment, or educational attainment as proxies for this. Although this data was sometimes reported at baseline across all participants, it was reported by trial arm in very few studies.

The Cochrane reviews were not consistent in the methods used for sensitivity analysis according to risk of bias: the review of practitioner delivered interventions restricted to trials at low risk of bias, whereas the review of digitally delivered interventions omitted trials at high risk of bias (therefore including those at unclear risk of bias). Trials judged at unclear risk of bias are usually judged that way due to deficiencies in reporting, so these trials may actually be at high or low risk of bias, and one might assume that they include both. A sensitivity analysis that restricts to trials at low risk of bias is certain to contain only the most robust trials but is likely also to have missed some, whereas an analysis that omits trials at high risk of bias uses more of the relevant data but may incorporate some less robust trials.

Meta-regression analysis

The next paper describes an estimation of which components of these complex alcohol interventions, if any, were most associated with the intervention effect (Garnett *et al.*, 2018b). BCTs (section 1.5) (Michie *et al.*, 2013) were used to code intervention components, and meta-

Paper 3

Garnett CV, Crane D, Brown J, Kaner EFS, **Beyer FR**, Muirhead C, et al. Behavior Change Techniques used in digital behavior change interventions to reduce excessive alcohol consumption: a meta-regression. *Annals of Behavioral Medicine*. 2018;52(6):530-43 https://doi.org/10.1093/abm/kax029

regression analysis was used to explore the effectiveness of the intervention associated with each BCT (Garnett *et al.*, 2018b). This analysis was conducted only for the trials of digitally delivered interventions; other members of the team had previously performed a similar exercise with the practitioner delivered interventions (Michie *et al.*, 2012).

Detail of the intervention content is often lacking in journal articles because of the restricted word count, so we contacted authors to retrieve associated papers or manuals so that the full detail of the interventions could be coded. Not all responded, so some BCTs may have been missed or mis-coded for some of the interventions. Further, BCTs must be explicitly described to be coded as included, so even with a more thorough description of an intervention, if the appropriate language is not used, a technique that is present may not be coded.

The meta-regression analysis cannot work for BCTs which occur in all (or nearly all) of the interventions because there needs to be a set of interventions containing the BCT and a set of interventions without, in order to compare the effectiveness with and without it. The included set of digital alcohol interventions by definition had to give feedback (this was part of the eligibility criteria), so this analysis could not explore the impact of feedback on effectiveness.

Whereas BCTs were unlikely to be present in 'no input' control arms, where 'treatment as usual' was provided it was unclear which, if any, BCTs were represented because they were usually poorly reported. BCTs were rarely coded from control arms, but this was highly confounded by reporting bias because control group provision was even more poorly reported than the detail of the intervention.

For the next paper we carried out a meta-regression to explore whether using behaviour change theory to develop the digital interventions had any influence on their effectiveness (Garnett *et al.*, 2018a).

Paper 4

Garnett C, Crane D, Brown J, Kaner E, **Beyer FR**, Muirhead C, et al. Reported theory use by digital interventions for hazardous and harmful alcohol consumption, and association with effectiveness: meta-regression. *Journal of Medical Internet Research*. 2018;20(2) https://doi.org/10.2196/jmir.8807

The Theory Coding Scheme was used to articulate the use of theory for developing the digitally delivered interventions (Michie, Webb and Sniehotta, 2010). This was a useful framework because it had been tested previously on a set of complex interventions and can be grouped into six different categories of theory use including the way that theory is used, tested, or refined. A meta-regression was conducted to explore whether the use of theory was associated with intervention effectiveness.

3.2 Dissemination

Because Cochrane reviews follow gold standard methods in terms of both conduct and reporting for effectiveness questions using meta-analysis, they are long, detailed, and technical, and must be disseminated for different audiences. Cochrane reviews contain plain language summaries for lay readers, along with a short scientific abstract and summary of findings tables. They also feature in blogs and podcasts produced by Cochrane (Beyer, 2017). There is a gap for a succinct summary that contains a little more context and explanation. I wrote a distillation of the practitioner delivered review for a clinician audience (Beyer *et al.*, 2019) (paper 1b), along with a paper that drew together the evidence from both Cochrane reviews and called for more research to understand the relative effectiveness of practitioner versus digitally delivered interventions (Beyer, Lynch and Kaner, 2018).

Paper 5

Beyer FR, Lynch E, Kaner E. Brief interventions in primary care: an evidence overview of practitioner and digital intervention programmes. *Current Addiction Reports*. 2018;5(2):265-73
https://doi.org/10.1007/s40429-018-0198-7

This paper was useful for the policy maker because it was one of the first to bring together all the evidence about the effectiveness of both practitioner and digitally delivered interventions, and to consider them together. It also provides a concise description of some of the thornier issues in this field (discussed in more detail in this PhD), such as whether reported small effect sizes are meaningful, the fact that participants in control arms often decrease their consumption as well as intervention participants - making the difference between arms statistically insignificant, sources

of heterogeneity, difficulties with blinding those involved in trials, and of pooling trials when so many different outcome measures are reported.

3.3 Effectiveness of practitioner versus digitally delivered interventions

The pairwise meta-analysis could not answer the question of comparative effectiveness between practitioner and digitally delivered interventions because there were insufficient trials comparing them directly. It is unclear why so many RCTs of digitally delivered interventions used no or minimal intervention instead of practitioner delivered intervention as the comparator, given that the latter have been recommended in multiple guidelines for at least a decade. However, in the absence of many head-to-head trials I decided to conduct a network meta-analysis to make use of the available indirect evidence to inform this question (Beyer *et al.*, 2022).

Paper 6

Beyer FR, Kenny RPW, Johnson E, Caldwell DM, Garnett C, Rice S, et al. Practitioner and digitally delivered interventions for reducing hazardous and harmful alcohol consumption in people not seeking alcohol treatment: a systematic review and network meta-analysis. *Addiction*. 2022
https://doi.org/10.1111/add.15999

3.3.1 Systematic review methods – network meta-analysis

The protocol for the network meta-analysis was registered on Prospero (CRD42018089609), and a detailed methods and analysis plan were written and circulated to the author team for comment and improvement before embarking on the project. I presented the protocol to a PPI group and noted the issues that they raised in response (Table 3). A change that was actioned within the protocol as a result of the PPI input was to include heavy episodic drinking as one of the main outcomes, in addition to mean grams per week. Other points fed into the discussion section of the paper.

Recommendation	Response
<u>Outcomes</u> **Binge drinking** is an important outcome because it is a common pattern of social drinking.	Discussed whether to use binge drinking as primary outcome, but decided to use weekly consumption because it is a standardisable measure. We used binge drinking as the secondary outcome.
Blood markers might give a more accurate picture than self-reported consumption	Although it seems intuitive that blood tests would give a more accurate representation of alcohol consumed, research shows that they do not give a good estimate of binge drinking, only steadier drinking. Few of the studies reported the results of blood tests, but better tests may be used in future trials, so updates of this work may include more trials that report blood tests.
Some felt that **health outcomes** are just as important as consumption outcomes, people don't always make the connection between their drinking and impacts on their health, and this might motivate them to reduce consumption. However, others felt it best to focus on consumption in this project.	Although alcohol is a **risk factor** for many diseases, we often cannot point to alcohol as directly **causing** disease, and we focused on consumption outcomes as the most directly attributable impact of the interventions.
Longer term outcomes – it is important to understand how these interventions impact people in the longer term	For this study, we are restricted to the follow-up time of the trials, which is rarely more than one year due to the expense of running randomised controlled trials.
Economic outcomes and influences - do the price of a unit of alcohol or the available income of the drinker have an impact?	Part of this project aimed to develop an economic model to compare the cost-effectiveness of the different interventions. These specific questions are the subject of ongoing research by other teams (for example minimum pricing). This research aims to understand the impact of the brief interventions, and although drinking may also be affected by the price per unit and available income, the data is not available from the included trials to make any connection.

Recommendation	Response
Outcomes by **population characteristics**, for example impact of intervention may differ by gender, age, employment status, family situation (e.g. carers), preferred type of drink	We carried out subgroup analyses on these factors where reported data allowed.
Change in understanding or knowledge as an outcome – people do not understand how many units they drink, because it is unclear what a unit means, and often drinks are not measured out. Over time the volume of a 'glass' of wine has increased	This is not well reported in trials and is the subject of ongoing research by other teams.
Labelling of alcohol as an outcome - current labelling is not helpful on alcoholic drinks, and that better labelling might be a useful outcome	Labelling is an interesting issue and a topic of ongoing research by other teams. The interventions under investigation here aim to change people's drinking by providing them with feedback and advice.
<u>What is missing from this research proposal?</u> **Addiction versus habit** - people might think of their hazardous consumption as a habit rather than an addiction, and this affects the perceived danger to health because a 'habit' is harmless whereas an 'addiction' has connotations that are more negative.	It is true that if people don't feel negative consequences to their health and use the term 'habit' then they are unlikely to see the need for change. Part of the purpose of the interventions under investigation is to help people recognise if their drinking is hazardous.
Impact of written information - compared to smoking and the fact that smokers ignore warning messages on cigarette packets, you might imagine that drinkers would ignore the written information about the hazards of drinking received in control conditions in the trials. This would mean there would be no difference in outcome between the trials with alcohol-related control content and those with general health-related control content.	This provided one of the rationales for splitting the control group according to whether participants received alcohol-related information or not, and I tested this hypothesis that alcohol-related information in the control group does not influence consumption outcomes.

Recommendation	Response
Individual participant outcomes should be reported because an average is meaningless.	What the average pooled result of the previous reviews tells us is that brief interventions are likely to reduce consumption for at least 95% of people who drink hazardously or harmfully, so there is a very good chance that an individual would benefit. We can also say that because many people drink hazardously, if they all receive a brief intervention most of them are likely to reduce their drinking so there would be a relatively large impact on population health. There is a type of analysis that uses individual patient data, but this is beyond the scope of this project – it is something to consider in future work.
There is no good alternative to drinking alcohol, especially in pubs where binging is common. Non- or low-alcohol alternatives never taste good, and pop (another alternative) is expensive in pubs. They highlighted that now there is an alternative to smoking (e-cigarettes) and many people have quit; if there was a realistic and tasty alternative to drinking then people might be more likely to cut down their alcohol consumption.	The type of alcohol consumed is rarely reported in trials. [*In the years since this PPI panel took place many more alcohol-free alternatives have become available*].
<u>Concerns or queries about the project?</u> **Definition of intervention** - what does 'brief' mean, and does it include further support after the initial session? For example, in weight loss classes multiple sessions sustain motivation.	For the practitioner delivered interventions, 'brief' usually means 15-20 minutes for the advice-based interventions and up to 60 minutes for the counselling-based interventions (which are still called 'brief' in the literature). Some of the practitioner delivered interventions also provide what they call 'booster' sessions, usually in the form of between one and three 10-minute phone calls. For digitally delivered interventions people might use them once or many times but usage patterns are not usually reported.

Recommendation	Response
Implementation - how would a person receive a brief intervention from their GP when a standard consultation (in the UK) is not long enough? How would they receive a digital intervention?	It is true that standard consultations are not long enough for many brief interventions, and in practice, people are screened and identified as at-risk in the consultation but re-directed to a nurse or alcohol health worker for the actual intervention. This may be something that needs to be discussed in the guidelines that are produced at the end of the project. Digital interventions could be 'prescribed' in the same way as the face-to-face intervention. The doctor could provide a QR code (square barcode) in the surgery for reading onto a mobile device or email a link to the intervention.
Cultural, geographic and temporal differences - alcohol consumption is viewed very differently between the UK and other countries. People in different countries sometimes have a different (or non-existent) relationship with alcohol for religious or cultural reasons. Relationships with alcohol may change in a single location over time, for example because of war, economic crises, unemployment, or legislative changes. For these reasons, baseline consumption may vary a lot between different countries, and perceived 'problem' levels of consumption may be very different. Levels of 'social desirability bias', where people underestimate their consumption to a more 'acceptable' level, would vary between different cultures too.	In this project, we are only able to report as much as the included trials reported, but we will investigate whether included trials consider these issues and raise these points in the discussion. The aim of the randomisation process in these trials is to ensure people with different characteristics are equally shared between the control and intervention groups, so social desirability bias should affect both groups equally in randomised trials.
<u>Generalisability, reliability, validity of the data</u> Is it possible to define **one 'recommended' drinking pattern** for everyone, when there are so many different patterns and reasons for drinking?	In order to explore this area, we would need to carry out qualitative research, asking people what they think about the reasons for drinking and what would help them to cut down. The previous systematic reviews both reported that most people do cut down their consumption in response to these interventions, which suggests that, to some extent, whatever their reason for drinking, most people do respond.

Recommendation	Response
Is there a **large enough data set** to address the question? Is from so many different types of people in different contexts that it may not be reliable?	Some of the studies contained 'power calculations' designed to ensure that they recruited enough people to be able to observe a difference in consumption between groups, although others didn't. More than 65,000 participants took part in all of the included studies across both Cochrane reviews, but it is true that there was a range of baseline characteristics (e.g. alcohol consumption) and contexts.
<u>Availability and marketing</u> Panellists were concerned about how easily and cheaply alcohol is available, particularly with special offers, happy hours and so on.	This is the subject of research by other teams and is starting to work its way through to policy with initiatives such as minimum unit pricing. It is beyond the scope of this project, although these issues may provide some of the background section in the write-up.
Implementation It is great to have a project that is gathering many trials together and using their data because we see so much very small discrete research. It is important to apply and use previous research. Implementation: it is important to get the results of the project out of the research arena into the 'real world'. We should (i) produce a definitive statement of the best approach; and (ii) create a pathway, tool, or recommendations that can be used by health professionals to ensure the best ideas are implemented into care.	The results of this work will be presented through a conference presentation, journal article and policy briefing.

Table 3: Key points from presentation of the protocol to a PPI panel (note: because this pertained to the project protocol, some of the responses refer to work that 'will be' done).

Gathering relevant evidence

Given the number of existing systematic reviews and the workload involved in searching for trials across 30 years of both practitioner and digitally delivered interventions and all settings from scratch, I decided to utilise the searches carried out for existing published systematic reviews and to 'top them up' using my own searches from the earliest search date of these reviews. This was not as robust a procedure as for the Cochrane reviews – more robust would have been to carry out comprehensive searches of all the databases with no date limit, which may have been less likely to miss relevant studies but would have had an enormous impact in terms of time required to screen tens or hundreds of thousands of records. Critically appraising and harvesting eligible studies from candidate systematic reviews was a pragmatic way to utilise previous work and reduce the required resource.

The screening processes were followed as described in section 2.2.3.

Critical appraisal of included studies

The assumptions for a network meta-analysis begin with the same assumptions described for pairwise meta-analysis – that is, successful randomisation within trials, blinded collection of objective outcome data, and analysis by intention-to-treat. Therefore, for the trials from the Cochrane reviews that were also included in the NMA these same risk of bias assessments were used. The risk of bias of other included trials were conducted by two reviewers using the Cochrane Risk of Bias tool version 1.0 to be consistent with the previous assessments.

The variability in follow-up time (one month to four years, with almost every week being represented up to three months, and every month being represented up to 12 months) made it difficult to know how to combine the results for pooling, since the Cochrane reviews of practitioner and digitally delivered interventions suggested a decay in effect. Splitting into too many different time points could mean that there would be too few trials to give a robust result at each time point, but combining into too few would disguise any decay, which is important to facilitate understanding of whether and when repeat treatment can be best administered. Another challenge with this body of evidence was that trials of digital interventions tended to report from one month and up to six months, whereas trials of practitioner delivered interventions tended to report from six months and up to 12 months. At one month follow-up results from ten trials of practitioner delivered and 24 of digitally delivered interventions were

available for the analysis, whereas at 12 months 35 trials of practitioner delivered and nine of digitally delivered interventions were available. I carried out analyses combining trials that reported at one month, six months, 12 months, and in order to capture most data a separate analysis was carried out capturing the longest follow-up from all trials reporting *up to* 12 months.

Synthesis: network meta-analysis

The network meta-analysis contained a narrative synthesis at the top of the results that considered the direction of effect from all included studies, before going on to report the results of the meta-analysis. This made it more robust as it accounted for all the available evidence rather than only those studies in the meta-analysis, and in fact demonstrated that 94% of all the included studies showed a decrease in consumption in the intervention arms.

The simplest network for the trials available to compare practitioner with digitally delivered interventions would have been to incorporate evidence from all the many trials that compare either of the interventions to a control group (Figure 5a). This is equivalent to Figure 3b, and so if the transitivity assumption were to hold then the control groups in all trials would be equivalent. This was clearly not the case for these trials; control participants received a variety of input across the trials, from baseline assessment only to direct advice about reducing their alcohol consumption. Between these extremes are 'attention control' conditions, where control participants take part in an activity unrelated to alcohol but similar in duration to the intervention group (such as a video game or discussion about their general health), and written information about resources for those looking to reduce their alcohol consumption (which may account for the fact that all participants often reduce their consumption in alcohol trials). To mitigate this, I split the control conditions into two separate nodes as shown in Figure 5b: control conditions that received baseline assessment only, and control conditions that provided alcohol-related input. This allowed me to test the hypothesis that material related to heavy drinking in the control groups would affect drinking levels and so these trials would report smaller effect sizes than those with non-alcohol-related input. It wasn't clear where to allocate the 'attention control' condition because they provide neither assessment only nor alcohol-related information. They could have gone in a node of their own, but this would have meant that there were more 'control' nodes than 'intervention' nodes in the network. This felt like too much of an artificial split and risked having too little data in some of the 'control' nodes. Alternatively, they could have been 'lumped' with the 'minimal intervention' node: this makes logical sense because it would explore the placebo

effect, given that it would compare participants that received the *minimum* amount of input (assessment only) with those receiving *some* kind of input. However, I decided to lump the 'attention control' condition with the assessment only group because this achieved the secondary aim of isolating the trials that provided some kind of alcohol input to work out whether this could have an independent effect on alcohol consumption compared to the interventions.

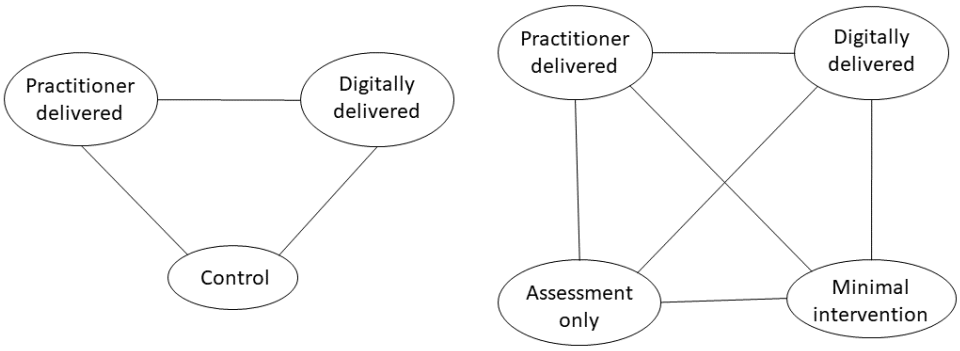


Figure 5a (left): simplest network for the alcohol trials
Figure 5b (right): network accounting for the fact that 'control' arms are not all comparable; 'minimal intervention' includes alcohol-containing comparator groups
Circles or 'nodes' denote the interventions; lines show the trials that exist comparing different interventions

Figure 5: Options for the network for practitioner and digitally delivered alcohol interventions

3.3.2 Exploring heterogeneity

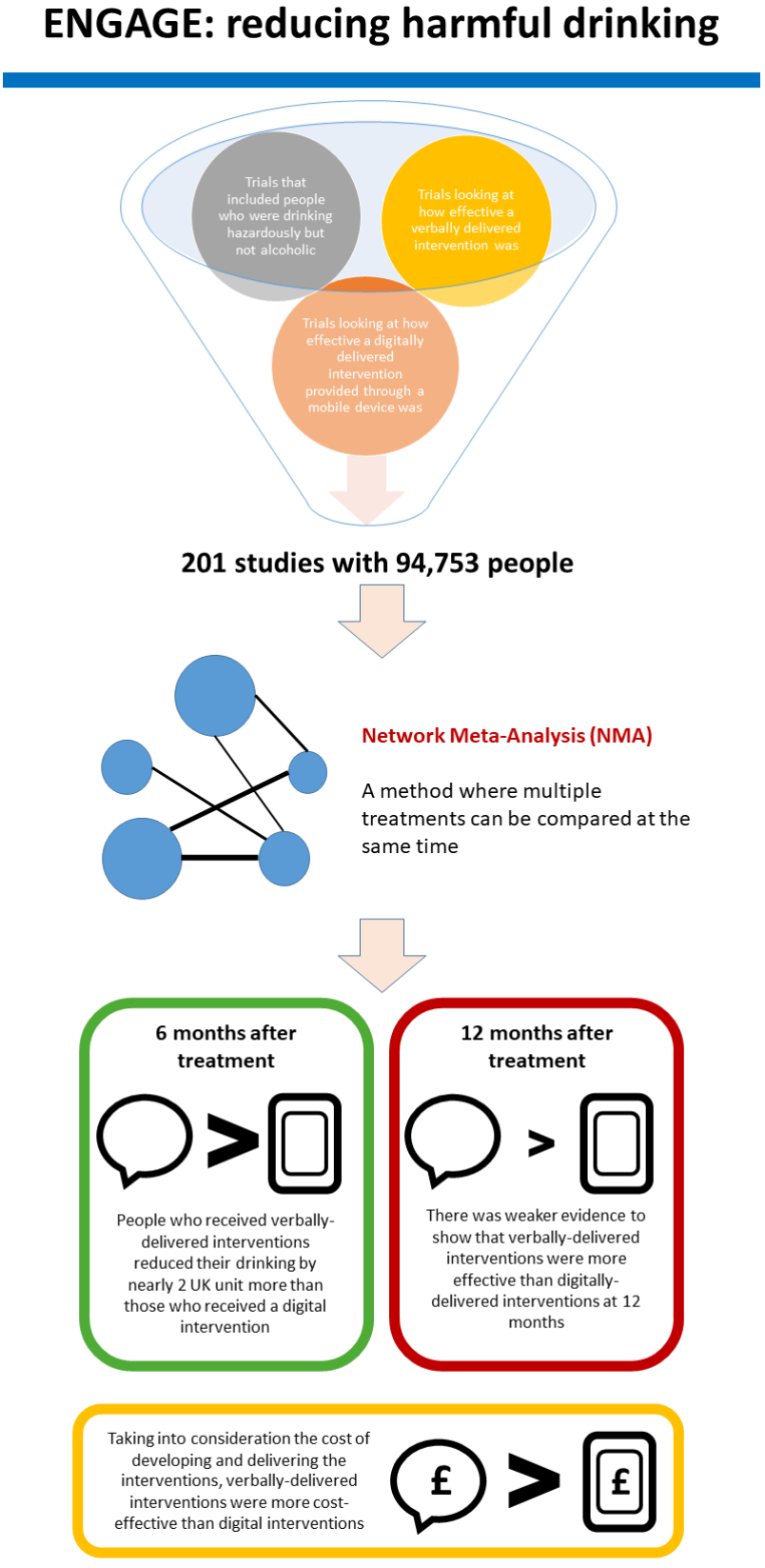
For the NMA the same issues of heterogeneity between different interventions and populations in the individual trials need to be considered as in the pairwise analysis, as well as the consistency of trials across the network. Unsurprisingly, given that the analysis included many of the same trials, heterogeneity was a concern and it fed into the CINeMA confidence in results.

I had planned to conduct a component analysis as an alternative way (compared to BCT meta-regression) of evaluating whether specific elements of interventions were associated with effectiveness. However, on reflection a large proportion of the included RCT data were already missing from the NMA due to the varied outcomes measures, and there seemed a high likelihood that even more of the available data would be lost to a component analysis because of the lack of clarity in reporting of intervention detail in trial reports. I therefore concluded that a component

analysis, although potentially interesting to explore, would contain too much uncertainty and lack robustness due to missing data – it would be difficult to understand how much bias such missing data would introduce.

3.4 Knowledge translation

The NMA project involved a multidisciplinary team that included expertise in systematic review methods, information retrieval, statistics, health economics, policy making, and experts by experience. Colleagues from Public Health England and a Director of Public Health contributed to the protocol and the interpretation of the findings. Two members of the PPI group that commented on the protocol continued to meet up with me to discuss interpretation, and they contributed to a lay visualisation of the project (Figure 6).



Created by Eugenie Johnson, Lois Neal, and Sandy Harvey

Figure 6: Lay visualisation of the NMA

3.5 Results of effectiveness reviews

Here I collate results from all the effectiveness reviews (Kaner *et al.*, 2017; Kaner *et al.*, 2018; Beyer *et al.*, 2022) and consider implications.

3.5.1 Effectiveness of alcohol interventions

Table 4 shows the results from the pairwise meta-analyses taken from both Cochrane reviews and the results from the NMA for the grams/week outcome. Five studies with only 390 participants provided a direct comparison of practitioner versus digitally delivered interventions at 12 months, and even less evidence was available at six months. The results were correspondingly imprecise and gave no useful information. However, the network meta-analysis was able to use data from many more trials to inform these estimates (52 trials and 26,777 participants at 6 months; 45 trials and 25,288 participants at 12 months). The resulting more precise estimates showed a statistically significant decrease in weekly consumption at six months and less evidence of an effect at 12 months. I carried out a sensitivity analysis at six months including only trials that reported at both 6 and 12 months, to ensure that any apparent decay was in the same participants and not lower consumption from a different group of participants by chance. The point estimate was similar although the confidence interval crossed the line of no effect (-15 grams/week [95% CI -33 to 3]). It is very likely, then, that practitioner delivered interventions are more effective than digitally delivered at 6 months, but that this difference wears off by 12 months.

Comparison	Pooled estimate (grams/week)	95% CI	I^2 (%)	k	n
6 months					
Practitioner delivered versus control (Kaner *et al.*, 2018)	-22	-32 to -12	70	21	10,313
#Digitally delivered versus control (Kaner *et al.*, 2017)	-12	-16 to -7	26	19	12,822
Practitioner versus digitally delivered from pairwise meta-analysis (Kaner *et al.*, 2017)	7	-51 to 64	-	1	113

Comparison	Pooled estimate (grams/week)	95% CI	I^2 (%)	k	n
Practitioner versus digitally delivered from network meta-analysis (Beyer *et al.*, 2022)	-14	-25 to -3	-	52	26,777
12 months					
Practitioner delivered versus control (Kaner *et al.*, 2018)	-20	-28 to -12	73	34	15,197
Digitally delivered versus control (Kaner *et al.*, 2017)	-23	-30 to -15	78	41	19,241
*Practitioner versus digitally delivered from pairwise meta-analysis (Kaner *et al.*, 2017)	1	-25 to 26	0	5	390
Practitioner versus digitally delivered from network meta-analysis (Beyer *et al.*, 2022)	-6	-24 to 12	-	45	25,288

*CI = confidence interval; k = number of studies; n = number of participants; # = from 3 to 6 months not 6 months; * = longest follow-up not 12 months*

Table 4: Effectiveness of practitioner and digitally delivered interventions on consumption at 6 and 12 months

There was little evidence of a reduction in heavy drinking episodes across all the analyses, so it seems that both practitioner and digitally delivered interventions have more impact on chronic drinking than binging. Perhaps heavy drinking episodes are more likely to be social occasions or special events, and therefore less acceptable to change than regular routine drinking.

There was little evidence to suggest that the alcohol-related information provided to control groups made a difference to people's drinking.

The levels of reduction in drinking, although statistically significant according to the confidence intervals, could be considered small in comparison to the baseline levels of some of the trial participants. The mean baseline consumption (where reported) across trials in the NMA was 216 (SD = 110) grams/week, whereas the mean reduction achieved by interventions compared to no or minimal intervention ranged from 14 to 40 grams/week at the different time points. In many fields there is an agreed clinically or minimally important difference in at least one outcome that interventions need to deliver before they are designated as useful and worth promoting. However, there is no commonly understood value for the minimally important difference for reduction in alcohol consumption. There is unlikely to be a single value: it is likely to depend on the amount of time since intervention and the characteristics of the drinker or other interested stakeholder (Grant *et al.*, 2016). The relative risk of some alcohol-associated diseases have been shown to be dose dependent so that the higher the levels of drinking the worse the disease (Zakhari and Li, 2007). If this is reversible then it could then be that even reducing by a small amount might benefit the drinker.

I subsequently worked with health economists to model the cost effectiveness of these interventions in terms of the impact on six conditions commonly affected by alcohol consumption. Chronic conditions were represented by alcoholic liver disease, hypertension, and colorectal cancer; acute conditions by intoxication, assaults, and falls. The results of the network meta-analysis were fed into a model showing the progression of each condition, and the impact of the assumed reduction in consumption was demonstrated. A cost consequence analysis (CCA) estimated the reduction in cases associated with the reduction in consumption and found that practitioner delivered interventions were associated with a greater case reduction than digitally delivered because they showed a greater effect size. A cost utility analysis (CUA) estimated the change in Quality Adjusted Life Years (QALYs) showed that although the change in QALYs per person is very small, the intervention costs are small enough to make both interventions cost effective (paper in preparation).

3.5.2 Exploring heterogeneity

Subgroup analyses

Table 5 shows the results of subgroup analyses across the two Cochrane reviews (Kaner *et al.*, 2017; Kaner *et al.*, 2018). For practitioner delivered interventions the estimated reduction in consumption was lower for women, but a test for subgroup differences showed no significant

difference between the estimates (p = 0.52). The previous version of this review showed a difference in the effectiveness between men and women (Kaner *et al.*, 2007), demonstrating how insufficient data can show a spurious effect (or lack of it).

However, these subgroup analyses also demonstrate a lack of robustness when so few trials can be included. For example, 34 studies were in the main meta-analysis for the practitioner delivered review, of which 11 provided data for men only and 7 for women only. Both of these latter estimates are considerably larger than in the main analysis (for everyone), which doesn't make sense when everyone in these trials was identified as either a man or a woman. Only one of these studies statistically compared the results for men and women (as opposed to just reporting them separately), and reported that men reduced their consumption more than women at six months but not 12 months (Richmond *et al.*, 1995).

The opposite is the case for the digitally delivered trials (lower consumption for both men and women than overall), and there are clearly not enough trials reporting by gender to produce a robust result here.

	Practitioner delivered			Digitally delivered		
	Mean g/wk (95% CI)	I^2 (%)	k	Mean g/wk (95% CI)	I^2 (%)	k
Main analysis	-20 (-28 to 12)	73	34	-23 (-30 to -15)	78	42
Men only	-42 (-65 to -20)	67	11	-9 (-32 to 14)	77	4
Women only	-30 (-59 to -2)	78	7	-10 [-22 to 2)	0	4
Adolescents/young adults	-7 (-17 to 3)	0	3	-13 (-19 to -8)	52	28
Not restricted to adolescents/ young adults	-23 (-32 to -13)	75	31	-56. (-82 to -30)	89	14
General practice	-26 (-37 to -14)	79	24		-	-
Emergency care	-10 (-18 to -2)	0	10		-	-

CI = confidence interval; g/wk = grams/week; k = number of studies

Table 5: Subgroup analyses to explore heterogeneity

In many of the subgroup analyses carried out for the review of practitioner delivered interventions, there initially appeared to be an impact according to subgroup (for example trials that restricted to adolescents and young adults compared to trials with no age limitations; interventions in general practice compared to emergency care). However, these changes were confounded with publication year because more recent trials are more likely to recruit only younger people and be set in emergency care departments. In meta-regression analyses of both age and setting that adjusted for publication date, the apparent differences disappeared, demonstrating the large impact of heterogeneity on this body of literature.

Although subgroup analysis can provide estimates for particular groups of trials, it assesses one specific subgroup at a time, and has nothing to say about the intersection of different elements, for example a population subgroup in a particular setting.

Meta-regression – BCTs

Of the possible 93 BCTs in the taxonomy, 49 (53%) were not used in the digital interventions. The mean number used was 9.2 (SD = 5.3), the median was 9 and the range 1–22. In the adjusted

model, three out of 44 BCTs that were used at least once in the digital interventions (7%) demonstrated high effectiveness compared to the lower confidence interval of the pooled overall estimate from the early version of the practitioner delivered review (-23 grams/week, Table 6).

BCT	Pooled estimate (g/wk)	95% CI	k
Behaviour substitution *'Prompt substitution of the unwanted behaviour with a wanted or neutral behaviour'*	-95	-163 to -27	4
Problem solving *'Analyse, or prompt the person to analyse, factors influencing the behaviour and generate or select strategies that include overcoming barriers and/or increasing facilitators'*	-46	-91 to 1	14
Credible source *'Present verbal or visual communication from a credible source in favour of or against the behaviour"*	-32	-61 to 4	13

CI = confidence interval; g/wk = grams per week; k = number of studies

Table 6: Association of BCTs with effectiveness according to the adjusted model

It is notable that the 'behaviour substitution' BCT has a pooled estimate much larger than the overall pooled estimate. Only four trials contributed to its analysis and three were published by the same research team; it would be more robust if more studies had contributed and an independent team could substantiate the result. However, other factors provide support for it having a larger impact than other BCTs: the adjusted R^2 was the largest for this BCT; and it was the only one whose 95% confidence interval did not come close to the line of no effect. It's possible that this larger impact of 'behaviour substitution' was offset by other BCTs that have a much smaller (or even negative) impact on the effectiveness estimate.

The adjusted model reduced the measure of heterogeneity from I^2 = 78% in the base case analysis containing all trials to 67%. This suggests that these BCTs account for a small amount of the variation in effect size but a moderate to large amount of heterogeneity remains; other BCTs must also be having an effect. The model for Control Theory BCTs produced a poor index of fit and did not improve heterogeneity compared to the primary analysis of all trials (I^2 = 78.0%, R^2 = 1.81%). A systematic review addressing the same question and published around the same time as ours used different eligibility criteria and BCT taxonomy subset and reported a different set of BCTs associated with effectiveness, demonstrating the difficulties with consistency and robustness in these reviews of complex interventions (Black, Mullan and Sharpe, 2016).

The BCT analysis in this paper used the generic 93-item taxonomy (Michie *et al.*, 2013) rather than the 42-item alcohol taxonomy previously developed by team members using the intervention manuals from trials included in an early (2007) version of the Cochrane review of practitioner-delivered interventions (Michie *et al.*, 2012). It would have seemed logical to use the latter here, given that it was designed for brief alcohol interventions. However, it has not been validated with more recent interventions, and as discussed elsewhere interventions are more varied in content and delivered in different settings compared to trials included in the 2007 version of the Cochrane review. The 93-item taxonomy was developed after the 42-item alcohol-focused taxonomy and could therefore be considered a more complete tool; indeed, the BCT 'credible source' that was shown in the analysis to be associated with intervention effectiveness does not appear in the alcohol taxonomy.

All of this relates to consumption as outcome – it does not explain whether the BCTs affected earlier outcomes on the pathway and therefore moved people along the pathway *towards* a reduction in consumption. It was also unknown from this analysis whether the BCTs could have cancelled each other out rather than having a synergistic effect to provide more than the sum of their individual effects.

Meta-regression – theory

The lack of evidence of association of theory items with intervention effectiveness chimes with other studies investigating theory use in behaviour change interventions, where theories were not consistently articulated, or studies did not consistently report how or whether they used theory to develop an intervention (Prestwich *et al.*, 2014; Tebb *et al.*, 2016). Several studies found weak or

no association between use of theory in intervention development and the effectiveness of the intervention (Prestwich *et al.*, 2014; Tebb *et al.*, 2016; Dalgetty, Miller and Dombrowski, 2019).

However, there are several potential reasons that studies may not find an association between theory use and intervention effectiveness, other than the possibility that it doesn't exist. Firstly, reporting of theory use can be poor – not only *whether* theory was used to inform interventions but *how* it was used. If reviews assume that theory is being used thoroughly to inform intervention development whereas in fact the intervention developers have done little more than mention it in the introduction to the paper, then any impact of theory use would be underestimated. My theory paper suggests that, for digital interventions at least, theory is often not used. The theories reported to be used in the trials may not be the most appropriate or be biased; many focus on individual psychological processes to the neglect of environmental and contextual influences that may influence alcohol consumption (Moore and Evans, 2017). Reviews would underestimate the effect of theoretically developing an intervention if the theory ignored many of the influences on heavy drinking. Where the reviews did not formally assess or incorporate estimation of risk of bias of included studies, a presumed lack of influence of theory may in fact be due to poor quality studies, or poor-quality theorising and construct development. Finally, if the population that was used to test the theory-developed intervention was already in the 'desired behaviour' space, there would be nowhere to go in terms of achieving the desired behaviour – no detectable impact of the intervention.

Does this mean that using theory in intervention development is misguided, and that it's better to use common sense, as Oxman and colleagues have suggested (Oxman, Fretheim and Flottorp, 2005)? 'Common sense' is to 'literature review' as 'theory' is to 'systematic review': literature reviews are undoubtedly useful to summarise an area of research, but their methods are not transparent and so it is usually unclear whether their conclusions are biased by missing data or researcher prejudice. A good quality systematic review, on the other hand, follows clear methods (including those for assessing the robustness of the conclusions) and transparent reporting guidelines that minimise bias in the results. In the same way, common sense is clearly valuable and necessary in conducting research but is often not well articulated or reasoned, and history is littered with examples where 'common sense' as understood at the time now looks decidedly non-sensible. Evidence-based theory, on the other hand, is a way to articulate transparently and objectively why and how interventions are likely to impact behaviour. One might argue that

common sense and theory are not in fact on opposite sides of the argument (Oxman, Fretheim and Flottorp, 2005); rather, common sense is 'black box' theory – we know the starting and ending points but often not how the pathway works.

In the Garnett 2018b paper we looked at 'theory' in the sense of articulating and basing intervention development on formal psychological theories which articulate a purely individual notion of behaviour change without accounting for environmental and contextual factors. However, there are at least two other ways to think about theory: theories of practice and programme theory.

Theory needs to encompass not only proposed changes at the individual level but accommodate complexity and contextual influences. Theories of practice shift the focus away from 'individual level behaviour and explanatory variables' to 'practices' which look at behaviours across groups of people (Meier, Warde and Holmes, 2018). Viewing behaviour as a social activity, where the individual is often influenced by the group as much as their own individual motivations, seems particularly sensible for hazardous alcohol consumption.

Programme theory is an intervention-specific description and visualisation of how an intervention is proposed to 'work' or influence behaviour or other key outcomes (and may be based on formal psychological theories) (Rogers, 2008; Kneale, Thomas and Harris, 2015). Visualisations of programme theory are sometimes expressed as logic models, which can be useful to guide the conduct and design of eligibility criteria in a systematic review (Kneale, Thomas and Harris, 2015). Programme theory for a complex intervention needs to incorporate contextual features such as characteristics and environment of the drinker along with proposed mechanisms of changing hazardous consumption, and to explicitly lay out the expected or potential behaviour change pathways, including the mechanism of how each part of an intervention is hypothesised to affect behaviour. It seems that until recently the field of alcohol interventions has focused on drinking as an event that can be changed with a brief interventions – hence the focus on consumption as an outcome. However, hazardous drinking is more a pathway than an event because it is dependent on so many things other than the alcohol itself – past relationship with drinking, social and environmental cues, or the person's mood. It's possible that brief interventions have an influence on the person's intentions or understanding of themselves and their consumption that moves them along the pathway without having an influence on actual consumption in the follow-up period of the trials. Complex interventions addressing behaviour change need to take into account

contextual factors which could include the drinker's own assessment of their drinking (i.e. problematic or not), relationship with a practitioner or opinion of a digital intervention, their environment, social life, and so on.

Programme theory explaining how an intervention works needs to explicitly incorporate contextual methods (Mills *et al.*, 2022) and look more broadly at all elements of context. Conversely, many alcohol trials appear to have considered the intervention in a one-dimensional and time-constrained way – focusing solely on the interaction between drinker and brief intervention provider (be that a person or a machine) in the room where it happens.

Testing further assumptions: transitivity and consistency

The transitivity assumption holds for the network meta-analysis - the eligibility criteria stated that eligible participants must be screened as hazardous or harmful but not dependent drinkers. Although participants across trials were heterogeneous in terms of some characteristics (such as age), all interventions targeted people who were drinking within a similar range (described as drinking over recommended limits or AUDIT score 8-20). Some trials, because they didn't impose a higher cut-off with respect to alcohol consumption, included a minority of participants who were tending towards dependency. I decided that the proportion of participants overall was not important enough to make a difference, and trials with more than 10% drinking at levels suggesting dependency at baseline were excluded. Additionally, this issue was dealt with through assessment of effect modifiers in terms of baseline consumption.

Effect modifiers should be the same across trials. This is not the case for this population of trials because baseline consumption of participants in these trials has reduced over the years as defined cut-offs for hazardous drinking have reduced. For the 100 trials that were included in all analyses, there was a statistically significant moderate negative correlation between baseline consumption and year of publication ($r=-0.37$, $p<0.001$), in other words older papers tended to include participants with a higher baseline consumption. To account for this, I carried out a meta-regression with baseline consumption as a covariate at each time point to see whether and by how much baseline consumption affected the pooled effect estimate. At one and six months, for every unit change in baseline consumption, the consumption effect size (grams/week) decreased for digital interventions; in other words, there was an association between heavier drinking at baseline and a bigger reduction in consumption as a result of a digital intervention after one and six months. There was no statistically significant impact of baseline consumption on grams/week

from practitioner delivered interventions at one or six months. At 12 months the opposite was the case: practitioner delivered interventions appeared to reduce consumption more for those with higher baseline consumption, but there was no apparent difference in effect size by baseline consumption for those that received digitally delivered interventions. These results are not robust because there were few trials of digitally delivered interventions reporting at 12 months and few trials of practitioner delivered interventions reporting at one month.

The statistical tests for consistency described in section 2.3.1 provided different results for the analyses at different follow up time points, which is logical given that different combinations of trials contributed data to each time point. At one and six months there was some evidence of inconsistency around parts of the network, whereas at 12 months there was little. This could be because there were more studies in the analysis at 12 months.

It is important not to base the assessment of consistency solely on p-values as the tests may be underpowered, so other methods must be used in parallel. In this case, the methods were strengthened by an a priori assessment of transitivity (section 2.3.1), ensuring that the control groups were not lumped together, the populations were eligible across the network, and the impact of baseline consumption was considered. Also, the CINeMA output for the global test of incoherence showed low concerns with a p-value of 0.759; this incorporated the maximum number of studies by using the analysis of studies with longest follow-up.

There was less inconsistency between direct and indirect evidence when the attention comparator was lumped with the alcohol containing comparator rather than being lumped with the assessment only group. This suggests that the alcohol-related material in the comparator groups provided no more effect than the placebo effect.

3.5.3 Confidence in the conclusions of the reviews

Robustness of pairwise estimates: sensitivity analyses

Sensitivity analyses for grams/week for practitioner delivered interventions restricting to studies at low risk of bias due to allocation concealment, and to studies at low risk of bias due to attrition, showed little difference in the final estimate at 12 months. It is therefore likely that this result was robust. For the digitally delivered interventions the results were less robust: sensitivity analyses omitting studies at high risk of performance bias and of attrition bias provided lower estimates than the main analysis.

GRADE assessment for the Cochrane reviews

We judged the primary outcomes to be of moderate quality according to GRADE. The biggest issues with risk of bias of individual studies were blinding and attrition bias as described in section 3.1.1.

Inconsistency was an issue for both reviews: for example, point effectiveness estimates varied by the equivalent of several drinks across the different studies. Although there was substantial heterogeneity in the main analyses (section 3.1.2), we did not downgrade for inconsistency. This was firstly because the heterogeneity was not unexplained: it was clear that there were large differences in the content and duration of interventions and control group content. Secondly, a large majority of the included studies showed that the intervention reduced consumption, so the results of the studies were consistent in this sense, and the lack of statistical significance was often due to the control arm decreasing their consumption in parallel with the intervention arm, not because of a lack of effect in the intervention arm.

Indirectness was not considered a problem for this body of studies – the eligibility criteria dictated that all participants must be drinking at hazardous or harmful levels, and all the interventions aimed to reduce alcohol consumption for people living freely and functioning in the community. Imprecision is considered a problem if a clinical decision would differ depending on whether one considered the upper or lower boundary of the CI, and whether there are caveats due to cost of or adverse events from the intervention, in which case there would have to be a higher benefit to decide to use the intervention. Publication bias was not considered to be a problem here because of the comprehensive search strategy that covered published and unpublished literature from various bibliographic and grey sources with no language restriction.

CINeMA assessment for the NMA

For the within-study risk of bias, the individual study risk of bias assessments were summarised using a weighted average score pertaining to the percentage contribution of studies at each level of bias; this delivered a judgement of 'some concerns' , downgraded to 'major concerns' where more than 50% of the studies in a comparison were at high risk of bias. We also assigned a judgement of 'some concerns' for reporting bias because in many of the trials the results were not reported in a format that could be included in the analysis. Indirectness was not considered an

issue because all participants were heavy but not dependent drinkers and most of the trial results pointed in the same direction – to lower consumption for intervention participants.

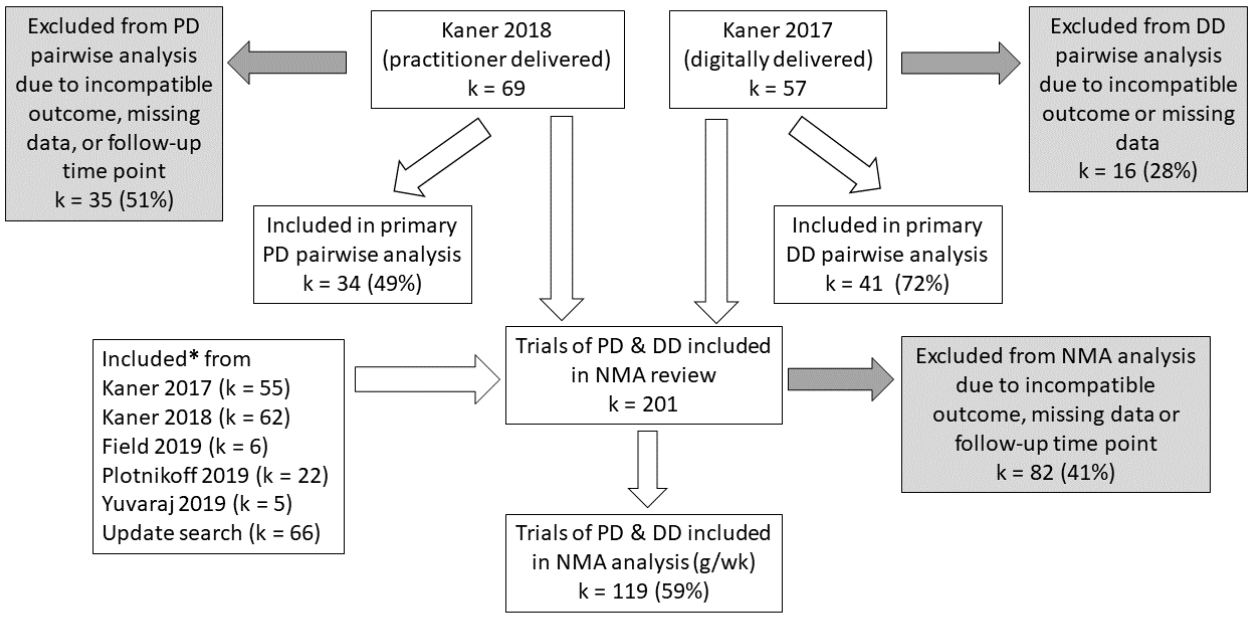
The global test of incoherence between direct and indirect evidence in the network for both outcomes, based on a random-effects design-by-treatment interaction model in CINeMA, showed low concerns in this domain (grams/week $\chi 2$ statistic: 1.174 (3 degrees of freedom), p-value: 0.759; binge episodes $\chi 2$ statistic: 0.058 (2 degrees of freedom), p-value: 0.972).

3.6 Meta-analysis for estimating effectiveness of complex interventions – summary and reflection

In the work discussed so far, I used robust Cochrane methods to report that both practitioner and digitally delivered alcohol interventions reduced hazardous or harmful alcohol consumption more than doing nothing (Kaner *et al.*, 2017; Kaner *et al.*, 2018). It was important to demonstrate the effectiveness of the interventions compared to no intervention, but insufficient trials were available comparing the two types of intervention directly to answer the question of which of the two was most effective – surely the more relevant question for decision makers. I led a paper that provided a high-level summary of the evidence from these Cochrane reviews and distilled the main outstanding and methodological issues (Beyer, Lynch and Kaner, 2018). I went on to lead a network meta-analysis to fill the gap and provide a comparison in effectiveness between the two styles of intervention. This showed that for up to six months practitioner delivered interventions reduced hazardous consumption more than digitally delivered interventions, and that it is possible that digitally delivered interventions worked better for 'heavier' than 'lighter' hazardous drinkers (Beyer *et al.*, 2022). An analysis using behaviour change techniques suggested that substituting an alternative to drinking, adopting a problem-solving approach, and ensuring the advice was from a credible source were most likely to be associated with the impact of digitally delivered interventions (Garnett *et al.*, 2018b).

However, there are some important caveats to consider with these results. Firstly, the primary analyses from the two pairwise and network meta-analyses could only include 49%, 72% and 59% of available trials respectively, due to the different consumption outcomes reported, as revealed by the ORBITAL group (Shorter *et al.*, 2019b), and the follow-up time points (Figure 7). Inspection of the results of the trials that could not be included in the analyses showed that a large majority of them (94%) showed the same direction of effect – that is, a decrease in consumption in the

group that received the intervention - but the fact remains that much of the available data was lost to the analyses, even using recommended 'gold standard' evaluation methods.



DD = digitally delivered intervention; g/wk = grams per week; k = number of trials; NMA = network meta-analysis; PD = practitioner delivered intervention; percentages are of total trials included in reviews
** = before de-duplication*

Figure 7: Journey of trials through the Cochrane and NMA reviews

Secondly, these analyses were carried out to answer the question about whether interventions as a whole work for reducing hazardous and harmful alcohol consumption, and as such they were analysed as 'black boxes' to inform their integration (or not) into national and international guidelines and policy (Gough, Oliver and Thomas, 2012). The fact that despite this integration into various guidelines, brief interventions have had consistent problems with implementation (Wilson *et al.*, 2011; Brown *et al.*, 2016; Rosario *et al.*, 2021), speaks to the fact that an 'effective' result from multiple systematic reviews of effectiveness does not tell the whole story and more was needed to really understand how brief alcohol interventions work in the real world.

It is not enough, then, to evaluate a complex behaviour change intervention as a 'black box': trials of behaviour change interventions tend to be much more heterogeneous than pharmaceutical trials due to differences in the way interventions are administered among participants and

between trials. A range of different styles of intervention were present in both practitioner and digital modes, and the fact that the interventions were broadly targeted (e.g. many targeted all heavy drinking adults) meant that there were also a range of participant characteristics. These intervention and participant characteristics can be conceptualised as 'contextual factors'. Given the levels of reported heterogeneity and the range of the confidence intervals it seems that the mean pooled estimates could disguise a lot of difference across trials that was difficult to elucidate using standard methods of exploring heterogeneity. An analysis using BCTs highlighted some components that were potentially associated with intervention effectiveness, but these also focus on the individual and do not account for external influences (Moore and Evans, 2017). To make the best use of the interventions we need to open the 'black box' and understand some of these issues.

RCTs are not the ideal study design to explore contextual factors. The aim of a well-conducted RCT is to ensure that everything is equal between the groups except for the intervention under investigation. This makes intuitive sense as the only way to know whether it is the intervention that makes the difference between intervention and control groups. However, the unintended side effect is that RCTs 'control out' contextual factors. They tend to exclude patients that might react differently or have safety issues (such as older or frail people, pregnant women, children, or people with specific needs). They tend to view contextual factors as 'confounding factors' that disguise the 'real' impact of the intervention (Shoveller *et al.*, 2016). The effectiveness of complex behavioural interventions can be influenced much more than pharmaceutical interventions by people's attitudes and daily life. For example, the effectiveness of a drug is much less likely than the effectiveness of an alcohol intervention to be influenced by a person's mood or whether they are in a public drinking venue or at home. Methods for meta-analysis of RCTs were developed in the health field by organisations like Cochrane using pharmaceutical interventions as their model, and although they allow for the exploration of heterogeneity, they stop short of being able to fully explore complexity and contextual factors. Although there are various examples of network meta-analysis feeding into decisions (Dias *et al.*, 2011; Gallos *et al.*, 2019; Thomas *et al.*, 2021), this work suggests that more is required for complex behaviour change interventions – and in fact more recent examples do incorporate different types of data (Meader *et al.*, 2022).

To sum up part I, conventional best practice meta-analyses of RCTs provided estimates of overall effectiveness of brief alcohol interventions but could neither make use of all the available data nor

fully explore contextual factors that may influence intervention effectiveness or implementation, nuances that may influence policy decisions. Part II further explores contextual factors.

PART II: Contextual factors: beyond RCTs

Part II explores elements of complexity and contextual factors and considers other types of studies that could be incorporated to help to understand how and when interventions work.

Chapter 4. Background (Part II)

This chapter describes why it is important when studying complex behaviour change interventions to explore and highlight, rather than try to 'control out', contextual factors. It introduces elements of context that are important in alcohol intervention trials and discusses methods of incorporating these, from designing them into interventions to using different types of data to analyse them.

4.1 Why context matters

The Cochrane reviews (Kaner *et al.*, 2017; Kaner *et al.*, 2018) demonstrated that there are plenty of published trials evaluating the effectiveness of alcohol interventions for reducing hazardous consumption. The first RCTs took place in the 1980s (Kristenson *et al.*, 1983; Heather *et al.*, 1987) and trials of brief alcohol interventions to reduce alcohol consumption are still being registered in 2022 (Ondersma, 2022; Woodward, 2022). Trials often aim to evaluate the effectiveness of these complex interventions for different types of people (e.g. students, pregnant women, veterans), dealing with different circumstances (e.g. comorbid conditions), or in different settings (geographically or culturally); in other words, they try to account for contextual factors by carrying out individual trials for each element of context. This ongoing plethora of trials raises questions: why are there so many RCTs in this field over such a long timeframe? Why is it that these trials did not satisfactorily answer the question 'are alcohol interventions effective' years ago? Could or should we have stopped doing trials a lot earlier? It may be more efficient and cost-effective to use other types of data to inform these questions of context than to keep designing RCTs to address them, especially considering the deficiencies of trials of complex interventions noted in section 3.6.

4.2 Contextual factors

Context is operationalised in many ways and at different levels of detail but is not always described using the term 'context' itself. A scoping review of 17 frameworks that describe

contextual determinants for implementation outcomes found that only three provided a specific definition of context but all described various contextual determinants using terms such as 'environment' and 'setting' (Nilsen and Bernhardsson, 2019). Domains were categorised as micro (patient level – 11 frameworks included these), meso (different organisational elements, from which all frameworks contained organisational support), macro (wider environment, ten frameworks), and multiple level (including interpersonal processes, leadership, and physical environment). Shoveller noted in an examination of representation of context in population health interventions that socio-demographic profiles are often used as proxy markers for context (Shoveller *et al.*, 2016). Brief alcohol interventions have historically tended to focus on the micro, patient level but this articulates the broader elements of context that can influence the implementation of interventions, including organisational support, interpersonal issues (which could extend to the interaction with a digital intervention), and the physical environment.

4.2.1 Personal characteristics and motivations as contextual factors

Motivations for heavy drinking vary (Table 7), and can influence people's ability or willingness to respond to interventions by reducing their drinking (Bresin and Mekawi, 2021). Drinking motivations can be articulated as the intersection between internal or external motivation, and positive or negative alcohol expectancy (the drinker's expectation of a positive or negative impact of their drinking) (Cox and Klinger, 1988; Cooper, 1994).

Reason for drinking	Motivation/expectancy	Assumptions
Improve social relationships	External/positive	Either the person overestimates how much their peers drink and drinks to match (perhaps if the 'normative group' is very large, like 'students') Or their immediate peer group does drink hazardously, and they are not overestimating (more likely if the 'normative group' is their actual drinking buddies)
Reduce social rejection	External/negative	People drink to be social but from a negative, anxious perspective
Enhance positive mood	Internal/positive	People feel positively about drinking and use it for mood enhancing and sociability, or to lower inhibitions
Coping – reduce negative effects	Internal/negative	People use alcohol to cope with stress
Habit	Internal/neutral	Drinking is part of the daily routine, the time and amount someone drinks is regular and considered 'normal' behaviour

Table 7: Reasons for hazardous alcohol consumption

Some of these motivations may lead to a person being resistant to recognising their own hazardous consumption, which may influence the effectiveness of an intervention. They may perceive no problem and therefore no need to change, rendering the intervention irrelevant from their perspective. Motivational interviewing style interventions aim to work with this resistance and elicit change talk (Rollnick and Miller, 1995). Similarly, if they accept the need for change from

a health perspective but assess either that their drinking is 'worth it' for its perceived benefits, or that changing would be too difficult, any advice provided by the intervention would likely be rejected. However, in this scenario the intervention may have set them on the pathway to change by allowing them to recognise that they are risking their health; although this would not be measured as a consumption outcome in a trial, it could still be a positive step towards considering change and the link to their health. If a person recognises that their drinking is hazardous, feels positive about changing, and accepts some of the advice and tools provided in the intervention, there may still be various stumbling blocks to reducing consumption and maintaining that reduction. For example, if the person habitually drinks to cope with stress, or their social life or identity are interlinked with drinking, good intentions may flounder in the face of real life.

Not all the reasons for heavy drinking listed in table 7 are necessarily considered a 'problem' by the drinker, so it makes sense that interventions designed and presented as 'solving' a drinking 'problem' would not always hit the mark, leading to reduced acceptability of the intervention.

Other differences in characteristics that are not directly associated with drinking but have been found to influence health and responses to interventions can be expressed using the PROGRESS-Plus acronym: **P**lace of residence; **R**ace, ethnicity, or culture; **O**ccupation; **G**ender or sex; **R**eligion; **E**ducation; **S**ocioeconomic status; **S**ocial capital; other factors may include age and disability (O'Neill *et al.*, 2014). RCTs sometimes report some of these characteristics at baseline but rarely include them in their results (at least for alcohol intervention trials).

4.2.2 Intervention context

As discussed in section 2.4.2 there is a lot of heterogeneity in brief alcohol interventions. Several contextual features of interventions are relevant here. They include where the intervention is received, which may describe both the physical setting (primary care, secondary care, community, or home) and for digitally delivered interventions the type of device or software used. Who delivers the intervention may also have an impact; for example, a health care professional, counsellor, or peer may have different impact for different people, and the success of a digitally delivered intervention is likely to depend at least to an extent on how comfortable the person is with the technology. These contextual factors may impact how well the intervention is received by different people and so they should be explored for a full picture of how an intervention works.

4.3 Evidence synthesis – beyond systematic review and meta-analysis of RCTs

Returning to the Gough spectrum of evidence synthesis methods (Figure 2), on the other end of the spectrum from a positivist, aggregative approach used in part I is a constructivist, configurative approach (Gough, Oliver and Thomas, 2012). A constructivist approach allows for people to fashion different versions or perspectives of reality according to their own experiences, which consequently 'permits' more than one reaction to the intervention (i.e. more nuance than a single estimate of effect). Whereas the aggregative approach predominantly 'adds up' the data to provide a single pooled estimate, a configurative approach focuses more on 'arranging' the data to develop theory about how interventions create change (or don't). The configurative approach is used to develop theory about the impact (or otherwise) of interventions, rather than testing hypotheses about their effectiveness. It allows space, therefore, for the influence of contextual factors on the intervention. It also allows for a broader perspective that allows us to consider not only the intervention's direct influence on individuals, but how it interacts with other interventions and elements of complexity using a 'systems perspective' (McGill *et al.*, 2021; Skivington *et al.*, 2021). For example, a heavy drinker will experience conflicting influences from alcohol marketing through mass and social media, along with other public health alcohol-related interventions such as pricing and restriction of availability (Petticrew *et al.*, 2017).

The methodological driver for constructivist synthesis is about the relevance of data to the research question rather than robustness of study design, because there is no single estimate that is susceptible to bias (although this is not to say that bias is not a potential issue in constructivist synthesis). A carefully purposive approach to data collection can be taken over a comprehensive one because the focus is less on gathering all evidence to minimise bias, and the process can be more iterative and less protocol-driven (although using a protocol is always helpful to aid with planning and transparency). This provides a framework under which to incorporate varied study designs and real-world evidence into analyses about the effectiveness of interventions.

Whilst acknowledging that meta-analysis of RCTs is important to demonstrate the overall effectiveness (or not) of interventions, other types of data can be used to understand more about the participants and influence of the intervention in context.

Researchers commonly incorporate process evaluations alongside RCTs, collecting quantitative or qualitative data from those involved in a trial to understand how an intervention works and what

intervention recipients, significant others and health professionals think of them (Oakley *et al.*, 2006; Petticrew *et al.*, 2013; Moore *et al.*, 2015). However, a relatively small proportion of alcohol trials have reported an associated process evaluation. Data from process evaluations and qualitative or quantitative data unrelated to specific trials can be incorporated along with trial data into mixed methods syntheses (Hong *et al.*, 2020), to construct a much richer evaluation of how the intervention 'works'.

Real-world data is derived from multiple sources outside typical clinical research settings. It can include routine data from patient records, disease registries, and data automatically gathered as people use personal devices and apps (Sherman *et al.*, 2016; Clarke *et al.*, 2019). Increasing capability to process and analyse big data presents novel opportunities to add data from these sources to understand more than the mean effect size obtained from a meta-analysis (Schaub *et al.*, 2020; Eichler *et al.*, 2021). These data become real-world evidence when they are incorporated into an analysis plan with appropriate design and evaluation, and can add to trial evidence (de Lusignan, Crawford and Munro, 2015). Real-world evidence has been used to enhance understanding from RCTs (Franklin *et al.*, 2021; Liu *et al.*, 2022; Nguyen *et al.*, 2022) and in evidence synthesis (Jenkins *et al.*, 2014), but not so far in the alcohol field.

Real-world evidence clearly lacks the robustness of RCTs for the purposes of evaluating effectiveness under the paradigm discussed in part I. In an RCT, the aim is to ensure all is equal except the intervention, and researchers work hard to control the participant characteristics and their journey in the trial to ensure this is the case. Real-world data is collected for a different purpose than evaluation and as the name suggests is susceptible to the messiness of the real world; for example, it has no pre-set eligibility criteria for the population characteristics or their treatment. The disadvantage is that there is little or no researcher control and there may be known and unknown sources of bias and confounding which reduce the capacity to assert causality (Créquit *et al.*, 2020). However, the advantage is that the data is likely to be truly relevant to a real-world population and will not have complex cases 'controlled out' of it, so it can enhance our understanding of how the intervention is working for the whole population.

Chapter 5. Exploring contextual factors

One way to address contextual factors in alcohol trials is to devise interventions that account for them in their design. Two of my papers synthesised trials of such interventions, and this chapter critiques their methods and discusses the pros and cons of meta-analysis versus narrative synthesis in a group of heterogeneous trials.

5.1 Personal characteristics and motivations

The first paper in this section synthesised trials that evaluated interventions targeting linked unhealthy eating and hazardous drinking, including or based on elements of brief alcohol interventions (Scott *et al.*, 2018). The eligibility criteria for interventions were broader here than those in the systematic reviews in part I: they needed to target alcohol consumption, some of them included elements of brief intervention such as feedback on consumption, but some interventions were more intensive than those included in part I. This review focused on a population of young adults because this is where a relationship between unhealthy linked drinking and eating patterns often begins.

Paper 7

Scott S, **Beyer FR**, Parkinson K, Muir C, Graye A, Kaner E, et al. Non-pharmacological interventions to reduce unhealthy eating and risky drinking in young adults aged 18–25 years: a systematic review and meta-analysis. *Nutrients*. 2018;10(10):1538 https://doi.org/10.3390/nu10101538

The second paper incorporating context included trials of interventions specifically targeting hazardous drinking alongside depression (O'Donnell *et al.*, 2022). Although the direction of cause or effect between hazardous drinking and mental health is uncertain (Cobiac and Wilson, 2018), it is still valuable to explore both.

Paper 8

O'Donnell A, Schmidt CS, **Beyer FR**, Schrietter MS, Anderson P, Jané-Llopis E, et al. Effectiveness of digital interventions for people with comorbid heavy drinking and depression: A systematic review and narrative synthesis. *Journal of Affective Disorders*. 2022;298:10-23

https://doi.org/10.1016/j.jad.2021.11.039

5.1.1 Systematic review methods

Both review protocols were published on Prospero, and the protocol for the comorbid depression review was published in a journal as well (Schulte *et al.*, 2019). Publishing in a journal is valuable because of the extra layer of independent peer review that comes with it, but the delays in response and peer review can be difficult to satisfactorily respond to if the review is well underway before peer review comments arrive. Uploading to a service like Open Science Registries provides a time-stamped version of the protocol (in more detail than the Prospero record) that allows a reader of the final report to compare the plans with the final report. Both of my reviews used a comprehensive search including multiple sources and grey literature, and independent duplicate screening at both stages. Both used the Cochrane tool for evaluating the risk of bias in the included trials.

5.2 Intervention context

The aim of these two reviews (Scott *et al.*, 2018; O'Donnell *et al.*, 2022) was to eliminate some of the heterogeneity and account for some contextual factors by selecting interventions that were designed to target hazardous or harmful drinking alongside the linked unhealthy eating behaviour or co-morbid depression respectively, in the same intervention. Focusing on targeted interventions like this did not resolve the problem of heterogeneity as much as expected, because the interventions still addressed the drinking, eating and depression in different ways.

5.3 Results from reviews of interventions accounting for contextual factors

Results of both reviews were inconclusive and based on a small number of trials. Three of eight (38%) of the included studies testing drinking/eating interventions were judged to be at low risk of bias, whereas there were 'some concerns' with all five of the studies evaluating

drinking/depression interventions (some of which provided more sessions than the 'brief interventions' included in the Cochrane reviews and NMA). The interventions targeting heavy drinking and linked unhealthy eating were more likely to increase the young people's fruit and vegetable consumption than to reduce their alcohol intake. As with the previous analyses, there was no evidence of impact on heavy drinking episodes. There were several methodological issues with the drinking/eating trials. They appear not to have screened the participants for hazardous alcohol consumption or unhealthy diet at recruitment, and in at least two of the trials the participants' behaviours with respect to either or both were deemed 'healthy' or were within recommendations, leaving little space for a change in those behaviours following intervention. Another issue is that the chosen measure of average fruit and vegetable consumption over a week is irrelevant to the consumption of unhealthy food in a binge context; a person could eat a very healthy diet containing lots of fruit and vegetables through most of the week and still exhibit the linked unhealthy drinking and eating behaviours in the 24 hours around a binge episode. A more relevant outcome would be the dietary intake for the 24 hours around the heavy drinking episode.

The interventions that targeted heavy drinking and depression together influenced both control and intervention groups in most of the trials, as noted in previously discussed reviews. Interventions were multi-component and mostly included counselling techniques, sometimes alongside feedback. The control groups were varied as discussed in the previous reviews (section 2.4.3). One trial reported a decrease in both alcohol consumption and depression outcomes, but only after one month.

5.4 Synthesis: narrative and meta-analysis

The appropriateness of meta-analysis for synthesis of trial data is a subjective decision, particularly for complex interventions and in the presence of different outcome measures assessing the same construct. The benefit of using as much data as possible to contribute to an overall pooled estimate in a meta-analysis needs to be balanced against how meaningful it is to combine interventions that, although aiming to influence the same outcome, provide different mechanisms for doing so. In the comorbid depression review, meta-analysis was not used because there were only five eligible studies that reported a range of outcomes. However, a similar systematic review of digital interventions for people with co-occurring depression and heavy drinking did carry out a meta-analysis (Schouten *et al.*, 2022). Despite similar eligibility criteria to our review, the Schouten review included two studies that we had excluded because they included interventions aiming to

maintain abstinence in a population that had been previously alcohol dependent. Our review aimed to evaluate interventions in people who were non-treatment seeking heavy but not dependent drinkers, because their motivation might be different and they require a different type of treatment. In order to carry out a meta-analysis, the Schouten review used standardised mean difference which allowed them to incorporate units per drinking day, drinks per week, and mean number of alcohol use occasions per day, into the same analysis and report a Hedges *g* effect size for alcohol consumption. The difficulty with combining alcohol outcomes together is that they represent different types of drinking that may respond to different stimuli. For example, units per drinking day and mean number of drinking occasions per day are both measures of heavy episodic or binge drinking, whereas drinks per week can be a measure of consistent drinking. According to the Cochrane reviews of brief alcohol interventions (Kaner *et al.*, 2017; Kaner *et al.*, 2018), there is evidence of an impact on the latter but not the former, so the benefit of a pooled estimate obtained by combining them in a standardised mean difference is perhaps offset by the fact this conflates two different types of outcomes that respond differently to the intervention. A narrative synthesis, whilst missing the simplicity of a single estimate, can be better at capturing complexity. However, a meta-analysis can increase the power of the synthesis – as demonstrated here where the narrative synthesis reported the studies suggested a direction of effect whereas the meta-analysis pooled them together and reported a statistically significant result.

5.5 Beyond RCTs

My last two reviews (Scott *et al.*, 2018; O'Donnell *et al.*, 2022) synthesised data from RCTs that attempted to deal with complexity by designing it into the interventions themselves. However, synthesising RCTs alone, whether narratively or by meta-analysis, still misses a lot of the nuance of how interventions would work for different people in different contexts.

Why are there so many RCTs in the alcohol intervention field over such a long timeframe, and why have they not answered the question of effectiveness of alcohol interventions? My hypothesis is that alcohol trials continue to be registered in an attempt to address multiple contextual factors separately: in other words, to test specific interventions on specific types of people (Norris, 2021; Ondersma, 2022; Woodward, 2022) or settings (Man-Ping, 2021). It may be less costly and more effective to use other types of data, including routinely collected data, to provide a richer assessment of whether and how interventions work best for different people in different contexts, something that is not common in the field of alcohol interventions. Alternatively, instead of

designing a separate intervention for each circumstance perhaps a modular intervention could be designed (also using varied rich evidence) that allows someone to select the parts that will most likely help them in their own circumstances. A similar suggestion involved openly available curated 'libraries of content' (McCambridge, 2021).

Chapter 6. Discussion and conclusions

6.1 Main synthesis of findings and relation to other work

This work aimed to explore the use and limitations of evidence synthesis methods for informing decisions about complex behaviour change interventions, using brief alcohol interventions as a case study. Despite dozens of systematic reviews and over 200 RCTs conducted over the last three decades that mostly reported that these interventions are effective in reducing heavy drinking, coupled with national guidance recommending their implementation, they are not consistently administered, and new RCTs continue to be registered in 2022. This suggests that current 'gold standard' evidence synthesis methods are not enough on their own to evaluate and make decisions about whether and how to use complex behaviour change interventions; if they were then they would have made sense of the complexities and settled the questions many, many trials ago. Decisions about complex behaviour change interventions require a broader perspective and different types of data to fully understand when and how they might work.

6.1.1 Effectiveness of alcohol interventions

I published two Cochrane reviews that demonstrated that practitioner and digitally delivered interventions each reduce consumption more than no or minimal intervention (Kaner *et al.*, 2017; Kaner *et al.*, 2018). We reported moderate confidence in these results and they are in agreement with other systematic reviews that often focused on specific populations, settings or types of interventions, that showed a decrease in consumption for those receiving alcohol interventions compared to no or minimal intervention (Hennessy *et al.*, 2019; Plotnikoff *et al.*, 2019; Yuvaraj *et al.*, 2019). It is possible that behaviour substitution, problem solving, and a transparent credible source may be more associated with effectiveness than other BCTs in digitally delivered interventions, but a large amount of heterogeneity remained in the analysis. This is not definitive; other effective BCTs have been identified from a different team in a partially overlapping set of trials (Black, Mullan and Sharpe, 2016).

I published a high level summary and discussion of all the above evidence (Beyer, Lynch and Kaner, 2018), followed by the first published analysis that directly compared practitioner with digitally delivered alcohol interventions using NMA (Beyer *et al.*, 2022). This showed that practitioner delivered interventions reduced hazardous consumption more than digitally delivered interventions up to the first six months after intervention, and that by 12 months post-

intervention the results were similar. There was no evidence of an impact of these interventions on heavy episodic drinking in any review, so it appears that they influence chronic drinking more than binging behaviour. There was also no evidence to suggest that alcohol-related information given to control group participants influenced their consumption, suggesting that observed consumption effects on many control groups are more likely to be an artefact of regression to the mean and/or assessment reactivity. Reporting of outcomes was not sufficient to explore inequalities in the impact of interventions.

Finally, I published two systematic reviews that integrated contextual factors into the design of interventions that were mostly based on brief alcohol interventions: for linked unhealthy eating behaviours and co-occurring depression respectively (Scott *et al.*, 2018; O'Donnell *et al.*, 2022). A similar picture emerged of small reductions in consumption in intervention groups compared to control.

6.1.2 Methods of evaluating effectiveness of complex behaviour change interventions

The results reported in the previous section used gold standard methods for evaluating the effectiveness of alcohol interventions. Meta-analysis of RCTs is the generally accepted method for evaluating the effectiveness of interventions as recommended by leading evidence synthesis organisations (Centre for Reviews and Dissemination, 2009; Aromataris and Munn, 2020; Higgins *et al.*, 2022b). Although meta-analyses of RCTs to estimate effectiveness of interventions as a 'black box' are necessary to discard interventions that are not effective, my work has shown that they were a blunt tool that was insufficient to fully explore how complex interventions worked, because there were too many outcome measures and too many dimensions of heterogeneity (in both interventions and populations). The Cochrane reviews and meta-analyses of RCTs contributed to national recommendations to use brief alcohol interventions widely, but they could not provide the nuance required to understand or predict the challenges of implementation in the real world. Only recently have systematic reviews been published seeking to address implementation of interventions, and they incorporated mixed study designs (Chan *et al.*, 2021; Rosario *et al.*, 2021). Pairwise and network meta-analyses were neither able to use all relevant data (not even all relevant RCT data) nor to provide a complete and rounded assessment of the effects of different intervention elements on different people.

RCTs aim to smooth out complexity and so they are by definition ill-equipped to study it. More and more RCTs over time, and multiple syntheses of them, have not moved the field of alcohol interventions forward very much. Despite trials and systematic reviews published over several decades demonstrating effectiveness, RCTs of brief alcohol interventions are still being published in 2022, seemingly to test out more and more individual types of intervention for different subgroups of people in specific settings. The first iteration of the Cochrane review of practitioner delivered interventions was published in 2007, it cited previous reviews reporting effectiveness of brief interventions going back to 1993, and most reviews from then right up to 2022 report a small effect of interventions compared to no or minimal treatment.

Rather than seeking to generate more and more precise answers for different subgroups of people using trial data, it would be more efficient to use existing non-trial data to provide a deeper and more holistic understanding of how elements of interventions work (or don't) for different people, and just as importantly how to get heavy drinkers and those engaged in their care to engage with the interventions. Logic models provide one mechanism by which to draw on other types of existing data to help explain the mechanism by which an intervention works, propose how contextual factors might influence its impacts, and try to foresee unintended consequences (Kneale, Thomas and Harris, 2015). In fact, existing guidance for narrative synthesis proposes a similar approach (described at the time not as a 'logic model' but as 'developing a theory of how the intervention works, why and for whom') (Popay *et al.*, 2006). Realist review also draws on quantitative and qualitative data as appropriate to build theory about how an intervention works, capturing upstream and broader determinants of health in its explanation by explicitly seeking contextual factors and linking them to outcomes in 'context-mechanism-outcome' configurations (Pawson *et al.*, 2005; Jagosh, 2019).

Other types of evidence are required that may be less easy to synthesise but are more relevant because they do not try to cancel out complexity like RCTs do. Other work published since I began this programme of work, notably the MRC guidance framework (Skivington *et al.*, 2021), agrees that many types of data are required when evaluating complex interventions. Other work has also addressed methods for incorporating different types of data into effectiveness evaluations. For example, guidance has been published on methods of synthesising quantitative evidence to explore complex systems (Higgins *et al.*, 2019). Higgins et al suggest using logic models to describe and understand the system, and also considering different behaviours or effects than the key

intended ones as separate outcomes for meta-analysis or meta-regression, in order to understand the influence of the intervention better. Other types of studies than RCTs can provide rich detail: for example, process evaluations associated with RCTs use quantitative or qualitative methods to explain the impact of an intervention; and cohort studies are likely to be of longer duration than RCTs and so are more likely to capture adverse events and longer-term impacts. Guidance has also been published for choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions (Booth *et al.*, 2016). Interviews and focus groups provide rich data about how accessible and acceptable an intervention is for those it targets, and about how those who are advised to use an intervention feel about it, thus drawing out barriers to use and implementation. Organisations like NICE and the FDA have published guidance and plans for using real-world evidence (Klonoff, 2019; NICE, 2022b). Researchers are exploring different ways to incorporate real-world evidence, for example to inform clinical development and RCT design (Dagenais *et al.*, 2022) or to augment the control arm of an RCT (Liu *et al.*, 2022). Real-world evidence is starting to be used to enhance RCTs (Franklin *et al.*, 2021) and in evidence synthesis - although not so far in the alcohol field (Kinast, Lutz and Schreiweis, 2021; Zeng *et al.*, 2021). Alcohol consumption data is regularly collected as part of routine care and via apps so it may be a useful untapped resource of real-world data.

6.2 Strengths and limitations of the portfolio as a whole

The systematic reviews and syntheses that comprise this portfolio were based on robust and well-established methods. Despite my critique that these methods do not go far enough to fully inform decisions about the use of complex behaviour change interventions, I do not dispute that they are robust and necessary methods for initial, broad evaluation of interventions. My review protocols were published in advance, comprehensive searches were conducted to reduce bias, and screening and critical appraisal were conducted by multiple reviewers independently. Comprehensive searches were not conducted from scratch to inform the NMA analysis, which may have introduced a higher risk of missing relevant studies. However, I felt that the saving in resource from harvesting studies from recent good quality systematic reviews and searching only recent years to bring these up to date, instead of having to screen tens or hundreds of thousands of results, outweighed the risks.

My paper comparing practitioner with digitally delivered interventions directly was the first to do so with network meta-analysis, using all available data to provide a more precise estimate than

was previously available from head-to-head trials. The reviews of interventions of co-morbid depression and linked unhealthy eating were the first to draw together trials of interventions targeting these issues together.

Although my work using these robust methods did provide estimates of effectiveness of alcohol interventions, they lacked the ability to incorporate all the available data and to provide full understanding of the complexities of the interventions and of the influences on the participants. This was most obviously demonstrated by the fact that these systematic reviews contributed to decisions and recommendations to roll out alcohol interventions but failed to predict or account for difficulties in implementation and were not able to fully explore differences in participants and interventions.

The main challenges with the analyses concerned characteristics of the trials that meant they couldn't be fully incorporated. A wide range of reported outcomes that could not be combined or integrated restricted the amount of data available to the analyses. Differences in the content of interventions, and the fact that they targeted a wide range of people, meant that there was a lot of different types of heterogeneity in the trials. Meta-analysis of trials tends to try to reduce and cancel out heterogeneity but for complex interventions we rather need tools to reveal and explore it, because it could inform decisions about how, where and for whom the interventions should be used.

Another issue was reporting of data. Many trials did not report sufficient detail about the content of interventions or participants to be able to confidently extract data about BCTs, theory or characteristics of participants.

Some of the analyses only took place for digitally delivered and not practitioner delivered interventions. Sometimes this was because the work had already been completed (for example BCT analysis of practitioner delivered interventions), but otherwise it was because it was designed as part of the review work on digitally delivered interventions and resource did not stretch to duplicating in the practitioner delivered space.

6.3 Implications for policy, research, and practice

Practitioner and digitally delivered alcohol interventions on average reduced overall weekly consumption for hazardous and harmful drinkers, but there was little evidence that they impacted

heavy episodic or binge drinking. Practitioner delivered interventions reduced consumption by more than digitally delivered interventions up to the first six months post intervention, but this difference disappeared by 12 months. Both practitioner and digitally delivered interventions should continue to be recommended for people drinking at hazardous or harmful levels. Digitally delivered interventions can be more accessible and should be used in the absence of practitioner delivered interventions; although they reduce consumption by less on average immediately after delivery, they have a similar impact over the longer term. Digitally delivered interventions may work better for higher than lower baseline levels of hazardous drinking.

Future studies of alcohol interventions should follow the ORBITAL core outcome set (Shorter *et al.*, 2021) to provide standardised data that allows synthesis of studies, thus avoiding the issue of missing data from analyses. The minimally important difference in consumption, measured according to one or more of these core outcomes, should be calculated for different baseline levels of drinking. This would demonstrate whether alcohol interventions have a meaningful impact on consumption for all hazardous and harmful drinkers, or only a subset, and allow policy makers to make clearer recommendations about how to target interventions.

Evaluation of effectiveness using meta-analysis of RCTs is necessary (where possible) but not sufficient to answer questions about complex behaviour change interventions where the interventions and the characteristics of the target population differ amongst trials. For these types of interventions, 'gold standard' methods need to be further developed to enable all relevant evidence to feed into decisions about interventions. For example, MRC guidance suggests extensions to standard designs of RCTs or the use of non-randomised designs and modelling approaches, and also advocates the use of mixed methods research to enhance evaluation of complex interventions (Skivington *et al.*, 2021). Real-world evidence could be used to help understand alcohol interventions more fully in terms of the influence of the complex differences in intervention content and drinker characteristics on how they work. For example, routinely collected data from electronic health records or apps could inform this work.

There is a lot of heterogeneity in the interventions and populations in the trials included in my analyses. However, more than 200 trials and over 40 systematic reviews across 35 years have consistently reported that these interventions reduce consumption, so no further 'conventional' RCTs are required. More recent RCTs tend to focus on specific groups of people with different iterations of interventions, but conventional RCT design is unable to explore all the complexity.

The pertinent question is not ‘are alcohol interventions effective?’ but rather ‘how do we get the best out of them?’, and this involves using different types of data to enrich our understanding of what works best for whom. Future syntheses should ask questions not about overall effectiveness, but rather about how different characteristics of drinkers and elements of context influence how interventions work, and should be enhanced by qualitative and observational real-world data. This will provide a richer analysis than a single average pooled value and allow policy makers to make more nuanced recommendations about who should use interventions in what circumstances. It is preferable in the context of providing evidence about complex interventions for policy making to use or adapt methods that allow all relevant data to contribute, regardless of how messy or difficult it is, rather than omit data that cannot be incorporated into a particular method.

References

Alcohol Change UK (2018) *The alcohol change report*. Available at: https://alcoholchange.org.uk/about-us/the-alcohol-change-report (Accessed: 29 Oct 2022).

Allsop, J. (2013) 'Competing paradigms and health research: design and process', in Saks, M. and Allsop, J. (eds.) *Researching health: qualitative, quantitative and mixed methods*. London: Sage, pp. 18-41.

Altman, D.G. (1991) 'Randomisation', *BMJ*, 302(6791), pp. 1481-2.

Anderson, P., O'Donnell, A., Kaner, E., Llopis, E.J., Manthey, J. and Rehm, J. (2021) 'Impact of minimum unit pricing on alcohol purchases in Scotland and Wales: controlled interrupted time series analyses', *Lancet Public Health*, 6(8), pp. e557-e65.

Aromataris, E. and Munn, Z. (2020) *JBI manual for evidence synthesis*. Available at: https://wiki-jbi.atlassian.net/wiki/spaces/MANUAL/overview (Accessed: 31 Oct 2022).

Barnett, A.G., van der Pols, J.C. and Dobson, A.J. (2004) 'Regression to the mean: what it is and how to deal with it', *Int J Epidemiol*, 34(1), pp. 215-20.

Barry, K.L., Blow, F.C., Willenbring, M., McCormick, R., Brockmann, L.M. and Visnic, S. (2004) 'Use of alcohol screening and brief interventions in primary care settings', *Subst Abus*, 25(1), pp. 27-36.

Beeston, C., McAdams, R., Craig, N., Gordon, R., Graham, L., MacPherson, M., et al. (2016) *Monitoring and evaluating Scotland's alcohol strategy. Final report*. Edinburgh: NHS Health Scotland. Available at: https://www.drugsandalcohol.ie/25250/1/NHS_Scotland__Final%20annual%20report.pdf (Accessed: 29 Oct 2022).

Beeston, C., Robinson, M., Giles, L., Dickie, E., Ford, J., MacPherson, M., et al. (2020) 'Evaluation of minimum unit pricing of alcohol: a mixed method natural experiment in Scotland', *Int J Environ Res Public Health*, 17(10), p. 3394.

Bellis, M.A., Hughes, K., Nicholls, J., Sheron, N., Gilmore, I. and Jones, L. (2016) 'The alcohol harm paradox: using a national survey to explore how alcohol may disproportionately impact health in deprived individuals', *BMC Public Health*, 16(1), pp. 1-10.

Berglund, M., Thelander, S., Salaspuro, M., Franck, J., Andréasson, S. and Öjehagen, A. (2003) 'Treatment of alcohol abuse: an evidence-based review', *Alcohol Clin Exp Res*, 27(10), pp. 1645-56.

Beyer, F., Campbell, F., Bertholet, N., Daeppen, J., Saunders, J., Pienaar, E., et al. (2019) 'The Cochrane 2018 review on brief interventions in primary care for hazardous and harmful alcohol consumption: a distillation for clinicians and policy makers', *Alcohol Alcohol*, 54(4), pp. 417-27.

Beyer, F., Lynch, E. and Kaner, E. (2018) 'Brief interventions in primary care: an evidence overview of practitioner and digital intervention programmes', *Curr Addict Rep*, 5(2), pp. 265-73.

Beyer, F.R. (2017) *Does personalised advice via computer or mobile devices reduce heavy drinking?* Available at: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011479.pub2/related-content/podcast/54003/ (Accessed: 10 Nov 2022).

Beyer, F.R., Kenny, R.P.W., Johnson, E., Caldwell, D.M., Garnett, C., Rice, S., et al. (2022) 'Practitioner and digitally delivered interventions for reducing hazardous and harmful alcohol consumption in people not seeking alcohol treatment: a systematic review and network meta-analysis', *Addiction*.

Black, N., Mullan, B. and Sharpe, L. (2016) 'Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory', *Health Psychol Rev*, pp. 1-33.

Boden, J.M., Fergusson, D.M. and Horwood, L.J. (2012) 'Alcohol misuse and violent behavior: Findings from a 30-year longitudinal study', *Drug Alcohol Depend*, 122(1), pp. 135-41.

Boden, J.M., Fergusson, D.M. and Horwood, L.J. (2013) 'Alcohol misuse and relationship breakdown: findings from a longitudinal birth cohort', *Drug Alcohol Depend*, 133(1), pp. 115-20.

Booth, A. (2004) 'Formulating answerable questions', in Booth, A. and Brice, A. (eds.) *Evidence-based Practice for Information Professionals: A Handbook*. Facet, pp. 61-70.

Booth, A., Noyes, J., Flemming, K., Gerhardus, A., Wahlster, P., van der Wilt, G.J., et al. (2016) *Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of*

complex interventions. Bremen: INTEGRATE-HTA. Available at: http://www.integrate-hta.eu/downloads/.
Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009a) 'Chapter 11: Fixed-effect model', in *Introduction to meta-analysis*. Hoboken NJ: John Wiley & Sons, pp. 63-7.
Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009b) 'Chapter 12: Random-effects model', in *Introduction to meta-analysis*. Hoboken NJ: John Wiley & Sons, pp. 69-75.
Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009c) 'Chapter 13: Fixed-effect versus random-effects models', in *Introduction to meta-analysis*. Hoboken NJ: John Wiley & Sons, pp. 77-86.
Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009d) 'Chapter 19: Subgroup analyses', in *Introduction to meta-analysis*. Hoboken NJ: John Wiley & Sons, pp. 149-86.
Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009e) 'Chapter 20: Meta-regression', in *Introduction to meta-analysis*. Hoboken NJ: John Wiley & Sons, pp. 187-203.
Bray, N., Kolehmainen, N., McAnuff, J., Tanner, L., Tuersley, L., Beyer, F.R., et al. (2020) 'Powered mobility interventions for very young children with mobility limitations to aid participation and positive development: the EMPoWER evidence synthesis', *Health Technol Assess*, 24(50), pp. 1-194.
Bresin, K. and Mekawi, Y. (2021) 'The "Why" of drinking matters: a meta-analysis of the association between drinking motives and drinking outcomes', *Alcohol Clin Exp Res*, 45(1), pp. 38-50.
Brignardello-Petersen, R., Bonner, A., Alexander, P.E., Siemieniuk, R.A., Furukawa, T.A., Rochwerg, B., et al. (2017) 'Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis', *J Clin Epidemiol*, 93, pp. 36-44.
Brown, J., West, R., Angus, C., Beard, E., Brennan, A., Drummond, C., et al. (2016) 'Comparison of brief interventions in primary care on smoking and excessive alcohol consumption: a population survey in England', *Br J Gen Pract*, 66(642), pp. e1-e9.
Caldwell, D.M., Ades, A.E. and Higgins, J.P. (2005) 'Simultaneous comparison of multiple treatments: combining direct and indirect evidence', *BMJ*, 331(7521), pp. 897-900.
Campbell, M., McKenzie, J.E., Sowden, A., Katikireddi, S.V., Brennan, S.E., Ellis, S., et al. (2020) 'Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline', *BMJ*, 368.
Centre for Reviews and Dissemination (2009) *Systematic Reviews: CRD's guidance for undertaking systematic reviews in health care*. York: University of York.
Chaimani, A., Caldwell, D.M., Li, T., Higgins, J.P.T. and Salanti, G. (2022) 'Chapter 11: Undertaking network meta-analyses', in Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A. (eds.) *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.3 (updated February 2022).
Chalmers, I., Bracken, M.B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A.M., et al. (2014) 'How to increase value and reduce waste when research priorities are set', *Lancet*, 383(9912), pp. 156-65.
Chan, P.S.-F., Fang, Y., Wong, M.C.-S., Huang, J., Wang, Z. and Yeoh, E.K. (2021) 'Using Consolidated Framework for Implementation Research to investigate facilitators and barriers of implementing alcohol screening and brief intervention among primary care health professionals: a systematic review', *Implement Sci*, 16(1), p. 99.
Chapman, E., Haby, M.M., Toma, T.S., de Bortoli, M.C., Illanes, E., Oliveros, M.J., et al. (2020) 'Knowledge translation strategies for dissemination with a focus on healthcare recipients: an overview of systematic reviews', *Implement Sci*, 15(1), p. 14.
Clarke, G.M., Conti, S., Wolters, A.T. and Steventon, A. (2019) 'Evaluating the impact of healthcare interventions using routine data', *BMJ*, 365.
Cobiac, L. and Wilson, N. (2018) *Alcohol and mental health: a review of evidence, with a particular focus on New Zealand*. Wellington, NZ: University of Otago.
Colom, J., Scafato, E., Segura, L., Gandin, C. and Struzzo, P. (2014) 'Brief interventions implementation on alcohol from the European health systems perspective', *Front Psychiatry*, 5(161).
Cooper, M.L. (1994) 'Motivations for alcohol use among adolescents: Development and validation of a four-factor model', *Psychol Assess*, 6(2), pp. 117-28.
Cox, W.M. and Klinger, E. (1988) 'A motivational model of alcohol use', *J Abnorm Psychol*, 97(2), p. 168.

Créquit, P., Boutron, I., Meerpohl, J., Williams, H.C., Craig, J. and Ravaud, P. (2020) 'Future of evidence ecosystem series: 2. current opportunities and need for better tools and methods', *J Clin Epidemiol*, 123, pp. 143-52.

Dagenais, S., Russo, L., Madsen, A., Webster, J. and Becnel, L. (2022) 'Use of real-world evidence to drive drug development strategy and inform clinical trial design', *Clin Pharmacol Ther*, 111(1), pp. 77-89.

Dalgetty, R., Miller, C.B. and Dombrowski, S.U. (2019) 'Examining the theory-effectiveness hypothesis: A systematic review of systematic reviews', *Br J Health Psychol*, 24(2), pp. 334-56.

de Lusignan, S., Crawford, L. and Munro, N. (2015) 'Creating and using real-world evidence to answer questions about clinical effectiveness', *BMJ Health Care Inform*, 22(3), pp. 368-73.

Deeks, J.J., Higgins, J.P.T. and Altman, D.G. (2022) 'Chapter 10: Analysing data and undertaking meta-analyses', in Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A. (eds.) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.3 (updated February 2022).

DerSimonian, R. and Laird, N. (1986) 'Meta-analysis in clinical trials', *Control Clin Trials*, 7(3), pp. 177-88.

Dias, S., Welton, N.J., Sutton, A.J. and Ades, A. (2011) *NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials*. National Institute for Health and Clinical Excellence. Available at: https://research-information.bris.ac.uk/ws/portalfiles/portal/7215331/TSD2_General_meta_analysis.final.08.05.12.pdf (Accessed: 29 Oct 2022).

Efthimiou, O., Debray, T.P.A., van Valkenhoef, G., Trelle, S., Panayidou, K., Moons, K.G.M., et al. (2016) 'GetReal in network meta-analysis: a review of the methodology', *Res Synth Methods*, 7(3), pp. 236-63.

Egger, M., Davey-Smith, G. and Phillips, A.N. (1997) 'Meta-analysis: principles and procedures', *BMJ*, 315(7121), pp. 1533-7.

Eichler, H.-G., Pignatti, F., Schwarzer-Daum, B., Hidalgo-Simon, A., Eichler, I., Arlett, P., et al. (2021) 'Randomized controlled trials versus real world evidence: neither magic nor myth', *Clin Pharmacol Ther*, 109(5), pp. 1212-8.

Evidence Synthesis International (2021) *What is evidence synthesis?* Available at: https://evidencesynthesis.org/what-is-evidence-synthesis/.

Field, M., Campbell, F., Hock, E. and Wong, R. (2019) *Digital interventions to reduce alcohol-related harm: a rapid horizon scanning review*. Alcohol Change UK. Available at: https://www.drugsandalcohol.ie/30801/1/Field-Campbell-Hock-Wong-Digital-Interventions.pdf (Accessed: 29 Oct 2022).

Fowler, L.A., Holt, S.L. and Joshi, D. (2016) 'Mobile technology-based interventions for adult users of alcohol: a systematic review of the literature', *Addict Behav*, 62, pp. 25-34.

Franklin, J.M., Patorno, E., Desai, R.J., Glynn, R.J., Martin, D., Quinto, K., et al. (2021) 'Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative', *Circulation*, 143(10), pp. 1002-13.

Gallos, I., Williams, H., Price, M., Pickering, K., Merriel, A., Tobias, A., et al. (2019) 'Uterotonic drugs to prevent postpartum haemorrhage: a network meta-analysis', *Health Technol Assess*, 23(9).

Garnett, C., Crane, D., Brown, J., Kaner, E., Beyer, F., Muirhead, C., et al. (2018a) 'Reported theory use by digital interventions for hazardous and harmful alcohol consumption, and association with effectiveness: meta-regression', *J Med Internet Res*, 20(2).

Garnett, C.V., Crane, D., Brown, J., Kaner, E.F.S., Beyer, F.R., Muirhead, C.R., et al. (2018b) 'Behavior Change Techniques used in digital behavior change interventions to reduce excessive alcohol consumption: a meta-regression', *Ann Behav Med*, 52(6), pp. 530-43.

Gough, D., Oliver, S. and Thomas, J. (2012) *An introduction to systematic reviews*. London: Sage.

The Government's Alcohol Strategy (2012). The Stationery Office. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/224075/alcohol-strategy.pdf (Accessed: 29 Oct 2022).

Grant, S., Pedersen, E.R., Osilla, K.C., Kulesza, M. and D'Amico, E.J. (2016) 'It is time to develop appropriate tools for assessing minimal clinically important differences, performance bias and quality of evidence in reviews of behavioral interventions', *Addiction*, 111(9), pp. 1533-5.

Grégoire, G., Derderian, F. and Le Lorier, J. (1995) 'Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias?', *J Clin Epidemiol*, 48(1), pp. 159-63.

Griswold, M.G., Fullman, N., Hawley, C., Arian, N., Zimsen, S.R., Tymeson, H.D., et al. (2018) 'Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016', *Lancet*, 392(10152), pp. 1015-35.

Guillemont, J., Cogordan, C., Nalpas, B., Viêt, N.-T., Richard, J.-B. and Arwidson, P. (2017) 'Effectiveness of a web-based intervention to reduce alcohol consumption among French hazardous drinkers: a randomized controlled trial', *Health Educ Res*, 32(4), pp. 332-42.

Guyatt, G., Oxman, A.D., Akl, E.A., Kunz, R., Vist, G., Brozek, J., et al. (2011) 'GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables', *J Clin Epidemiol*, 64(4), pp. 383-94.

Haynes, R.B. (2006) 'Forming research questions', *J Clin Epidemiol*, 59(9), pp. 881-6.

Heather, N. (2014) 'Interpreting null findings from trials of alcohol brief interventions', *Front Psychiatry*, 5, p. 85.

Heather, N., Campion, P.D., Neville, R.G. and Maccabe, D. (1987) 'Evaluation of a controlled drinking minimal intervention for problem drinkers in general practice (the DRAMS scheme)', *Br J Gen Pract*, 37, pp. 358-63.

Hennessy, E.A., Tanner-Smith, E.E., Mavridis, D. and Grant, S.P. (2019) 'Comparative effectiveness of brief alcohol interventions for college students: results from a network meta-analysis', *Prev Sci*, 20(5), pp. 715-40.

Higgins, J.P., Jackson, D., Barrett, J.K., Lu, G., Ades, A.E. and White, I.R. (2012) 'Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies', *Res Synth Methods*, 3(2), pp. 98-110.

Higgins, J.P.T., Altman, D.G. and Sterne, J.A.C. (2011) 'Chapter 8: Assessing risk of bias in included studies', in Higgins, J.P.T. and Green, S. (eds.) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011).

Higgins, J.P.T. and Li, T. (2022) 'Exploring heterogeneity', in Egger, M., Higgins, J.P.T. and Smith, G.D. (eds.) *Systematic Reviews in Health Research*. Hoboken, NJ: Wiley, pp. 185-203.

Higgins, J.P.T., López-López, J.A., Becker, B.J., Davies, S.R., Dawson, S., Grimshaw, J.M., et al. (2019) 'Synthesising quantitative evidence in systematic reviews of complex health interventions', *BMJ Glob Health*, 4(Suppl 1), p. e000858.

Higgins, J.P.T., Savović, J., Page, M.J., Elbers, R.G. and Sterne, J.A.C. (2022a) 'Chapter 8: Assessing risk of bias in a randomized trial', in Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A. (eds.) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.3 (updated February 2022).

Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., et al. (2022b) *Cochrane handbook for systematic reviews of interventions, version 6.3 (updated February 2022)*. Cochrane.

Hill, K.M., Foxcroft, D.R. and Pilling, M. (2018) '"Everything is telling you to drink": understanding the functional significance of alcogenic environments for young adult drinkers', *Addict Res Theory*, 26(6), pp. 457-64.

Hoffmann, S. and Kiefer, F. (2020) *Screening, diagnosis and treatment of alcohol-related disorders*. Available at: https://www.dgppn.de/_Resources/Persistent/5e704820632c4f4ec9c635b61e2d9e483917ef1c/S3-Leitlinienreport%20Alkohol_AWMF%2076-001.pdf (Accessed: 29 Oct 2022).

Hong, Q.N., Pluye, P., Bujold, M. and Wassef, M. (2017) 'Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence', *Syst Rev*, 6(1), p. 61.

Hong, Q.N., Rees, R., Sutcliffe, K. and Thomas, J. (2020) 'Variations of mixed methods reviews approaches: A case study', *Res Synth Methods*, 11(6), pp. 795-811.

Huckle, T., Huakau, J., Sweetsur, P., Huisman, O. and Casswell, S. (2008) 'Density of alcohol outlets and teenage drinking: living in an alcogenic environment is associated with higher consumption in a metropolitan setting', *Addiction*, 103(10), pp. 1614-21.

Hutton, B., Salanti, G., Caldwell, D.M., Chaimani, A., Schmid, C.H., Cameron, C., et al. (2015) 'The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations', *Ann Intern Med*, 162(11), pp. 777-84.

Jagosh, J. (2019) 'Realist synthesis for public health: building an ontologically deep understanding of how programs work, for whom, and in which contexts', *Annual Review of Public Health*, 40, pp. 361-72.

Jansen, J.P. and Naci, H. (2013) 'Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers', *BMC Med*, 11(1), p. 159.

Jenkins, D., Czachorowski, M., Bujkiewicz, S., Dequen, P., Jonsson, P. and Abrams, K.R. (2014) 'Evaluation of methods for the inclusion of real world evidence in network meta-analysis - a case study in multiple sclerosis', *Value Health*, 17(7), p. A576.

Joseph, J. and Basu, D. (2017) 'Efficacy of brief interventions in reducing hazardous or harmful alcohol use in middle-income countries: systematic review of randomized controlled trials', *Alcohol Alcohol*, 52(1), pp. 56-64.

Kaiser Permanente (2020) *Unhealthy drinking in adults screening and intervention guideline* Available at: https://wa.kaiserpermanente.org/static/pdf/public/guidelines/alcohol-adult.pdf (Accessed: 29 Oct 2022).

Kaner, E.F.S., Beyer Fiona, R., Brown, J., Crane, D., Garnett, C., Hickman, M., et al. (2015) 'Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations (Protocol)', *Cochrane Database Syst Rev*, 2.

Kaner, E.F.S., Beyer Fiona, R., Garnett, C., Crane, D., Brown, J., Muirhead, C., et al. (2017) 'Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations', *Cochrane Database Syst Rev*, 9.

Kaner, E.F.S., Beyer, F.R., Muirhead, C.R., Campbell, F., Pienaar, E.D., Bertholet, N., et al. (2018) 'Effectiveness of brief alcohol interventions in primary care populations', *Cochrane Database Syst Rev*, 2.

Kaner, E.F.S., Campbell, F., Pienaar Elizabeth, D., Heather, N., Schlesinger, C. and Saunders John, B. (2004) 'Brief interventions for excessive drinkers in primary care health settings', *Cochrane Database Syst Rev*, 2.

Kaner, E.F.S., Dickinson, H.O., Beyer, F.R., Campbell, F., Schlesinger, C., Heather, N., et al. (2007) 'Effectiveness of brief alcohol interventions in primary care populations', *Cochrane Database Syst Rev*, 1.

Keurhorst, M., van de Glind, I., do Amaral-Sabadini, M.B., Anderson, P., Kaner, E., Newbury-Birch, D., et al. (2015) 'Implementation strategies to enhance management of heavy alcohol consumption in primary health care: a meta-analysis', *Addiction*, 110(12), pp. 1877-900.

Khadjesari, Z., Freemantle, N., Linke, S., Hunter, R. and Murray, E. (2014) 'Health on the web: randomised controlled trial of online screening and brief alcohol intervention delivered in a workplace setting', *PLoS One*, 9(11), p. e112553.

Khan, S. (2020) 'Meta-analysis of weighted mean difference', in *Meta-Analysis*. Singapore: Springer, pp. 195-216.

Kinast, B., Lutz, M. and Schreiweis, B. (2021) 'Telemonitoring of real-world health data in cardiology: a systematic review', *Int J Environ Res Public Health*, 18(17).

Klonoff, D.C. (2019) 'The expanding role of real-world evidence trials in health care decision making', *J Diabetes Sci Technol*, 14(1), pp. 174-9.

Kneale, D., Thomas, J. and Harris, K. (2015) 'Developing and optimising the use of logic models in systematic reviews: exploring practice and good practice in the use of programme theory in reviews', *PLoS One*, 10(11), p. e0142187.

Kreitman, N. (1986) 'Alcohol consumption and the preventive paradox', *Br J Addict*, 81(3), pp. 353-63.

Kristenson, H., Öhlin, H., Hultén-Nosslin, M.-B., Trell, E. and Hood, B. (1983) 'Identification and intervention of heavy drinking in middle-aged men: results and follow-up of 24–60 months of long-term study with randomized controls', *Alcohol Clin Exp Res*, 7(2), pp. 203-9.

Kypri, K., Langley, J.D., Saunders, J.B. and Cashell-Smith, M.L. (2007) 'Assessment may conceal therapeutic benefit: Findings from a randomized controlled trial for hazardous drinking', *Addiction*, 102, pp. 62-70.

Kypri, K., Wilson, A., Attia, J., Sheeran, P., Miller, P. and McCambridge, J. (2016) 'Social desirability bias in the reporting of alcohol consumption: a randomized trial', *J Stud Alcohol Drugs*, 77(3), pp. 526-31.

Lapham, G.T., Hawkins, E.J., Chavez, L.J., Achtmeyer, C.E., Williams, E.C., Thomas, R.M., et al. (2012) 'Feedback from recently returned veterans on an anonymous web-based brief alcohol intervention', *Addict Sci Clin Pract*, 7(1), p. 17.

Lefebvre, C., Glanville, J., Briscoe, S., Featherstone, R., Littlewood, A., Marshall, C., et al. (2022) 'Chapter 4: Searching for and selecting studies', in Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A. (eds.) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.3 (updated February 2022).

Liu, Y., Lu, B., Foster, R., Zhang, Y., Zhong, Z.J., Chen, M.-H., et al. (2022) 'Matching design for augmenting the control arm of a randomized controlled trial using real-world data', *J Biopharm Stat*, 32(1), pp. 124-40.

Man-Ping, W. (2021) *Alcohol brief intervention integrated with mobile chat-based support for risky drinkers in Emergency Departments*. Available at: https://www.clinicaltrials.gov/ct2/show/NCT05018624 (Accessed: 10 Nov 2022).

Mastroleo, N.R. (2018) 'Brief interventions to reduce college student drinking', in *Brief Interventions for Adolescent Alcohol and Substance Abuse*. New York, NY: The Guilford Press, pp. 405-25.

McCambridge, J. (2021) 'Reimagining brief interventions for alcohol: towards a paradigm fit for the twenty first century?', *Addict Sci Clin Pract*, 16(1), p. 41.

McCambridge, J. and Kypri, K. (2009) 'Does simply asking questions change behavior? A systematic review of brief alcohol intervention trial data', *Alcohol Clin Exp Res*, 33, p. 275A.

McCambridge, J., Kypri, K. and McElduff, P. (2014) 'Regression to the mean and alcohol consumption: A cohort study exploring implications for the interpretation of change in control groups in brief intervention trials', *Drug Alcohol Depend*, 135, pp. 156-9.

McGill, E., Er, V., Penney, T., Egan, M., White, M., Meier, P., et al. (2021) 'Evaluation of public health interventions from a complex systems perspective: a research methods review', *Soc Sci Med*, 272, p. 113697.

McGowan, J., Sampson, M., Salzwedel, D.M., Cogo, E., Foerster, V. and Lefebvre, C. 'PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement', *J Clin Epidemiol*, 75, pp. 40-6.

McKenzie, J.E., Brennan, S.E., Ryan, R.E., Thomson, H.J., Johnston, R.V. and Thomas, J. (2022) 'Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis', in Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A. (eds.) *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.3 (updated February 2022).

Meader, N., Melton, H., Evans, C., Wright, K., Shiers, D., Ratschen, E., et al. (2022) 'Multiple versus single risk behaviour interventions for people with severe mental illness: a network meta-analysis and qualitative synthesis', *Health Soc Care Deliv Res*, 10(6).

Meier, P.S., Warde, A. and Holmes, J. (2018) 'All drinking is not equal: how a social practice theory lens could enhance public health research on alcohol and other health behaviours', *Addiction*, 113(2), pp. 206-13.

Meinert, C.L. (1986) *Clinical trials: design, conduct and analysis*. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780195035681.001.0001.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., et al. (2013) 'The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions', *Ann Behav Med*, 46(1), pp. 81-95.

Michie, S., Webb, T.L. and Sniehotta, F.F. (2010) 'The importance of making explicit links between theoretical constructs and behaviour change techniques', *Addiction*, 105(11), pp. 1897-8.

Michie, S., Whittington, C., Hamoudi, Z., Zarnani, F., Tober, G. and West, R. (2012) 'Identification of behaviour change techniques to reduce excessive alcohol consumption', *Addiction*, 107(8), pp. 1431-40.
Michie, S., Wood, C.E., Johnston, M., Abraham, C., Francis, J. and Hardeman, W. (2015) 'Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data)', *Health Technol Assess*, 19(99).
Miller, W.R. (1983) 'Motivational interviewing with problem drinkers', *Behavioural Psychotherapy*, 11(02), pp. 147-72.
Miller, W.R. and Sanchez, V.C. (1993) 'Motivating young adults for treatment and lifestyle change', in Howard, G. (ed.) *Issues in Alcohol Use and Misuse in Young Adults*. University of Notre Dame Press.
Mills, T., Shannon, R., O'Hara, J., Lawton, R. and Sheard, L. (2022) 'Development of a 'real-world' logic model through testing the feasibility of a complex healthcare intervention: the challenge of reconciling scalability and context-sensitivity', *Evaluation*, 28(1), pp. 113-31.
Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015) 'Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement', *Syst Rev*, 4(1), p. 1.
Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., et al. (2015) 'Process evaluation of complex interventions: Medical Research Council guidance', *BMJ*, 350.
Moore, G.F. and Evans, R.E. (2017) 'What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions', *SSM Popul Health*, 3, pp. 132-5.
Nguyen, H.S., Vo, N.-P., Chen, S.-Y. and Tam, K.-W. (2022) 'The optimal strategy for pertussis vaccination: a systematic review and meta-analysis of randomized control trials and real-world data', *Am J Obstet Gynecol*, 226(1), pp. 52-67.e10.
NHS (2021) *Alcohol units*. Available at: https://www.nhs.uk/live-well/alcohol-advice/calculating-alcohol-units/.
NICE (2010) *Alcohol-use disorders: prevention* (NICE Public Health Guideline PH24). London: National Institute for Health and Clinical Excellence. Available at: https://www.nice.org.uk/guidance/ph24 (Accessed: 29 Oct 2022).
NICE (2011) *Alcohol-use disorders: diagnosis, assessment and management of harmful drinking (high-risk drinking) and alcohol dependence* (NICE clinical guideline CG115). Available at: https://www.nice.org.uk/guidance/CG115 (Accessed: 29 Oct 2022).
NICE (2012) *The guidelines manual*. Available at: https://www.nice.org.uk/process/pmg6/chapter/introduction (Accessed: 29 Oct 2022).
NICE (2021) *NICE strategy 2021 to 2026: dynamic, collaborative, excellent*. Available at: http://www.nice.org.uk/strategy (Accessed: 29 Oct 2022).
NICE (2022a) *NICE health technology evaluations: the manual* (NICE process and methods guidance PMG36). Available at: https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation (Accessed: 29 Oct 2022).
NICE (2022b) *NICE real-world evidence framework*. Available at: www.nice.org.uk/corporate/ecd9 (Accessed: 29 Oct 2022).
Nikolakopoulou, A., Higgins, J.P.T., Papakonstantinou, T., Chaimani, A., Del Giovane, C., Egger, M., et al. (2020) 'CINeMA: an approach for assessing confidence in the results of a network meta-analysis', *PLoS Med*, 17(4), p. e1003082.
Nilsen, P. and Bernhardsson, S. (2019) 'Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes', *BMC Health Serv Res*, 19(1), p. 189.
Noblit, G.W. and Hare, R.D. (1988a) 'Chapter 4: Refutational synthesis', in *Meta-ethnography: synthesizing qualitative studies*. London: Sage, pp. 47-62.
Noblit, G.W. and Hare, R.D. (1988b) *Meta-ethnography: synthesizing qualitative studies*. London: Sage.

Normand, S.-L.T. (1999) 'Meta-analysis: formulating, evaluating, combining, and reporting', *Stat Med*, 18(3), pp. 321-59.
Norris, A.L. (2021) *Preventing alcohol misuse and consequences in vulnerable women*. Available at: https://clinicaltrials.gov/ct2/show/NCT05010187?term=brief+intervention&recrs=ab&cond=alcohol+misuse&draw=2&rank=7 (Accessed: 10 Nov 2022).
Noyes, J., Booth, A., Moore, G., Flemming, K., Tunçalp, Ö. and Shakibazadeh, E. (2019) 'Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods', *BMJ Glob Health*, 4(Suppl 1), p. e000893.
O'Donnell, A., Anderson, P., Newbury-Birch, D., Schulte, B., Schmidt, C., Reimer, J., et al. (2014) 'The impact of brief alcohol interventions in primary healthcare: a systematic review of reviews', *Alcohol Alcohol*, 49(1), pp. 66-78.
O'Donnell, A., Schmidt, C.S., Beyer, F.R., Schrietter, M., Anderson, P., Jané-Llopis, E., et al. (2022) 'Effectiveness of digital interventions for people with comorbid heavy drinking and depression: a systematic review and narrative synthesis', *J Affect Disord*, 298, pp. 10-23.
O'Neill, J., Tabish, H., Welch, V., Petticrew, M., Pottie, K., Clarke, M., et al. (2014) 'Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health', *J Clin Epidemiol*, 67(1), pp. 56-64.
Oakley, A., Strange, V., Bonell, C., Allen, E. and Stephenson, J. (2006) 'Process evaluation in randomised controlled trials of complex interventions', *BMJ*, 332(7538), pp. 413-6.
Office for National Statistics (2020) *Health Survey for England 2019. Main findings: alcohol consumption*. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2019/main-findings#alcohol-consumption (Accessed: 29 Oct 2022).
Ogilvie, D., Fayter, D., Petticrew, M., Sowden, A., Thomas, S., Whitehead, M., et al. (2008) 'The harvest plot: A method for synthesising evidence about the differential effects of interventions', *BMC Med Res Methodol*, 8(1), p. 8.
Oliver, K., Innvar, S., Lorenc, T., Woodman, J. and Thomas, J. (2014) 'A systematic review of barriers to and facilitators of the use of evidence by policymakers', *BMC Health Serv Res*, 14(1), p. 2.
Ondersma, S.J. (2022) *Scaling up: a multi-site trial of e-SBI for alcohol use in pregnancy (e-Health)*. Available at: https://www.clinicaltrials.gov/ct2/show/NCT04332172 (Accessed: 10 Nov 2022).
Oxman, A.D., Fretheim, A. and Flottorp, S. (2005) 'The OFF theory of research utilization', *J Clin Epidemiol*, 58(2), p. 113.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al. (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, 372.
Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K. (2005) 'Realist review - a new method of systematic review designed for complex policy interventions', *J Health Serv Res Policy*, 10(suppl_1), pp. 21-34.
Petticrew, M., Rehfuess, E., Noyes, J., Higgins, J.P.T., Mayhew, A., Pantoja, T., et al. (2013) 'Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute', *J Clin Epidemiol*, 66(11), pp. 1230-43.
Petticrew, M., Shemilt, I., Lorenc, T., Marteau, T.M., Melendez-Torres, G.J., O'Mara-Eves, A., et al. (2017) 'Alcohol advertising and public health: systems perspectives versus narrow perspectives', *J Epidemiol Community Health*, 71(3), pp. 308-12.
Piantadosi, S. (2020) *Principles and practice of clinical trials*. London: Springer Nature.
Plotnikoff, R.C., Costigan, S.A., Kennedy, S.G., Robards, S.L., Germov, J. and Wild, C. (2019) 'Efficacy of interventions targeting alcohol, drug and smoking behaviors in university and college students: A review of randomized controlled trials', *J Am Coll Health*, 67(1), pp. 68-84.
Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., et al. (2006) 'Guidance on the conduct of narrative synthesis in systematic reviews', *A product from the ESRC methods programme. Version*, 1.
Prestwich, A., Sniehotta, F.F., Whittington, C., Dombrowski, S.U., Rogers, L. and Michie, S. (2014) 'Does theory influence the effectiveness of health behavior interventions? Meta-analysis', *Health Psychol*, 33(5), pp. 465-74.

Public Health England (2016) *The public health burden of alcohol and the effectiveness and cost-effectiveness of alcohol control policies: an evidence review*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733108/alcohol_public_health_burden_evidence_review_update_2018.pdf (Accessed: 29 Oct 2022).

Public Health England (2019a) *PHE strategy 2020-25*. London: Public Health England. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/830105/PHE_Strategy__2020-25__Executive_Summary.pdf (Accessed: 06 Nov 2022).

Public Health England (2019b) *The range and magnitude of alcohol's harm to others: a report delivered to the Five Nations Health Improvement Network*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/806935/Alcohols_harms_to_others-1.pdf (Accessed: 29 Oct 2022).

Rethlefsen, M.L., Kirtley, S., Waffenschmidt, S., Ayala, A.P., Moher, D., Page, M.J., et al. (2021) 'PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews', *Syst Rev*, 10(1), p. 39.

Richardson, M., Garner, P. and Donegan, S. (2019) 'Interpretation of subgroup analyses in systematic reviews: A tutorial', *Clin Epidemiol Glob Health*, 7(2), pp. 192-8.

Richmond, R., Heather, N., Wodak, A., Kehoe, L. and Webster, I. (1995) 'Controlled evaluation of a general practice-based brief intervention for excessive drinking', *Addiction*, 90(1), pp. 119-32.

Ring, N., Jepson, R. and Ritchie, K. (2011) 'Methods of synthesizing qualitative research studies for health technology assessment', *Int J Technol Assess Health Care*, 27(04), pp. 384-90.

Rogers, P.J. (2008) 'Using programme theory to evaluate complicated and complex aspects of interventions', *Evaluation*, 14(1), pp. 29-48.

Rohwer, A., Taylor, M., Ryan, R., Garner, P. and Oliver, S. (2021) 'Enhancing public health systematic reviews with diagram visualization', *Am J Public Health*, 111(6), pp. 1029-34.

Rollnick, S. and Miller, W.R. (1995) 'What is motivational interviewing?', *Behav Cogn Psychother*, 23, pp. 325-34.

Rosario, F., Santos, M.I., Angus, K., Pas, L., Ribeiro, C. and Fitzgerald, N. (2021) 'Factors influencing the implementation of screening and brief interventions for alcohol use in primary care practices: a systematic review using the COM-B system and Theoretical Domains Framework', *Implement Sci*, 16(1), p. 6.

Royal Society and Academy of Medical Sciences (2018) *Evidence synthesis for policy: a statement of principles*. Available at: https://royalsociety.org/-/media/policy/projects/evidence-synthesis/evidence-synthesis-statement-principles.pdf (Accessed: 29 Oct 2022).

Salanti, G. (2012) 'Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool', *Res Synth Methods*, 3(2), pp. 80-97.

Salanti, G., Higgins, J.P., Ades, A.E. and Ioannidis, J.P. (2008) 'Evaluation of networks of randomized trials', *Stat Methods Med Res*, 17.

Saunders, J.B., Aasland, O.G., Babor, T.F., De La Fuente, J.R. and Grant, M. (1993) 'Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II', *Addiction*, 88(6), pp. 791-804.

Savic, M., Barker, S.F., Hunter, B. and Lubman, D.I. (2016) ''Holy shit, didn't realise my drinking was high risk': an analysis of the way risk is enacted through an online alcohol and drug screening intervention', *Health Risk Soc*, 17(7-8), pp. 565-91.

Schaub, M.P., Berman, A.H., Pelayo, H.L., Boumparis, N., Khadjesari, Z., Blankers, M., et al. (2020) 'e-INEBRIA special interest group roadmap for best practices for research on brief digital interventions for problematic alcohol and illicit drug use', *J Med Internet Res*, 22(8), p. e20368.

Schou, L. and Moan, I.S. (2016) 'Alcohol use–sickness absence association and the moderating role of gender and socioeconomic status: A literature review', *Drug Alcohol Rev*, 35(2), pp. 158-69.

Schouten, M.J.E., Christ, C., Dekker, J.J.M., Riper, H., Goudriaan, A.E. and Blankers, M. (2022) 'Digital interventions for people with co-occurring depression and problematic alcohol use: A systematic review and meta-analysis', *Alcohol Alcohol*, 57(1), pp. 113-24.

Schulte, B., Kaner, E.F.S., Beyer, F., Schmidt, C.S. and O'Donnell, A. (2019) 'Study protocol for a systematic review of evidence for digital interventions for comorbid excessive drinking and depression in community-dwelling populations', *BMJ Open*, 9(10), p. e031503.

Scott, S., Beyer, F., Parkinson, K., Muir, C., Graye, A., Kaner, E., et al. (2018) 'Non-pharmacological interventions to reduce unhealthy eating and risky drinking in young adults aged 18–25 years: a systematic review and meta-analysis', *Nutrients*, 10(10), p. 1538.

Senn, S. (2021) 'Intention to treat, missing data and related matters', in Senn, S. (ed.) *Statistical Issues in Drug Development*. Chichester: Wiley, pp. 203-23.

Sherman, R.E., Anderson, S.A., Dal Pan, G.J., Gray, G.W., Gross, T., Hunter, N.L., et al. (2016) 'Real-world evidence—what is it and what can it tell us?', *N Engl J Med*, 375(23), pp. 2293-7.

Shokraneh, F. and Adams, C. (2018) 'Patient and public involvement in systematic reviews: a systematic review of literature', *25th Cochrane Colloquium*. Edinburgh, Scotland. Available at: https://abstracts.cochrane.org/patient-and-public-involvement-systematic-reviews-systematic-review-literature.

Shorter, G.W., Bray, J.W., Giles, E.L., O'Donnell, A.J., Berman, A.H., Holloway, A., et al. (2019a) 'The variability of outcomes used in efficacy and effectiveness trials of alcohol brief interventions: A systematic review', *J Stud Alcohol Drugs*, 80(3), pp. 286-98.

Shorter, G.W., Bray, J.W., Heather, N., Berman, A.H., Giles, E.L., Clarke, M., et al. (2021) 'The "Outcome Reporting in Brief Intervention Trials: Alcohol" (ORBITAL) Core Outcome Set: international consensus on outcomes to measure in efficacy and effectiveness trials of alcohol brief interventions', *J Stud Alcohol Drugs*, 82(5), pp. 638-46.

Shorter, G.W., Heather, N., Bray, J.W., Berman, A.H., Giles, E.L., O'Donnell, A.J., et al. (2019b) 'Prioritization of outcomes in efficacy and effectiveness of alcohol brief intervention trials: international multi-stakeholder e-Delphi consensus study to inform a core outcome set', *J Stud Alcohol Drugs*, 80(3), pp. 299-309.

Shoveller, J., Viehbeck, S., Di Ruggiero, E., Greyson, D., Thomson, K. and Knight, R. (2016) 'A critical examination of representations of context within research on population health interventions', *Crit Public Health*, 26(5), pp. 487-500.

Skinner, H.A. and Holt, S. (1983) 'Early intervention for alcohol problems', *Br J Gen Pract*, 33(257), pp. 787-91.

Skivington, K., Matthews, L., Simpson, S.A., Craig, P., Baird, J., Blazeby, J.M., et al. (2021) 'A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance', *BMJ*, 374.

Slattery, P., Saeri, A.K. and Bragge, P. (2020) 'Research co-design in health: a rapid overview of reviews', *Health Res Policy Syst*, 18(1), p. 17.

Stern, C. and Kleijnen, J. (2020) 'Language bias in systematic reviews: you only get out what you put in', *JBI Evid Synth*, 18(9).

Sterne, J.A.C., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., et al. (2019) 'RoB 2: a revised tool for assessing risk of bias in randomised trials', *BMJ*, 366, p. l4898.

Straus, S.E., Tetroe, J. and Graham, I. (2009) 'Defining knowledge translation', *CMAJ*, 181(3-4), pp. 165-8.

Sutcliffe, K., Thomas, J., Stokes, G., Hinds, K. and Bangpan, M. (2015) 'Intervention Component Analysis (ICA): a pragmatic approach for identifying the critical features of complex interventions', *Syst Rev*, 4(1), pp. 1-13.

Tebb, K.P., Erenrich, R.K., Jasik, C.B., Berna, M.S., Lester, J.C. and Ozer, E.M. (2016) 'Use of theory in computer-based interventions to reduce alcohol use among adolescents and young adults: a systematic review', *BMC Public Health*, 16(1), pp. 1-33.

Thomas, J. and Harden, A. (2008) 'Methods for the thematic synthesis of qualitative research in systematic reviews', *BMC Med Res Methodol*, 8, p. 45.

Thomas, K.H., Dalili, M.N., López-López, J.A., Keeney, E., Phillippo, D., Munafò, M.R., et al. (2021) 'Smoking cessation medicines and e-cigarettes: a systematic review, network meta-analysis and cost-effectiveness analysis', *Health Technol Assess*, 25, p. 59.

Thompson Coon, J., Gwernan-Jones, R., Garside, R., Nunns, M., Shaw, L., Melendez-Torres, G.J., et al. (2020) 'Developing methods for the overarching synthesis of quantitative and qualitative evidence: The interweave synthesis approach', *Res Synth Methods*, 11(4), pp. 507-21.

Thompson, S.G. and Higgins, J.P.T. (2002) 'How should meta-regression analyses be undertaken and interpreted?', *Stat Med*, 21(11), pp. 1559-73.

Toye, F., Seers, K., Allcock, N., Briggs, M., Carr, E. and Barker, K. (2014) 'Meta-ethnography 25 years on: challenges and insights for synthesising a large number of qualitative studies', *BMC Med Res Methodol*, 14(1), p. 80.

UK Chief Medical Officers' low risk drinking guidelines (2016). London: Department of Health.

Vevea, J.L., Coburn, K. and Sutton, A. (2019) 'Publication bias', in Cooper, H., Hedges, L.V. and Valentine, J.C. (eds.) *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation, pp. 383-430.

Voogt, C.V., Kleinjan, M., Poelen, E.A.P., Lemmers, L. and Engels, R. (2013) 'The effectiveness of a web-based brief alcohol intervention in reducing heavy drinking among adolescents aged 15-20 years with a low educational background: a two-arm parallel group cluster randomized controlled trial', *BMC Public Health*, 13(1), pp. 1-11.

Watson, H., Godfrey, C., McFadyen, A., McArthur, K., Stevenson, M. and Holloway, A. (2015) 'Screening and brief intervention delivery in the workplace to reduce alcohol-related harm: a pilot randomized controlled trial', *Int J Nurs Stud*, 52(1), pp. 39-48.

Welton, N.J., Caldwell, D.M., Adamopoulos, E. and Vedhara, K. (2009) 'Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease', *Am J Epidemiol*, 169(9), pp. 1158-65.

White, H., Albers, B., Gaarder, M., Kornør, H., Littell, J., Marshall, Z., et al. (2020) 'Guidance for producing a Campbell evidence and gap map', *Campbell Syst Rev*, 16(4), p. e1125.

White, I.R., Barrett, J.K., Jackson, D. and Higgins, J.P.T. (2012) 'Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression', *Res Synth Methods*, 3(2), pp. 111-25.

Whitty, C.J.M. (2015) 'What makes an academic paper useful for health policy?', *BMC Med*, 13(1), p. 301.

Wilson, G.B., Lock, C.A., Heather, N., Cassidy, P., Christie, M.M. and Kaner, E.F.S. (2011) 'Intervention against excessive alcohol consumption in primary health care: a survey of GPs' attitudes and practices in England 10 years on', *Alcohol Alcohol*, 46(5), pp. 570-7.

Wilson, P.M., Petticrew, M., Calnan, M.W. and Nazareth, I. (2010) 'Disseminating research findings: what should researchers do? A systematic scoping review of conceptual frameworks', *Implement Sci*, 5(1), p. 91.

Woodward, M. (2022) *A brief intervention for alcohol users with interpersonal trauma*. Available at: https://clinicaltrials.gov/ct2/show/NCT05414344 (Accessed: 10 Nov 2022).

World Health Organization (2001) *The Alcohol Use Disorders Identification Test: guidelines for use in primary care*. Geneva: World Health Organization.

World Health Organization (2006) *Public health problems caused by harmful use of alcohol: gaining less or losing more?* WHO Regional Office for South-East Asia.

World Health Organization (2014) *NCD country profile 2014: United Kingdom of Great Britain and Northern Ireland*. Geneva: WHO. Available at: https://www.euro.who.int/en/countries/united-kingdom-of-great-britain-and-northern-ireland/data-and-statistics/ncd-country-profile-2014-united-kingdom-of-great-britain-and-northern-ireland (Accessed: 29 Oct 2022).

World Health Organization (2019) *Global status report on alcohol and health 2018*. World Health Organization.

World Health Organization (2022) *ICD-11 for Mortality and Morbidity Statistics: 6C40.2 Alcohol dependence*. Available at: https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/1580466198.

Yuvaraj, K., Eliyas, S.K., Gokul, S. and Manikandanesan, S. (2019) 'Effectiveness of workplace intervention for reducing alcohol consumption: a systematic review and meta-analysis', *Alcohol Alcohol*, 54(3), pp. 264-71.

Zakhari, S. and Li, T.-K. (2007) 'Determinants of alcohol use and abuse: Impact of quantity and frequency patterns on liver disease', *Hepatology*, 46(6), pp. 2032-9.

Zeng, C., Doherty, M., Persson, M.S.M., Yang, Z., Sarmanova, A., Zhang, Y., et al. (2021) 'Comparative efficacy and safety of acetaminophen, topical and oral non-steroidal anti-inflammatory drugs for knee osteoarthritis: evidence from a network meta-analysis of randomized controlled trials and real-world data', *Osteoarthritis Cartilage*, 29(9), pp. 1242-51.