Does moderate alcohol consumption have long-term health benefits?: A causal inference perspective

Rachel Visontay

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

The Matilda Centre for Research in Mental Health and Substance Use

Faculty of Medicine and Health

The University of Sydney

August 2023

Abstract

Heavy and disordered alcohol consumption is a known risk factor for numerous health conditions and is associated with considerable disease burden. However, at low-to-moderate levels, a substantial body of evidence suggests drinking might be associated with reduced risk for certain health outcomes. Whether these findings represent genuine protective effects, or mere methodological artefacts, remains unclear – in large part due to the difficulty of drawing causal inferences from observational research. To address this knowledge gap, this thesis aims to apply novel statistical and design-based methods that promote causal inference to the analysis of alcohol–long-term health relationships.

Study 1 is the first paper to systematically review observational studies that employ enhanced causal inference methods in assessing alcohol's relationship with any long-term health outcome. Limited evidence is found for protective effects for certain health outcomes, with the key finding being a dearth of studies employing these approaches of interest and the identification of clear gaps where these methods must be applied. **Study 2** applies a recently-developed framework for assessing consistency of associational findings – a necessary precursor to causal inference – to the alcohol–inflammation relationship. The results of 'multiverse' analyses indicate that the association of moderate drinking with reduced levels of inflammatory biomarkers is robust to data processing and analysis decisions. **Study 3** tests whether the established association between moderate drinking and lower risk for depression is a causal one by employing modern statistical approaches. Specifically, a marginal structural model is used to account for dynamic drinking behaviours and time-varying confounding in a longitudinal context. This study demonstrates that a small, protective effect for occasional and moderate alcohol consumption on depression persists even when enhanced causal inference approaches are employed. **Study 4** applies an alternative observational design to promote causal inference in assessing the relationship between alcohol consumption and type 2 diabetes, leveraging genetic instrumental variables. This is the first study to implement recently-developed Mendelian Randomisation methods for non-linear relationships in assessing the functional form of the relationship between alcohol and type 2 diabetes. It found a small protective effect for low-level drinking against risk for incident diabetes, but results were not robust to all sensitivity analyses. This study indicates that, at most, any protective effect is limited to a small benefit in women only.

The findings from this thesis advance both content knowledge relating to several complex, unanswered questions about alcohol–long-term health relationships, as well as methodological applications, laying the foundations for causal inference approaches in this field.

Thesis statement of originality

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

This thesis has not been submitted for any other degree or purposes.

Rachel Visontay

August 31, 2023

Author attribution statement

The chapters of this thesis contain four published papers and one manuscript in preparation for submission to a journal. I am the lead author on all of these publications, and the corresponding author on all but one (the protocol paper for Chapter 2; permission to include this paper in the appendices has been granted by the corresponding author). Citations are provided on page x.

Study 1 / Chapter 2 reports on content published in two journals: a protocol paper for a systematic review published in *BMJ Open*, and the results of that review published in *BMC Medical Research Methodology*. For the former, the authors are Rachel Visontay (RV), Matthew Sunderland (MS), Tim Slade (TS), Jack Wilson (JW), and Louise Mewton (LM). RV conceptualised the study with assistance from MS, TS, and LM. All authors critically revised the manuscript and approved the final version. For the latter publication, the authors are Rachel Visontay (RV), Matthew Sunderland (MS), Tim Slade (TS), Jack Wilson (JW), and Louise Mewton (LM). RV conceptualised the study with assistance from MS, TS, and LM. RV and JW completed the screening, data extraction, and quality assessment process. RV conducted the narrative synthesis and drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Study 2 / Chapter 3 is published in *Drug and Alcohol Dependence*. The authors are Rachel Visontay (RV), Louise Mewton (LM), Matthew Sunderland (MS), Steven Bell (SB), Annie Britton (AB), Bridie Osman (BO), Hayley North (HN), Nisha Mathew (NM), and Tim Slade (TS). RV conceptualised the study with assistance from LM, MS, and TS. RV conducted the analyses and drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Study 3 / Chapter 4 is published in the *American Journal of Psychiatry*. The authors are Rachel Visontay (RV), Louise Mewton (LM), Tim Slade (TS), Izzuddin M. Aris (IMA), and Matthew Sunderland (MS). RV conceptualised the study with assistance from LM, TS, and MS. RV conducted the analyses and drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Study 4 / Chapter 5 is an unpublished manuscript, currently in preparation for submission. The authors are Rachel Visontay (RV), Louise Mewton (LM), Matthew Sunderland (MS), Tim Slade (TS), Steven Bell (SB), Jorien Treur (JT), and Amy Mason (AM). RV conceptualised the study with assistance from LM, TS, MS, SB, and AM. RV conducted the analyses and drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Rachel Visontay

August 31, 2023

As supervisors for the candidate upon which this thesis is based, we can confirm that the author attribution statement above is correct.

Louise Mewton **Matthew Sunderland** **Tim Slade**

August 31, 2023

Contents

List of abbreviations

apoB	Apolipoprotein B
AMI	Acute myocardial infarction
AUD	Alcohol use disorder
BCS70	1970 British Birth Cohort Study
BMI	Body mass index
CES-D-SF	Center for Epidemiologic Studies Depression Scale–Short Form
CHD	Coronary heart disease
CI	Confidence interval
CRP	C-reactive protein
CVD	Cardiovascular disease
DAG	Directed acyclic graph
DBP	Diastolic blood pressure
DZ	Dizygotic
GABA	Gamma-aminobutyric acid
GBD	Global Burden of Disease
GRS	Genetic risk score
GWAS	Genome-wide association study
HbA1C	Haemoglobin A1c
HDL	High-density lipoprotein
HED	Heavy episodic drinking
HIV	Human immunodeficiency virus
HOMA-beta	Homeostatic model assessment for insulin resistance
HOMA-IR	Homeostasis model assessment of β-cell function
hsCRP	High-sensitivity C-reactive protein
HR	Hazard ratio
ICC	Intraclass correlation coefficient
IL-6	Interleukin-6
IPCW	Inverse probability of censoring weight
IPTW	Inverse probability of treatment weight
IV	Instrumental variable
IVW	Inverse variance-weighted
LACE	Localised average causal effect
LATE	Local average treatment effect
LCGM	Latent class growth models
LDL	Low-density lipoprotein
Lp(a)	Lipoprotein(a)
MAF	Minor allele frequency
MHD	Mental health diagnoses
MR	Mendelian Randomisation
MSD	Musculoskeletal diagnoses
MSM	Marginal structural model
MZ	Monozygotic

NLSY-79	National Longitudinal Survey of Youth 1979 cohort
NOS	Newcastle-Ottawa Scale
OR	Odds ratio
P2hBG	2-hr post-load plasma glucose
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PS	Propensity score
RCT	Randomised controlled trial
RR	Relative risk
SBP	Systolic blood pressure
SE	Standard error
SES	Socio-economic status
SNP	Single-nucleotide polymorphism
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TC	Total cholesterol
TG	Triglycerides
TIA	Transient ischaemic attack
T2D	Type 2 diabetes
VoE	Vibration of effects
WC	Waist circumference
2SLS	Two-stage least-squares

Dissemination, funding, and awards during candidature

Publications associated with this thesis

Visontay, R., Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2021). Are there non-linear relationships between alcohol consumption and long-term health? Protocol for a systematic review of observational studies employing approaches to improve causal inference. *BMJ open*, *11*(3), e043985.

Visontay, R., Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2022). Are there non-linear relationships between alcohol consumption and long-term health?: a systematic review of observational studies employing approaches to improve causal inference. *BMC medical research methodology*, *22*(1), 1-28.

Visontay, R., Mewton, L., M., Slade, T., Aris, I. M., & Sunderland, M. (2023). Moderate alcohol consumption and depression: A marginal structural model approach promoting causal inference. *American Journal of Psychiatry, 180*(3), 209-17.

Visontay, R., Mewton, L., Sunderland, M., Bell, S., Britton, A., Osman, B., North, H., Mathew, N., & Slade, T. (2023). A comprehensive evaluation of the longitudinal association between alcohol consumption and a measure of inflammation: Multiverse and vibration of effects analyses. *Drug and Alcohol Dependence, 247*, 109886.

Additional publications during candidature

Visontay, R., Mewton, L., Sunderland, M., Prior, K., & Slade, T. (2020). Changes over time in young adults' harmful alcohol consumption: A cross-temporal meta-analysis using the AUDIT. *Drug and alcohol dependence*, *214*, 108172.

Marel, C., Mills, K., **Visontay, R.**, Wilson, J., Darke, S., Ross, J., Slade, T., Haber, P. S., Haasnoot, K., Keaveny, M., Tremonti, C., & Teesson, M. (2020). Australian treatment outcome study: protocol for the 18–20-year follow-up of a prospective longitudinal cohort examining the natural history of heroin dependence and associated mortality, psychiatric and physical health, and health service use. *BMJ open*, *10*(7), e039226.

Mewton, L., Hodge, A., Gates, N., **Visontay, R.**, Lees, B., & Teesson, M. (2020). A randomised double-blind trial of cognitive training for the prevention of psychopathology in at-risk youth. *Behaviour Research and Therapy*, *132*, 103672.

Visontay, R., Rao, R. T., & Mewton, L. (2021). Alcohol use and dementia: new research directions. *Current opinion in psychiatry*, *34*(2), 165-170.

Parmenter, B., Burley, C., Stewart, C., Whife, J., Champion, K., Osman, B., Newton N., Green O., Wescott A.B., Gardner L.A., **Visontay, R.**, et al. (2022). Measurement Properties of Smartphone Approaches to Assess Physical Activity in Healthy Young People: Systematic Review. *JMIR mHealth and uHealth*, 10(10), e39085.

Thornton, L., Osman, B., Champion, K., Green, O., Wescott, A. B., Gardner, L. A., Stewart, C., **Visontay, R.**, Whife, J., Parmenter, B., Birrell, L., Bryant, Z., Chapman, C., Lubans, D., Slade, T., Torous, J., Teesson, M., & Van de Ven, P. (2022). Measurement properties of smartphone approaches to assess diet, alcohol use, and tobacco use: systematic review. *JMIR mHealth and uHealth*, *10*(2), e27337.

Wilson, J., Mills, K., Freeman, T. P., Sunderland, M., **Visontay, R.**, & Marel, C. (2022). Weeding out the truth: a systematic review and meta-analysis on the transition from cannabis use to opioid use and opioid use disorders, abuse or dependence. *Addiction*.

Wilson, J., Mills, K., Freeman, T. P., Sunderland, M., **Visontay, R**., & Marel, C. (2022). Response to Bahji et al: Limitations of the available evidence that restrict our interpretation of the transition from cannabis to opioid use. *Addiction*.

Osman, B., Teesson, M., Marx, W., Thornton, L., Jacka, F. N., Hunter, E., **Visontay, R.,** & Sunderland, M. (2023). Elevated rates of systemic inflammation among adolescents pervade geography and time: A systematic review and meta-analysis of global mean C-reactive protein levels. *OSF Preprints*.

Sunderland, M., Olsen, N., **Visontay, R.**, Chapman, C., Mewton, L., Stapinski, L., Newton, N., Teesson, M., Slade, T. (2023). 'One metric to rule them all': a common metric for symptoms of depression and generalized anxiety in adolescent samples. *Clinical Psychological Science*, 21677026231168564.

Marel, C., Wilson, J., Darke, S., Ross, J., Slade, T., Haber, P. S., Haasnoot, K., **Visontay, R.**, Keaveny, M., Tremonti, C., Mills, K., & Teesson, M. (2023). Patterns and Predictors of Heroin Use, Remission, and Psychiatric Health Among People with Heroin Dependence: Key Findings from the 18–20-Year Follow-Up of the Australian Treatment Outcome Study (ATOS). *International Journal of Mental Health and Addiction*.

Mewton, L., **Visontay, R**., Hoy, N., Lipnicki, D.M.,, Sunderland, M., Lipton, R.B., Guerchet, M., Ritchie, K., Najar, J., Scarmeas, N., Kim, K-W., Riedel Heller, S., van Boxtel, M., Jacobsen, E., Brodaty, H., Anstey, K.J., Haan, M., Scazufca, M., Lobo, E., Sachdev, P.S., and Collaborators from the Cohort Studies of Memory in an International Consortium (COSMIC). (2023) The relationship between alcohol use and dementia in adults aged over 60 years: A combined analysis of prospective, individual-participant data from 15 international studies. *Addiction, 118*(3), 412-424.

Visontay, R., Mewton, L., & Teesson, M. (2023). Tobacco, gambling, now alcohol: why Australia needs mandated alcohol warnings. *The Policymaker*. January 2023. [Non-peer reviewed publication]

Halladay, J., Slade, T., Chapman, C., Mewton, L., O'Dean, S., **Visontay, R.**, Baillie, A., Teesson, M., & Sunderland, M. (2023). Is the association between psychological distress and risky alcohol consumption shifting over time? An age-period-cohort analysis of the Australian population. *Psychiatry Research*, *326*, 115356.

Presentations arising from this thesis

Visontay, R. (2020, September 24). The relationship between alcohol consumption and long-term health: A causal inference perspective [Seminar presentation – virtual]. Matilda Centre Seminar Series.

Visontay, R., Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2021, September 3-6). Are there non-linear alcohol–health relationships?: Review of observational studies employing enhanced causal methods [Conference presentation – on demand]. World Congress of Epidemiology 2021.

Visontay, R. (2022, July 21) Can a little bit of alcohol be good for you?: Disentangling association from causation. 3 Minute Thesis, Faculty of Medicine and Health Semi-Final. 2022 Faculty of Medicine and Health HDR Conference, Sydney, Australia.

Visontay, R. (2022, July 28) Can a little bit of alcohol be good for you?: Disentangling association from causation. Rapid Talk. 2022 Matilda Centre Conference, Sydney, Australia.

Visontay, R., Mewton, L., M., Slade, T., Aris, I. M., & Sunderland, M. (2022, October 12). Moderate alcohol consumption and depression: A marginal structural model analysis of a representative US cohort [Conference presentation]. Australasian Professional Society on Alcohol & other Drugs Conference 2022, Darwin, Australia.

Visontay, R. (2023, February 7) Alcohol and health: A causal inference perspective. Rapid Talk. 2023 Matilda Centre x Medical University of South Carolina Symposium, Sydney, Australia.

Additional presentations during candidature

Visontay, R. (2021, June 21). The relationship between alcohol use and dementia: A combined analysis of prospective, individual participant data from 15 international studies [Seminar presentation – virtual]. Centre for Health Brain Ageing UNSW Seminar Series.

Mewton, L., **Visontay, R.,** Hoy, N., Lipnicki, D., Crawford, J., & Sachdev, P. (2021, November 9). The relationship between alcohol use and dementia: A combined analysis of prospective, individual participant data from 15 international studies [Conference presentation – virtual]. Australasian Professional Society on Alcohol & other Drugs Conference 2021.

Visontay, R. (2022, May 10). Can a little alcohol be good for your health? [Public talk]. Pint of Science Festival, Sydney, Australia.

Visontay, R. (2022, May 17). Fair testing: An introduction to the scientific method. [Presentation to Year 6 students – virtual]. St Fiacre's Catholic Primary School, Sydney, Australia.

Visontay, R., & O'Dean, S. (2022, September 1 & 8). Matilda Centre Introduction to R Workshop [Workshop].

Visontay, R. (2022, November 2). Can a little alcohol be good for your health? [Public talk]. Future Science Talks, Sydney, Australia.

Scholarships and funding awarded during candidature

Visontay, R. Research Training Program Scholarship. *The University of Sydney*. $35,000 p.a. [Declined].

Visontay, R. (2020-2023). Postgraduate Research Scholarship. *National Health and Medical Research Council*. $90,047.

Visontay, R. (2020-2023). NHMRC Supplementary Scholarship. *The University of Sydney*. $7,401 p.a.

Visontay, R. (2021). PREMISE Career Development Support Grant. $4,999.

Visontay, R. (2021). Matilda Centre PhD Support Scheme. $2,500.

Visontay, R. (2021). The University of Sydney Postgraduate Research Support Scheme. $527.

Visontay, R. (2021). The University of New South Wales School of Psychiatry Research Support and Professional Development Funding. $1,170.

Visontay, R. (2021). Travel Scholarship. *Dementia Centre for Research Collaboration*. $1,910.

Visontay, R. (2022). Matilda Centre PhD Support Scheme. $1,000.

Visontay, R., Bryant, Z., & Sercombe, J. (2022). Mental health and substance use data visualisation competition. Student Life Grant, *The University of Sydney*. $2,000.

Visontay, R. (2022). The University of Sydney Postgraduate Research Support Scheme. $513.

Visontay, R. (2022). Travel & Accommodation Scholarship. *Australasian Professional Society on Alcohol & other Drugs*. ~$1,400.

Visontay, R. (2023). Matilda Centre PhD Support Scheme. $1,000.

Visontay, R. (2023). Science and Industry Endowment Fund-Australian Academy of Science Fellowship [to attend the Lindau Nobel Laureate meeting 2023]. $7,500.

Visontay, R. (2023). The University of Sydney Postgraduate Research Support Scheme. $1,060.

Awards arising from this thesis

Visontay, R. (2021). Visualise Your Thesis Viewers' Choice Award. *The University of Sydney*.

Visontay, R. (2022). Student Merit Award. *Research Society on Alcoholism Conference 2022*. [Declined]

Visontay, R. (2022). People's choice best oral presentation. *Matilda Centre Conference*.

Media coverage arising from this thesis

The study reported in **Chapter 4** and published in the *American Journal of Psychiatry* (Visontay, Mewton, Slade, et al., 2023) received print, online, radio, and television coverage. Highlights include the *Sydney Morning Herald* (Australia), *The Sun* (UK), *Channel Nine News* (Australia), *2GB* (Australia), and *ABC Radio Melbourne* (Australia).

The research contained in this thesis also informed the candidate's opinion piece ('How much should I drink? Why no single study will ever answer that question') published in the *Sydney Morning Herald* (April 10, 2022).

Chapter 1

General introduction

Preface

Identifying and quantifying the health impacts of modifiable lifestyle behaviours is a principal target of epidemiologic enquiry. Knowledge derived from such research is crucial to disease prevention, alleviating disease burden, and guiding public health policy. However, this enterprise is rarely straightforward; the data we use to answer research questions are often imperfectly suited to the task, necessitating elegant methodological innovation.

Arguably, these complexities are no more evident than in the study of relationships between alcohol consumption and long-term health outcomes. The harms associated with heavy and disordered alcohol consumption (i.e., a compulsive pattern of drinking associated with significant physical and social harms) are well-established and wide-ranging (Rehm, Gmel Sr, et al., 2017). However, somewhat counter-intuitively, drinking at low-to-moderate levels is often associated with *reduced* risk for many of those same conditions, when compared with abstinence. Elucidating the nature of these relationships has proven challenging, particularly when it comes to delineating association from causation. This is largely due to research in this space being limited – for ethical and practical reasons – to observational work, from

which it was long-considered impossible to draw inference about causation, and taboo to describe with causal language (Hernán, 2018).

However, recent advances in the development of enhanced methods to promote causal inference from observational evidence, combined with the increasing availability of large, comprehensive cohort data and sophisticated computing software, have the potential to change this. Such approaches, centred on using observational data to mimic the features of randomised controlled trials (Hernán & Robins, 2020), have only recently begun to be applied in the alcohol–long-term health space. Literature on how they might best be applied, and what the results of such work can tell us about supposed protective effects of moderate drinking, is sparse.

This thesis was designed to contribute to laying the foundations for causal inference frameworks in alcohol–long-term health research, as well as to generate substantive preliminary causal evidence on the nature of some of the epidemiologic relationships in question.

In the remaining sections of this first chapter, an overview of the evidence base on alcohol consumption and long-term health is provided, including supposed 'J-shaped' relationships (1.1). Reasons for scepticism of these relationships are then explored, including the methodological limitations of existing research (1.2). Causal inference frameworks and methodologies are then introduced (1.3), followed by a discussion of the urgent need for their application to alcohol–long-term health relationships (1.4). Finally, the aims of this thesis will be detailed, and an outline provided of its chapters (1.5).

1.1 Alcohol and health; the 'J-shaped' curve

1.1.1 Prevalence, volume, and patterns of alcohol consumption

Alcohol is the most used substance globally, with 25% and 39% of females and males currently drinking, respectively (Griswold et al., 2018). It is particularly prevalent in countries high on socio-demographic indices, where 72% of females and 83% of males are current consumers and the average number of standard drinks consumed per day by current drinkers is 2.2 for women and 3.3 for men (Griswold et al., 2018). Prevalence of alcohol consumption, and patterns of use – comprised of both quantity and frequency components – vary between different age- and sex-based groups.

In Australia for example, in 2019, 82% of males and 77% of females aged over 18 were current drinkers, with prevalence highest in the 50-59 age group for both sexes (Australian Institute of Health and Welfare, 2020). Daily drinking was most common in older adults, particularly for men, among whom 17% aged over 70 reported drinking alcohol every day. The prevalence of past-month heavy episodic drinking (HED) among drinkers aged over 14 (defined here as having >=60 grams of alcohol in a sitting; 'bingeing') was 36% in the USA, 41% in the UK, and 45% in Australia, in 2016 (World Health Organization, 2018). Past-year alcohol use disorder (AUD), a chronic, diagnosable condition characterised by problematic alcohol use leading to distress and harm, is estimated to occur in 18%, 13%, and 6% of males, and 10%, 5%, and 3% of females aged over 14 in the USA, UK, and Australia, respectively (World Health Organization, 2018).

At the same time, the prevalence of abstinence from alcohol has gradually risen, with drinking declines driven by younger generations (under 40; Livingston et al., 2018). In 1990, 27% of males and 31% of females aged 15-19 in countries high on socio-demographic indices abstained from alcohol; by 2016 that had risen to 32% of males and 39% of females (Griswold et al., 2018).

1.1.2 Acute effects of alcohol consumption

Upon entering the digestive tract, alcohol (of which the primary component is ethanol) begins to be metabolised by enzymes – most notably, alcohol dehydrogenase (Edenberg & McClintick, 2018). From the stomach, alcohol is absorbed into the blood stream to the liver, where most metabolism occurs (Edenberg & McClintick, 2018). This metabolism produces various by-products, including acetaldehyde (Wilson & Matschinsky, 2020) – the resulting mixture of remaining ethanol and acetaldehyde travels through the blood stream and is dispersed in body water (Kezer et al., 2021), travelling to other organs like the heart and brain.

Vasodilation – the widening of blood vessels – occurs within five-to-ten minutes, resulting in warmth in the skin, one of the earliest noticeable effects of drinking (Nutt, 2020). In the brain, alcohol has considerable effects on glutamate and gamma-aminobutyric acid (GABA), which are the primary excitatory and inhibitory neurotransmitters respectively (Nutt, 2020; Wilson & Matschinsky, 2020). Ethanol consumption upregulates GABA, which precipitates anxiolytic effects. Mood is also affected by alcohol's promotion of serotonin, dopamine, and endorphins. However, increasing amounts of alcohol, and as a consequence, GABA, begin to

depress activity in the brain. The frontal cortex is also affected early, leading to the impaired judgement associated with intoxication (Nutt, 2020).

Once blood alcohol concentration (BAC) has reached 80mg/100mL of blood, glutamate begins to be downregulated (Nutt, 2020), further depressing cerebral activity (McIntosh & Chick, 2004), and at 150mg/100mL, the capacity for forming new memories is also impaired (Nutt, 2020). Even higher blood alcohol concentrations lead to impaired co-ordination, anaesthesia, and loss of consciousness. Extreme intoxication over 300mg/100mL can lead to coma and slowed respiration, with death possible above 400mg/100mL due to breathing failure (McIntosh & Chick, 2004).

1.1.3 Harms associated with heavy and disordered alcohol consumption

These physical effects of drinking have tangible consequences for health: heavy alcohol use is known to increase the risk for a wide range of adverse health outcomes, generating significant financial, societal, and personal costs. Globally, alcohol consumption was the seventh leading risk factor for disability-adjusted life-years (DALYs) and death in 2016 (Griswold et al., 2018), and in 2020 led to an estimated 1.78 million deaths (Bryazka et al., 2022). As a proportion of age-standardised deaths globally, it is estimated that alcohol accounts for 2.2% and 6.8% of female and male deaths, respectively (Griswold et al., 2018). This risk is particularly pronounced in males aged 15-49, for whom alcohol is the leading risk factor for mortality (Bryazka et al., 2022), accounting for 12.2% of deaths in this group (Griswold et al., 2018).

Much of this burden is due to acute conditions or events that occur during intoxication, with strong associations between alcohol and self-injury, as well as fatal motor vehicle injury

(NHMRC Clinical Trials Centre at the University of Sydney, 2019). Other short-term harms include alcohol poisoning, self-harm, and interpersonal violence (National Health and Medical Research Council, 2020). These acute harms are particularly prevalent in those aged under 40, for whom injuries account for the greatest proportion of disability-adjusted life years due to alcohol (Bryazka et al., 2022).

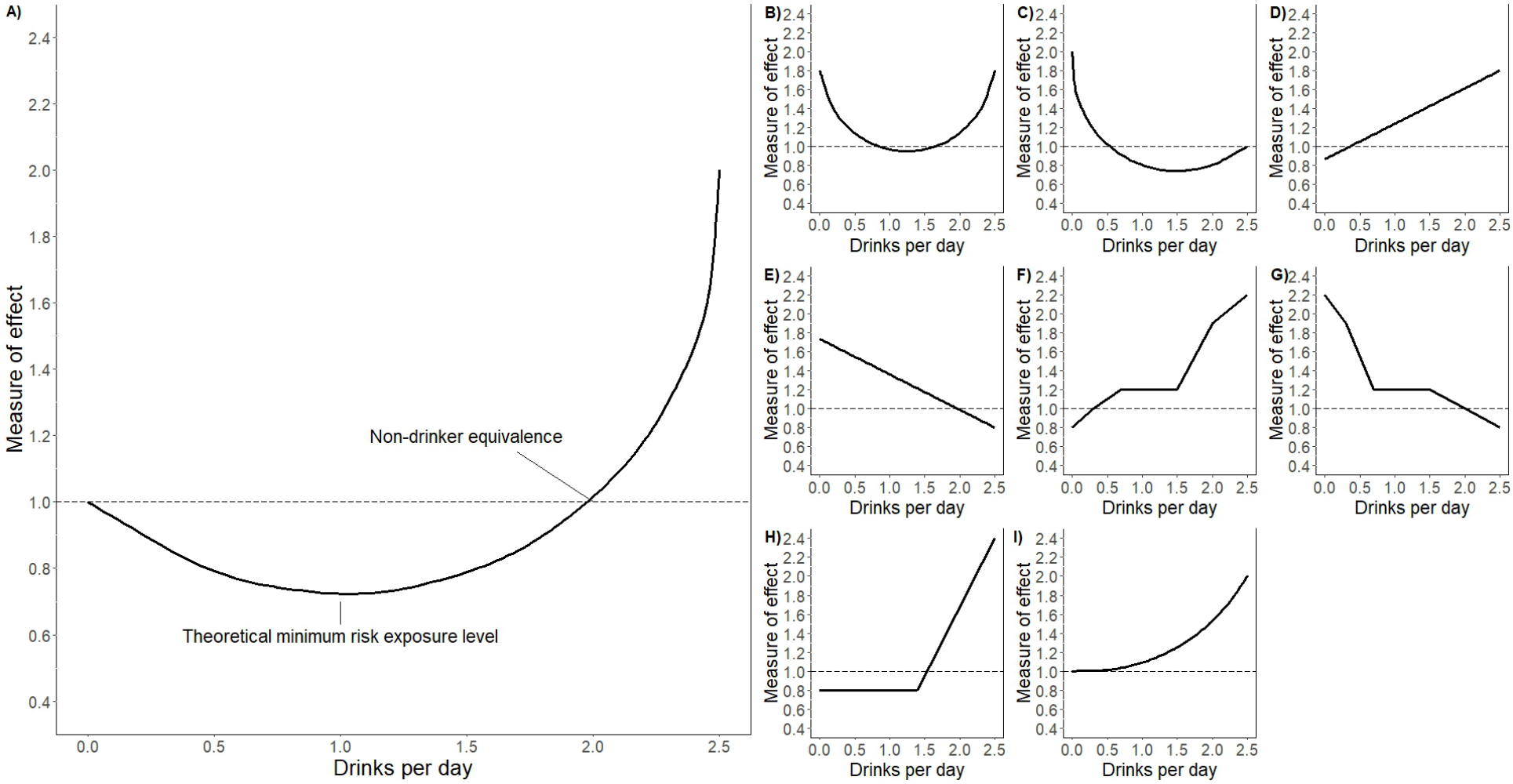
However, of particular focus for this thesis, there are also considerable long-term and chronic harms that arise from the cumulative effects of heavy alcohol consumption over time. Some of these conditions can be considered wholly-attributable to alcohol, and do not occur without drinking (Angus et al., 2019) – perhaps most notoriously, alcohol-related liver disease (Julien et al., 2020). However, most of these conditions are only partially-attributable to alcohol, meaning drinking is not necessary for them to develop, but that it increases risk (Angus et al., 2019). These harms, diverse in nature, generally begin to accrue in middle-to-older adulthood (Bryazka et al., 2022). For example, there is strong evidence for heavy alcohol consumption and AUD precipitating Wernicke-Korsakoff Syndrome (Visontay, Rao, et al., 2021) and increasing risk for all-cause dementia more generally (Schwarzinger et al., 2018). Heavy alcohol use is also known to increase risk for poor mental health, several cancers, and various cardiovascular diseases (CVD), among many other health outcomes across a range of bodily systems (Rehm, Gmel Sr, et al., 2017).

1.1.4 International drinking guidelines

Given the harms associated with heavy and disordered alcohol consumption, government health bodies in most countries prescribe both a standard drink size and recommend limitations on alcohol consumption. These guidelines vary considerably from country-to-

country, due to differing interpretations of the existing evidence base, currency (with a trend for decreasing upper limits over time), sex-specific/invariant guidelines, variability in what is considered an acceptable level of risk, and reliance on different metrics (i.e., lifetime risk of death vs years of life lost, consideration of mortality alone vs mortality *and* morbidity, and theoretical minimum risk exposure level [TMREL; the amount of alcohol consumption that coincides with the lowest risk; see Figure 1.1A] vs non-drinker equivalence [NDE; the amount of alcohol consumption that is equal to the risk associated with abstinence; see Figure 1.1A]).

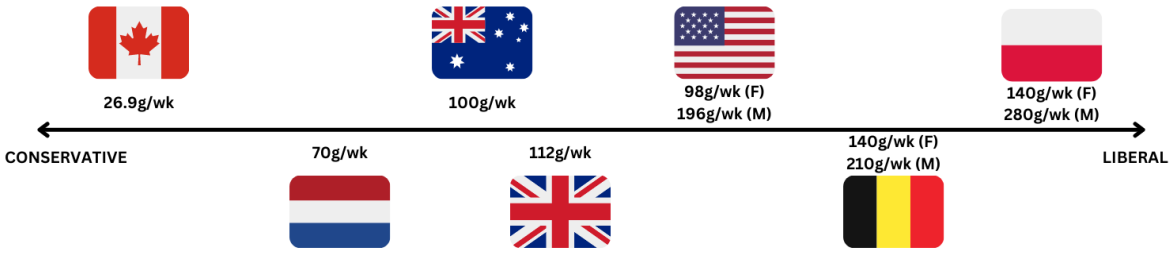
Figure 1.1. Exemplar functional forms between alcohol consumption and health outcomes.



A=J-shaped; B=U-shaped; C=reverse J-shaped; D=positive linear; E=negative linear; F= monotonically increasing; G=monotonically decreasing; H=threshold effect; I=exponential.

Figure 1.2 reflects these guidelines for several countries, with examples ranging from conservative (Canada and the Netherlands) to liberal (Belgium and Poland), as well as those for Australia, the UK, and USA – commonly used to define drinking categories in epidemiological research.

Figure 1.2. Low-risk drinking guidelines in a range of countries.



From left to right: Canada (Paradis et al., 2023), the Netherlands (Health Council of the Netherlands, 2015), Australia (National Health and Medical Research Council, 2020), the UK (Department of Health, 2016), USA (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020), Belgium (Sciensano, 2022), and Poland (State Agency for the Prevention of Alcohol-Related Problems [PARPA], 2019).

1.1.5 Evidence for protective effects of low-to-moderate alcohol consumption

However, evidence that heavy or disordered alcohol consumption leads to harm does not imply that *any* amount of alcohol is more harmful than abstinence. On one hand, there *are* several conditions where linear, accelerated/exponential, or monotonically increasing relationships with alcohol are observed (see Figure 1.1 for illustrations of common functional forms). These conditions include cancers at several sites, tuberculosis, and cirrhosis of the liver (Bryazka et al., 2022; Rehm, Gmel Sr, et al., 2017; Spanish Ministry of Health, 2020). For these health outcomes, the more one drinks, the more risk increases.

For a host of other conditions though, there is a substantial literature finding that low-to-moderate consumption is *protective* when compared to abstinence, particularly for certain

cardiometabolic, cognitive, and affective conditions, such as type 2 diabetes and depression. In fact, this phenomenon has even been observed for mortality from COVID-19 (Huang et al., 2022). These findings date back almost a century, when a study noted that moderate drinkers in a US sample had the lowest mortality risk (Pearl, 1926). When such protective effects are observed, the overall functional form of the relationship often resembles a 'J-shape' (Figure 1.1A), such that low-to-moderate consumption coincides with the nadir, compared to a slightly higher risk for alcohol abstainers and much greater risk for heavy consumers.

While long thought to be a consequence of the polyphenols found in red wine specifically, the proposed mechanisms for such protective effects are now largely attributed to ethanol itself – common to all alcoholic beverages (Krenz & Korthuis, 2012). These mechanisms (see Table 1.2) remain largely theorised rather than substantiated, given a lack of studies directly testing such mediation. In keeping with the J-shaped epidemiological relationship, it is interesting to note that alcohol at heavier doses also begins to exert opposite effects on many of these same mechanisms (Szabo, 2007; Tizabi et al., 2018).

1.2 Reasons for scepticism of alcohol's protective effects

J-shaped relationships, and the protective effects of low-to-moderate drinking that they entail, have long been taken for granted (Chikritzhs et al., 2009). They are embedded in the layperson's alcohol–health narrative (Barber, 2016; Bergeron et al., 2021), and are routinely taken into consideration by governments and health bodies when formulating public health policy and low-risk drinking guidelines.

However, scepticism within this field has been gradually developing, largely centred around key methodological limitations that restrict researchers' abilities to make causal inferences (i.e., confidence that a relationship's observed functional form and strength reflect a genuine *causal* effect of exposure on outcome) from the existing evidence base. Findings are inconsistent from study to study, and J-shaped relationships are sometimes found for certain outcomes (e.g., cirrhosis of the liver) that lack plausible biological mechanisms (Andreasson et al., 2014). This has prompted scrutiny of potential common biases that may be driving the identification of very similar functional forms for diverse health outcomes.

1.2.1 Methodological limitations and sources of bias

Limitations of traditional methods for observational research often lead to systematic errors in efforts to identify and estimate causal relationships, and thus biased findings. These biases are typically categorised as either measurement error, selection bias, or confounding (Kleinbaum et al., 1981). Not included in this classification scheme, but an important additional source of bias, is reverse causation. Each of these are nuisance factors in the study

of alcohol and its long-term health effects that are known to substantially impact observed effects (Callinan et al., 2019; T. S. Naimi et al., 2017; Shaper et al., 1988).

1.2.1.1 Measurement error/misclassification

Measurement error describes bias in the way exposure, covariate or outcome data are measured or recorded (Watkins, 2019). In the alcohol–long-term health context, measurement error is often present in the form of misclassification, a consequence of the categorisation of drinking status generally being based on one-time measurement. This fails to capture short-term variation in consumption (particularly 'regression dilution' – that extreme initial observations tend towards the mean on subsequent observations) as well as longer-term changes across the life course ('drinker drift'). This is particularly problematic in the case where misclassification likelihood is dependent on other variables associated with the outcome ('differential misclassification'; Watkins, 2019).

Respondents often misreport their own consumption, particularly in the form of underreporting. For example, longitudinal research has found that many who claim to be lifetime abstainers have reported drinking in previous surveys (Callinan et al., 2019). To this end, self-report survey data consistently result in lower estimates of population drinking than indicated by alcohol sales data (Kilian et al., 2020; Stockwell et al., 2018). The impact of misclassification can be severe – in one modelling study, former heavy drinking in all those misclassified as lifetime abstainers led to underestimation of alcohol-attributable all-cause mortality by roughly 15-20% (Rehm et al., 2008).

Additionally, many studies include former drinkers – whose abstinence may be precipitated by illness – in the abstaining reference group, thereby shifting poor health outcomes resulting

from drinking to the 'non-drinking' group (Shaper et al., 1988). Indeed, former drinkers have been found to have higher rates of cardiovascular disease, diabetes, anxiety, cancer, lung disease, and mortality (Ng Fat, 2014). This phenomenon is known as 'sick quitter bias' and is thought to be responsible for many previous J-shaped findings, prompting increasingly careful consideration of the composition of reference groups in these studies.

1.2.1.2 Reverse causation

In the case of sick quitters, misclassification introduces reverse causality, i.e., an unexpected causal relationship flowing from the designated outcome of interest to the designated predictor, either constituting the only causal relationship between these variables or existing in addition to a causal relationship flowing in the expected direction (see Figure 1.3). Here, poor health causes outright abstinence from alcohol consumption, or in a related phenomenon, substantial reductions in drinking frequency (Ng Fat et al., 2015). Similarly, reverse causality threatens causal inference in the presence of 'sick non-starters' (Ng Fat, 2014), where poor health from a young age may stop individuals from ever initiating drinking.

Figure 1.3. Depiction of a relationship with causation flowing in both directions

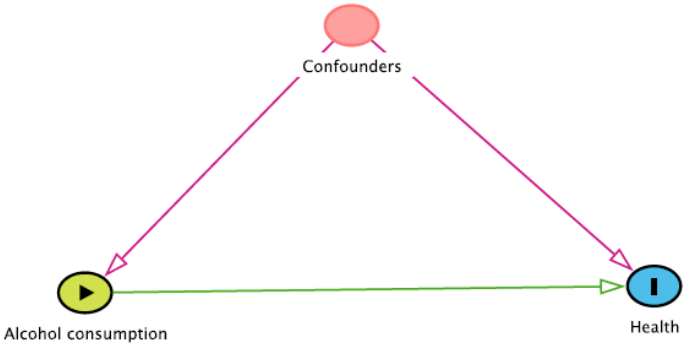
1.2.1.3 Selection bias

Selection bias captures a range of systematic errors in the selection of participants into a study itself, leading to a sample that is not representative of the target population (Delgado-Rodríguez & Llorca, 2004; Watkins, 2019). Selection bias may be present at a study's outset or may manifest later as certain participants fail to provide particular data or participate in follow-up waves entirely. In alcohol-health research, there are several selection biases at play. For example, 'healthy survivor bias' is likely when cohorts are too old at recruitment, meaning those whose drinking has already contributed to illness or death aren't represented (T. S. Naimi et al., 2017). More generally, drinkers who participate in research are likely to be healthier than those who don't (T. S. Naimi et al., 2017), as are those who remain participants compared to those lost to attrition.

1.2.1.4 Confounding

Confounding may constitute the biggest threat to causal inference. Whereas selection bias refers to skewed selection into/continuation in a study more generally, confounding results from non-random selection into a given exposure status within the study. When the variables that affect both exposure status and outcome are insufficiently accounted for, the results are said to be confounded. That is, while an exposure and outcome may *appear* causally related, this may only be due to some third variable that causes both exposure and outcome (see Figure 1.4). Indeed, several confounders, such as socio-economic status (SES; Kerr et al., 2017; T. S. Naimi et al., 2005) and limited health care access (T. S. Naimi et al., 2005), may be driving the relationship between alcohol abstention and poor health outcomes.

Figure 1.4. Directed acyclic graph (DAG) of a confounded relationship.



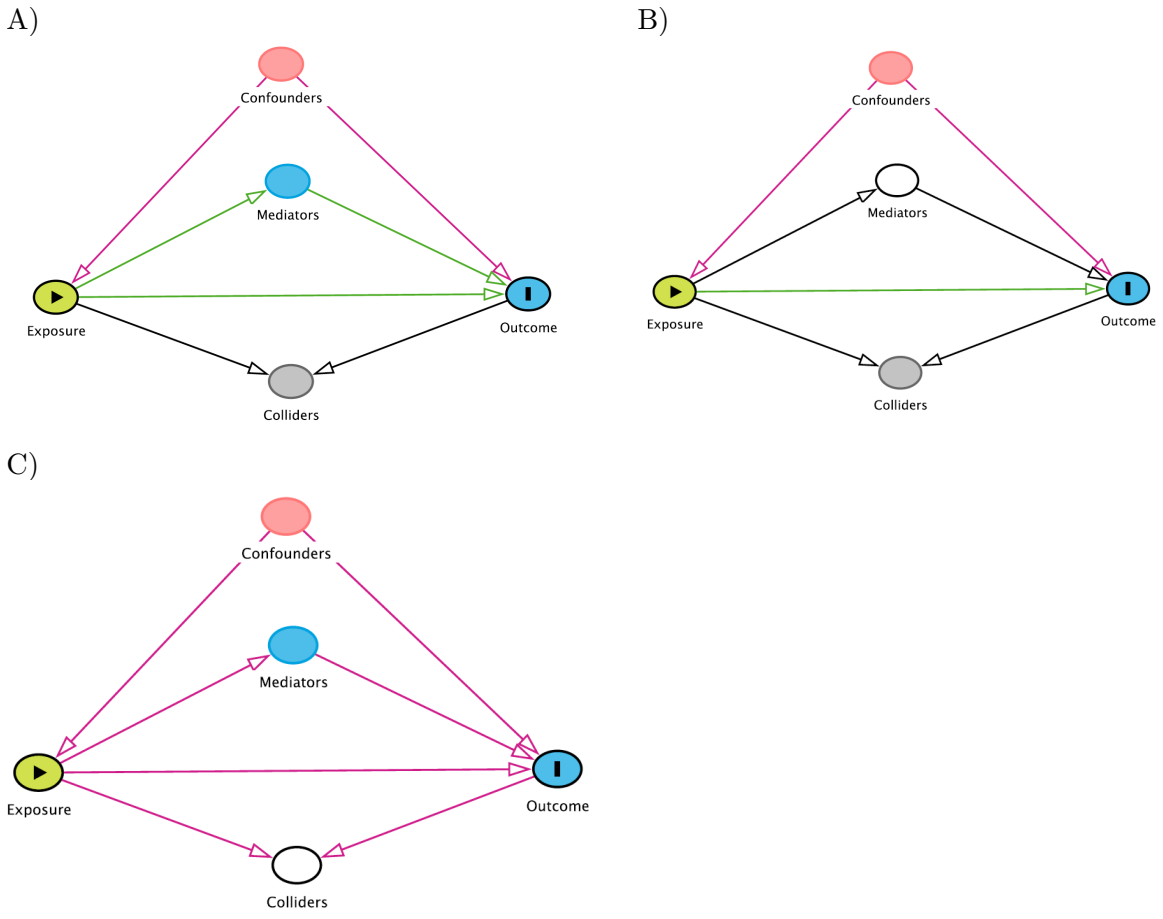
1.2.1.4.1 Limitations of standard approaches to address confounding

Uptake of improved approaches to mitigate confounding has been slow, with the standard methods in this field remaining measurement of and control for covariates via multivariable regression adjustment (entering covariates in the model, thereby holding them constant and isolating the effect of exposure), stratification (splitting the sample by levels of a given covariate and running the model separately within strata), or exact matching (making comparisons within subclasses comprised of individuals with identical values on a set of specified covariates; Wallach et al., 2020).

These methods have well-known limitations; regression with adjustment requires making distributional assumptions (Ali et al., 2014; Melberg, 2006) and correctly specifying the functional form (i.e., algebraic form in regression equations) of the relationship between covariates and outcome (Brookhart et al., 2013; Thoemmes & Ong, 2016). It also relies on extreme extrapolation when there are insufficient observations for all combinations of exposure, covariates, and outcomes (Greenland et al., 2016; Melberg, 2006; Thoemmes & Ong, 2016). Similarly, the 'curse of dimensionality' – that the number of groups to compare

increases exponentially with the inclusion of additional covariates, resulting in too few observations per group – is a limitation for matching and stratification methods (Melberg, 2006). Additionally, standard approaches generally cannot account adequately for time-varying exposures (as mentioned, exposures like alcohol consumption may not be stable over time) nor covariates which act as both confounders and mediators over time (Garcia-Huidobro & Michael Oakes, 2017).

Figure 1.5. DAGs of A) an exposure–outcome relationship with confounding and mediation; B) a confounded exposure–outcome relationship where mediators have been controlled for, thus blocking indirect effects; C) a confounded exposure–outcome relationship further biased by control for collider variables.



There is also the more basic problem of disagreement over which covariates should be controlled for when assessing alcohol and health outcomes (Wallach et al., 2020). Here, there is potential for relevant confounders to go unacknowledged and thus not controlled for (see Figure 1.5A, depicting in pink the biased pathways arising from confounding), as well as for over-adjustment for variables that are affected by/occur *after* exposure ('posttreatment' variables). These variables may be mediators, control for which will block the mediating path through which some or all of the effect occurs, and bias the effect estimate. In Figure 1.5B, the mediated path between exposure and outcome is blocked (represented by black lines), leaving only the direct effect from exposure to outcome open (in green).

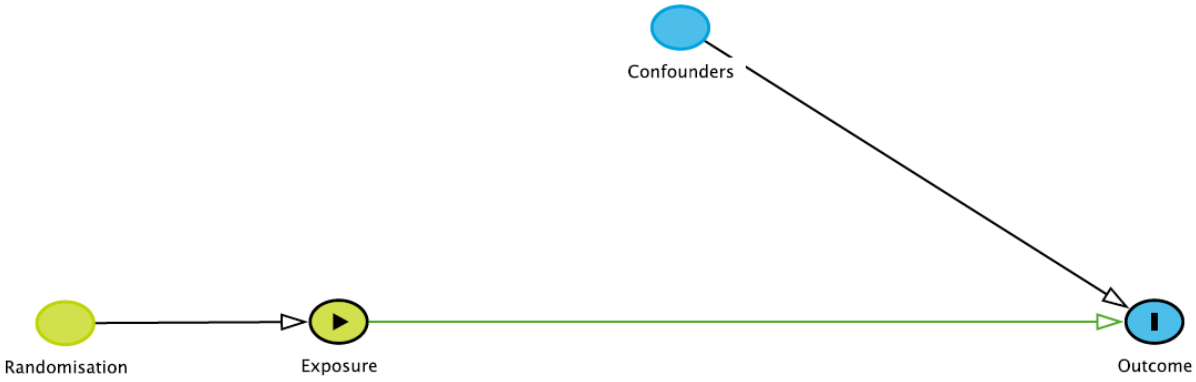
Alternatively, these variables may be 'colliders' (common effects of exposure and outcome), which when controlled for induce spurious correlations (see Figure 1.5C, in which adjustment for colliders opens additional biased pathways in pink; D'Onofrio et al., 2020; Garcia-Huidobro & Michael Oakes, 2017; Rohrer, 2018). Even if the right confounders are identified, they may be imperfectly measured or controlled for, leading to residual confounding. And in some cases, relevant confounders – particularly environmental ones like cultural or social factors – may be unmeasurable/difficult to measure (Ahlskog & Oskarsson, 2023).

1.2.2 Randomised controlled trials – the solution?

Many of these biases would be obviated by randomised controlled trials (RCT) in which participants are randomly assigned to different alcohol consumption levels and followed up on health outcomes. Figure 1.6 illustrates how the randomisation mechanism removes any association between potential confounders and exposure. And history alone gives us reason to suspect that results from an RCT might conflict with previous observational findings.

Traditional observational epidemiology has often erred in inferring mistaken causal conclusions from observed associations, with several high-profile, costly mistakes overturned by subsequent RCTs. These include the conclusions that hormone replacement therapy reduces risk for coronary heart disease, that antioxidant supplements protect against cancer, cardiovascular, and neurodegenerative diseases, and overestimation of the rates of adverse side effects associated with statins (Watkins, 2019).

Figure 1.6. DAG of a randomised controlled trial.



In short-term interventional studies, moderate alcohol consumption has been found to promote beneficial changes in several biomarkers of cardiovascular health (Brien et al., 2011). Similarly, a two-year RCT of people with diabetes found moderate wine consumption led to reductions in certain cardiometabolic risks (Gepner et al., 2015). While this provides some evidence for the benefits of moderate alcohol consumption, such benefits may only be transitory, and harms not yet apparent. Alcohol consumption can have complex cumulative and or/delayed effects (Kerr et al., 2019); without long-term follow-up, these cannot be observed.

But the complexities involved in conducting a long-term RCT evaluating alcohol and health outcomes, such as ethical concerns around long-term exposure to a known carcinogen, probable deviation from compliance to assigned long-term drinking/non-drinking levels, and immense cost, have proved insurmountable (Oppenheimer & Bayer, 2020). Indeed, a reliance on industry funding and subsequent undue influence on study design meant that the only planned long-term RCT – the MACH 15 trial, which was to run for roughly six years (Spiegelman et al., 2020) – was cancelled just three months after it began.

In summary, long-term RCTs are the ideal approach for exploring unanswered questions around alcohol and long-term health, but are not feasible. As will be discussed below, a practical solution lies in applying methods for analysing observational studies that mimic the key features of RCTs.

1.3 Causal inference frameworks and methodologies

Section 1.2 above outlines the many threats to valid causal inference in observational studies of alcohol–long-term health relationships, but also establishes that the field is reliant on these studies. What follows now is a basic introduction to dominant frameworks for conceiving of causality and causal inference, and to methods developed within these frameworks that promote causal inference from observational data.

1.3.1 Frameworks for thinking about causality

An early influential theory of causality in epidemiology emerged from a lecture given by Austin Bradford Hill in 1965 (Hill, 1965). Hill proposed nine principles of causation, akin to rule-of-thumb guidelines. He argued that each principle alone was neither necessary nor sufficient for an observed relationship to be deemed causal, but the more that were satisfied the stronger the case for causality. These principles are: strength of the association (with causality more likely when stronger effects are present), consistency (with causality more likely when associations replicate across time and contexts; see **Chapter 3**), specificity (with causality more likely if the relationship is a specific one), temporality (with causality requiring cause to precede effect), biological gradient (with, in general, causal relationships tending to demonstrate dose-response effects), plausibility (with causality more likely if there is a plausible underlying mechanism), coherence (with causality more likely if the relationship is consistent with other established, related knowledge), experiment (with causality more likely if manipulating the exposure changes the outcome as expected), and analogy (with causality more likely if the association is analogous to other, established causal relationships).

Hill's guidelines were, and continue to be, influential. However, in 1974, Donald Rubin proposed a more formal framework based on the concept of 'potential outcomes', or 'counterfactuals' (D. B. Rubin, 1974), that now dominates in the causal inference literature. This framework invokes the idea of counterfactual outcomes under alternative exposures/treatments to conceive of causal effects at the level of the individual. Accordingly, a definition of a causal effect is that an exposure has a causal effect for an individual if the

outcome that occurred for that individual under one exposure does not equal the outcome that *would have* occurred under another (Hernán & Robins, 2020; Watkins, 2019). The formal notation to express a causal effect is therefore:

$$Y_i^{a=1} \neq Y_i^{a=0}$$

Where Y_i is the outcome for a particular individual, and $a = 1$ and $a = 0$ are alternative values for exposure A.

This framework, which has also come to be known as the Rubin Causal Model, conceives of the challenge of causal inference as a missing data problem (Moodie & Stephens, 2022). That is, observational research limits the researcher to observing just one outcome, corresponding to one value of exposure, for any given individual (note that RCTs typically also only offer observation of outcomes under a single exposure for a given individual, but their design means that the missing counterfactual values arise by chance; Hernán & Robins, 2020).

Because of this, it is not possible to identify causal effects at the level of the individual. As such, researchers instead identify average causal effects: the difference between average outcome in those exposed and unexposed. At the population level, a causal effect exists if the expected values of an outcome under one level of exposure differ to the expected values under another level of exposure treatment (Hernán & Robins, 2020; Watkins, 2019). This can be formally expressed as:

$$\mathrm{E}\,[Y^{a=1}] \neq \mathrm{E}\,[Y^{a=0}]$$

This is crucially different from an *associational* effect at the population level, which exists if the expected values for an outcome under one level of exposure *for those who received that level of exposure* differ to the expected values under another level of exposure *for those who received that other level of exposure*. That is:

$$E[Y|A = 1] \neq \mathrm{E}\,[Y|A = 0]$$

Associations reflect conditional effects, i.e., they reflect the relationship between a level of treatment and an outcome only in those who actually received that level of treatment. In contrast, causal effects are marginal (unconditional) – they reflect the relationship between a level of treatment and an outcome in the entire population (Hernán & Robins, 2020).

1.3.2 Enhanced causal inference methods

Modern developments in causal inference from observational research can be viewed within the potential outcomes framework. They come in two main flavours, both of which aim to mimic or approximate the features of an RCT. These are statistical approaches and alternative observational designs (Hammerton & Munafò, 2021). The former comprises conventional cohort/case-control studies that apply enhanced analytical techniques. These make use of the same data as traditional observational studies but employ novel statistical methods. Alternative designs, on the other hand, require additional kinds of data or participants, such as the presence of a measured instrumental variable (IV; a randomised or as-if randomised variable that proxies for an exposure of interest), or related individuals (for use in within-family analyses). Of particular relevance for alcohol–long-term health research is a kind of instrumental variable method known as Mendelian Randomisation (MR; see **Chapter 5**), which uses genes to proxy for an exposure of interest. As genes are assigned at

birth, are distributed more-or-less randomly in the population with respect to covariates, and can't be affected by reporting bias, their use in place of an observed exposure largely avoids the problems of reverse causality, confounding, and measurement error (Burgess & Thompson, 2021; Ebrahim & Davey Smith, 2008). These features make MR a promising and cost-effective substitute for RCTs, and it has been an increasingly popular approach in recent years.

Table 1.1 lists key approaches in each of the statistical- and design-based methodological camps. Some of these, such as inverse probability of treatment weights (IPTWs), were developed directly within the potential outcomes framework (Moodie & Stephens, 2022). Others predate the Rubin Causal Model, but have been integrated into this framework, e.g., twin studies (Mcgue et al., 2010) and instrumental variables analysis (Angrist et al., 1996).

Table 1.1 Selected enhanced methods to improve causal inference from observational data

Statistical Approaches	
Propensity scores (PS) (Benedetto et al., 2018; Brookhart et al., 2013)	-The PS is a single value reflecting the probability of exposure for an individual given their values on all relevant covariates -Once generated, the PS can be used for matching, stratification, weighting (using IPTWs), or as a covariate for adjustment in regression
G-methods (Hernán & Robins, 2020; Mansournia et al., 2017; A. I. Naimi et al., 2017; Watkins, 2019)	-A family of methods developed as a solution to the problem of time-varying covariates affected by past exposure, including those that act as both confounders and mediators over time -The three G-methods are the G-formula, marginal structural models, and G-estimation, each relying on its own modelling assumptions
Doubly robust methods (Schuler & Rose, 2017)	-Incorporate both an estimation of the outcome mechanism (as in G-formula or regression adjustment) and the exposure mechanism (as in propensity scores) -Examples include targeted maximum likelihood estimation; augmented inverse probability weighting
Fixed effects regression (Fergusson et al., 2009, 2015; Gunasekara et al., 2014)	-A technique for use with longitudinal data with repeat outcome measurements, only using information on within-subject variation, thus controlling for all time-invariant sources of confounding -Each participant serves as their own control
Causal mediation analysis (Pearce & Lawlor, 2017; Richiardi et al., 2013; VanderWeele, 2016)	-Integrates traditional mediation analysis with the potential outcomes framework to allow for exposure-mediator interaction and non-linear relationships (i.e., is a non-parametric method) -Makes explicit underlying assumptions related to unmeasured confounding, and encourages sensitivity analyses to test robustness to assumption violations

Alternative observational designs	
Natural experiments (Davies et al., 2018; Dunning, 2012; Gage et al., 2016; Richmond et al., 2014)	-Mimic RCTs by exploiting exogenous events that are truly randomised/approximate random assignment -Differ from true experiments in that exposure is not assigned by the researcher -Assignment may be as a result of naturally occurring phenomena (e.g., a weather event), or of human intervention implemented for reasons other than the research question (e.g., army draft lottery) -Includes instrumental variable analysis, such as those using genetic variants as proxies for exposure (known as 'Mendelian Randomisation' studies)
Quasi-experiments (Dunning, 2012)	-Like natural experiments, exploit exogenous events to assess relationships between exposures and outcomes, but lack random or as-if random assignment
Family-based designs (Gage et al., 2016; Richmond et al., 2014)	-By comparing genetically related participants discordant for the exposure or outcome of interest, account for confounding from genetic or shared environmental sources
Negative controls (Gage et al., 2016; Lipsitch et al., 2010)	-Have the same confounding structures as the exposure–outcome relationship of interest, but lack a plausible causal mechanism -If association is greater for the relationship of interest than for the negative control, a causal relationship is likely; if not, suggests confounding/other shared biases responsible -May take the form of a negative control exposure or a negative control outcome

1.4 The need for causal inference approaches in alcohol–long-term health research

While some of the aforementioned statistical and design-based causal inference methods are becoming more popular in health research (Watkins, 2019), their application to alcohol–long-term health research specifically is only nascent, with calls for greater implementation (McQuire & de Vocht, 2019). In mitigating confounding and other biases that have plagued observational research to date, the research contained in this thesis and the future work it lays the foundations for will contribute to a greater understanding of causal relationships in this space.

1.4.1 Which alcohol–long-term health relationships are of greatest concern?

While J-shaped relationships have been found for a wide range of conditions, there are a select group of outcomes for which they are particularly consistent, but for which robust, causally-focussed analysis is particularly lacking.

The first of these is type 2 diabetes (T2D). As reviews, meta-analyses, and modelling work track the evolving literature over time, a clear trend has emerged wherein the protective effects of low-level drinking diminish in confidence and strength, if not vanish entirely. T2D is one of the only conditions to resist this trend (Rehm, Gmel Sr, et al., 2017) – indeed, the modelling evidence informing the world's most conservative guidelines (in Canada, just 2 drinks per week; Paradis et al., 2023) admits of benefits for this condition.

Relationships between drinking and mental health outcomes also require urgent analysis. Given the well-established and substantial harms to mental health caused by heavy and

disordered drinking (Boden & Fergusson, 2011; Fergusson et al., 2009; Li et al., 2020), potential protective effects at the lower end of consumption demand thorough scrutiny, critical interpretation, and cautious translation into recommendations. However, existing evidence is so sparse and of such low quality that mental health conditions could not be incorporated in the modelling underpinning the most recent Global Burden of Disease (GBD) alcohol report (Bryazka et al., 2022), Australian low-risk drinking guidelines (National Health and Medical Research Council, 2020), or Canadian low-risk drinking guidelines (Paradis et al., 2023).

Finally, while not an end-point health condition in itself, chronic inflammation is increasingly implicated in a host of serious downstream outcomes including cancer and cardiovascular disease, as well as both psychiatric disorders and T2D themselves (Ansar & Ghosh, 2013; Avan et al., 2018; Furman et al., 2019; Réus et al., 2015). These conditions all contribute substantially to disease burden across the life-course, so epidemiological work that can refine our understanding of their lifestyle risk and protective factors is a high priority.

Table 1.2 Key outcomes with observational evidence of protective effects for a non-zero level of alcohol consumption

Condition	**Recent meta-analysis or review suggesting protective effects**	**Proposed mechanisms for protection**[a]
Inflammation	N/A[b]	Alcohol may inhibit pro-inflammatory cytokines and chemokines (Muralidharan et al., 2014; Szabo, 2007). Less directly, it may also exert a protective effect mediated by stress reduction.
Depression	(Li et al., 2020)	Alcohol's effect on depression may be directly biologically mediated, i.e., via GABAergic and dopaminergic effects, increases in brain-derived neurotrophic factor, and reduced inflammation. Another mediating pathway may be stress reduction and facilitation of social interaction (Gémes et al., 2019; Tizabi et al., 2018).
Type 2 Diabetes	(Griswold et al., 2018; Rehm, Gmel Sr, et al., 2017)	Alcohol may improve insulin resistance/sensitivity (Krenz & Korthuis, 2012), potentially via reducing glucose uptake and increasing glucose absorption and glycogenolysis (Oh et al., 2021). Alcohol may also improve insulin resistance via improvements in inflammation, glucose metabolism, and endocrine functioning of fat tissue/adiponectin modulation (Hendriks, 2007).

[a] Derived from a mix of epidemiological, pre-clinical, and clinical (short-term administration of alcohol) evidence.

[b] There is also a dearth of longitudinal observational studies of this relationship (see Table 2.3).

1.4.2 What is at stake for policy, practice, and research?

Uncertainty in our understanding of alcohol–long-term health relationships muddies the waters for researchers, policymakers, clinicians, and individuals. Associations found between moderate alcohol consumption and good health have had, and will continue to have, a great impact on safe drinking guidelines and public health policy, often offsetting evidence on the

harms of heavier levels of drinking. For example, Figure 1.7 depicts results from modelling conducted for the recent National Health and Medical Research Council guidelines in Australia, which incorporates evidence of protective effects of low-level consumption. However, sensitivity analyses reveal that the number of standard drinks per week associated with an acceptable level of health risk is reduced severalfold when assuming no protective effects of lower-level consumption (Angus et al., 2019). Spread across three drinking days per week (the Australian average), the primary modelling indicates men can consume 12.5 drinks per week, and women 10.5 drinks per week, before lifetime risk of death from alcohol-related disease or injury exceeds 1 in 100. Testing a scenario in which protective effects are assumed erroneous – i.e., not causal – reduces the estimates to *just 2.5 drinks per week for both sexes*, far fewer than the 10 drinks per week that became the official guideline.

Figure 1.7. Modelling results from the Australian low-risk drinking guidelines

Table A2.2. Number of standard drinks per week associated with a 1 in 100 risk of mortality according to whether protective effects of alcohol are considered

Drinking frequency	Male		Female	
	Protective effects	No protective effects	Protective effects	No protective effects
	Standard drinks (per week)			
Daily	20.2	2.9	15.3	2.3
6 days/week	18.6	2.8	14.5	2.2
5 days/week	16.9	2.5	13.5	2.2
4 days/week	14.9	2.6	12.1	2.1
3 days/week	12.5	2.5	10.5	2.5
2 days/week	9.0	2.6	7.8	2.2
1 day/week	4.1	0	4.7	0.1

It is clear then that the potential policy and public health impacts hinging on this research are substantial. The results of this work are also essential to enabling informed decision-making by clinicians and individuals, who, as noted earlier, often presume benefits of low-level alcohol consumption when reasoning about drinking. Particularly important for this endeavour is research that provides granular detail, i.e., disaggregating risk by demographic and other characteristics, allowing for personalised decision-making. Finally, early work establishing the merits and limitations of various methodologies for a causally-focussed alcohol–health epidemiology will be of great value to this research area as it develops.

As such, the work in this thesis stands to make a substantial contribution to the future of public health and disease burden, policy, and research practice.

1.5 Thesis aims and outline

The overall aim of this thesis is to further the application of causal inference methodologies to alcohol–long-term health research, and to advance substantive knowledge of key complex epidemiological relationships within the field.

Given causal inference is rarely the focus of individual studies of alcohol-long-term health relationships, let alone syntheses of the literature, **Chapter 2** was designed to be the first systematic review of relationships between alcohol and all long-term health outcomes from studies employing causal inference approaches. A total of 16 studies are reviewed, covering a range of health outcomes and methodologies.

Each of **Chapters 3-5** – representing original research – applies a different sophisticated methodology promoting causal inference to a given alcohol–long-term health relationship requiring urgent elucidation.

Chapter 3 uses enhanced sensitivity frameworks ('multiverse analysis') to assess how robust the association between drinking and inflammation is to common variations in researcher-defined analysis parameters. This is valuable work given that a consistent, robust association is a necessary preliminary condition for causation, and is the first paper to apply multi-dimensional, multiverse analyses to a study of alcohol and long-term health in a single cohort.

Chapter 4 employs a marginal structural model to assess the longitudinal relationship between alcohol consumption and depression. Unlike existing work in this space, the robust methodology applied in this analysis adequately accounts for dynamic confounding variables and changes in alcohol consumption over time. The generated estimates have a straightforward interpretation for health policymakers and individuals.

Chapter 5 reports on a study applying non-linear MR analyses to the relationship between alcohol consumption and T2D. The focus on sex-stratification, exhaustive sensitivity analyses, and comparisons to results from both linear MR and traditional observational models make this one of the most robust and nuanced evaluations of this epidemiological relationship to date.

Chapter 6 contains a general discussion of the findings and implications of the thesis.

Chapter 2

Causal inference in studies of alcohol consumption and long-term health: A systematic review

Preface

When it comes to alcohol and long-term health, reviews of the evidence base with an *explicit* causal focus are sparse. Existing reviews are limited in scope to combinations of a specific causal inference method and specific health outcome (e.g., MR and cardiovascular disease; van de Luitgaarden et al., 2021). There have been no attempts at systematic reviews that are comprehensive in either their coverage of health outcomes or of enhanced causal inference approaches, let alone both. Consequently, a clear and broad picture of the extent to which such methods have been applied, and what the results from these studies show – most importantly, whether they are concordant with findings from traditional epidemiological methods – was missing. This chapter reports on a comprehensive systematic review that addressed this gap by identifying all observational studies employing causal inference approaches and by synthesising their findings on the functional form and strength of alcohol–long-term health relationships.

A variety of functional forms are found, including reverse J/J-shaped relationships for prostate cancer and related mortality, dementia risk, mental health, and certain lipids. However, results tend to be discrepant even within specific health outcomes, most outcomes are only evaluated by a single study, and few studies provide information on the role of

alcohol consumption pattern. This review makes plain that more research employing enhanced causal inference methods is urgently required to accurately characterise alcohol–long-term health relationships.

A slightly modified version of this chapter was published as: Visontay, R., Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2022). Are there non-linear relationships between alcohol consumption and long-term health?: a systematic review of observational studies employing approaches to improve causal inference. *BMC Medical Research Methodology, 22*(1), 1-28 (available in Appendix A). Its protocol has been published as: Visontay, R., Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2021). Are there non-linear relationships between alcohol consumption and long-term health? Protocol for a systematic review of observational studies employing approaches to improve causal inference. *BMJ Open*, *11*(3), e043985, and is available in Appendix B. Supplementary materials are available in Appendix 1.

2.1 Introduction

Chapter 1 established the muddied state of the evidence base on alcohol and long-term health: J-shaped relationships are often found between alcohol and a variety of outcomes, but findings are inconsistent, with many discrepant functional forms reported. Which of these reflect true causal relationships, and which are merely methodological artefacts, remains unclear, with several biases – in particular, confounding – likely at play.

As established, while confounding is generally obviated by the randomisation mechanism in RCTs, long-term RCTs in this context are not feasible. To investigate alcohol–long-term health relationships, the field is therefore limited to observational studies. As such, efforts to improve causal inference have centred on mitigating bias. Certain tools and strategies can assist with limiting bias, such as creating directed acyclic graphs (DAGs) at study outset, and, following primary analysis, assessing robustness to sample-specific confounding ('cross-cohort comparison') or research type ('triangulation'; Richmond et al., 2014).

Particularly promising, however, for addressing the identified limitations of existing research, are modern methods for data analysis and alternative observational designs, some of which were introduced in **Chapter 1**. Conventional designs (e.g., prospective cohort studies) can be enhanced with modern analysis methods, such as propensity scores used for matching or weighting, and 'G-methods' such as marginal structural models (MSMs; which can account for time-varying variables that act as both confounders and mediators). Regarding alternative designs, twin studies and other family-based designs control automatically for shared confounders, as do negative controls (Lipsitch et al., 2010). Natural experiments are another alternative, mimicking the random allocation of RCTs and thus guarding against

confounding and reverse causation. These include IV designs, where as-if/randomly allocated instruments for exposures are used in place of exposures themselves. Mendelian Randomisation (MR), a kind of IV analysis, offers particular promise given the potential of genetic proxies for alcohol consumption.

While none of these methods are guarantees against erroneous causal inferences, they represent significant improvements over conventional analyses (see Table 2.1 for a full list of methods of interest for this chapter, their advantages, and their limitations). Some of these approaches are gaining popularity (Watkins, 2019), but they are not routinely applied to alcohol–long-term health research (McQuire & de Vocht, 2019). Importantly, reviews in this area rarely focus on improved analytical methods to counter confounding, and tend to exclude novel study designs. This review therefore aims to identify all observational studies employing such approaches, and to synthesise their findings on the functional form and strength of alcohol–long-term health relationships.

Table 2.1. Methods to enhance causal inference from observational research

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
								
Statistically-based methods								
Propensity scores (PS) (Benedetto et al., 2018; Brookhart et al., 2013)	Covariate balancing propensity scores (CBPS)	-The PS is a single value reflecting the probability of exposure for an individual given their values on all relevant covariates -PS generation occurs as a data 'pre-processing' step prior to main analysis -Usually generated via logistic regression -Once generated, the PS can be used for matching, stratification, weighting, or as a covariate for adjustment in regression	✓				-Unlike standard methods, can handle large numbers of covariates -Not reliant on correctly modelling covariate-outcome relationships -Covariate balance after matching/weighting can be assessed	-Still relies on appropriate choice of covariates and accurate measurement -PS matching may not find matches for some treatment cases, leading to reduced sample size and limiting generalisability -Effect estimation with PS doesn't always perform better than regular adjustment -Most PS methods rely on manual covariate balance checking and refitting
G-methods (Hernán & Robins, 2020; Mansournia et al., 2017; A. I. Naimi et al., 2017; Watkins, 2019)		-A family of methods intended for use with time-dependent variables -Developed as a solution to the problem of time-varying covariates affected by past exposure, including those that act as both confounders and mediators over time -The three G-methods are the G-formula, marginal structural models, and G-estimation, each relying on its own modelling assumptions					-Unlike standard methods to control for confounding, G-methods do not fix values of covariates, thus do not block mediation via the covariate, and avoid introducing collider bias -Accounting for changes in variables over time mitigates misclassification	-Still relies on appropriate choice of covariates and accurate measurement
	G-formula (aka G-computation or G-standardisation) (Mansournia et al., 2017; Schuler & Rose, 2017; Watkins, 2019;	-First models relationship given observed data (using actual exposure for each individual), and then predicts outcomes under counterfactual exposures, with the difference taken as the causal effect -Is a generalisation of standardisation (conditioning on covariates and then	✓		✓[a]	✓	-Can be used to calculate risk ratios and risk differences -Well-suited to assess time-varying exposures	-Vulnerable to the 'g-null paradox'- null hypotheses tend to be rejected in large studies even when true -Usually requires specifying a statistical model, hence also being known as the 'parametric' G-formula

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
								
	Williamson & Ravani, 2017)	marginalising) that accounts for dynamic variables by considering covariate distribution over follow-up time						
	Marginal structural models (MSMs) (Mansournia et al., 2017; Watkins, 2019; Williamson & Ravani, 2017)	-Use weights based on inverse probability of exposure at each time point to create a pseudo-population where each combination of covariates is equally present in each exposure condition -Using these weights, MSMs then estimate the causal effect -The most popular of the G-methods	✓	✓	✓[a]	✓	-Simplest of the G-methods to understand and implement -Can also integrate censoring weights to account for differential attrition	-Not ideal for assessing exposure-confounder interactions, with standard MSMs unable to estimate interactions involving dynamic variables -Requires checking weight distribution, may require refitting (as with PS methods) -Cannot be used if all participants are exposed/unexposed on a particular level of a confounder
	G-estimation of structural nested models (Hernán & Robins, 2020; Mansournia et al., 2017; Williamson & Ravani, 2017)	-At each wave assesses the relationship between exposure and likelihood of outcome given covariates, adjusting for exposure and covariate values from past waves, thus accounting for dynamic confounders affected by past exposure -Considered semi-parametric in that mean counterfactual outcomes under no exposure are unspecified	✓			✓	-Can be used even if all participants are exposed/ unexposed on a particular level of a confounder -Can be used to assess exposure-confounder interactions	-Unlike the other G-methods, cannot account for selection bias arising from censoring, so data requires preliminary weighting to account for bias from censoring
Doubly robust methods (Schuler & Rose, 2017)	Targeted maximum likelihood estimation; Augmented inverse probability weighting	-Incorporates both an estimation of the outcome mechanism (as in regression adjustment) and the exposure mechanism (as in propensity scores)	✓				-Only the outcome mechanism or the exposure mechanism need be consistently estimated to generate an unbiased estimate of exposure effect	-Still relies on appropriate choice of covariates and accurate measurement
Fixed effects regression (Allison, 2005; Fergusson et		-A technique developed in the econometrics literature for use with longitudinal data with repeat outcome measurements, only	✓				-Removes the threat of all observed and unobserved time-invariant confounding	-Cannot overcome time-varying confounding without extending the model, and these variables must be observed/measured

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
al., 2007, 2009, 2015; Gunasekara et al., 2014)		using information on within-subject variation, thus controlling for all time-invariant sources of confounding -Treats time-invariant characteristics that differ between individuals as fixed parameters (unlike in mixed models), allowing estimation of parameters of interest net of stable confounders -Each participant serves as own control					-Models can be extended to include time-varying covariates	-Individuals with stable exposure values do not contribute to estimates; also leads to imprecision when exposures change little over time -Reducing confounding comes at the cost of more sampling variability -Cannot generate parameter estimates for stable characteristics like race
Causal mediation analysis (Pearce & Lawlor, 2017; Richiardi et al., 2013; VanderWeele, 2016)		-Integrates traditional mediation analysis (which separately estimates total effect of exposure on outcome, indirect effect via mediators, and direct effect unexplained by mediators) with the potential outcomes framework to allow for exposure-mediator interaction and non-linear relationships (i.e., is a non-parametric method) -Uses the concepts of 'controlled direct effect', 'natural direct effect', and 'natural indirect effect' -Makes explicit underlying assumptions related to unmeasured confounding, and encourages sensitivity analyses to test robustness to assumption violations	✓				-Effect decomposition is still possible given exposure-mediator interaction, non-linearity, and categorical variables -Makes underlying assumptions explicit -Helps identify the mechanism/s of an exposure's effect; especially useful when heterogenous causal mechanisms at play -Can be extended to situations with multiple mediators -If a variable completely mediates the exposure-outcome relationship and is shielded from confounders, confounder measurement is not needed	-In practice, likely there will always be exposure-mediator or mediator-outcome confounding, so still need to observe and accurately measure covariates -Analyses make strong assumptions, necessitating sensitivity analyses

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
								
Design-based methods								
Natural experiments (Davies et al., 2018; Dunning, 2012; Gage et al., 2016; Richmond et al., 2014)		-Mimic RCTs by exploiting exogenous events that are truly randomised/approximate random assignment -Differ from true experiments in that exposure is not assigned by the researcher -Assignment may be as a result of naturally occurring phenomena (e.g., a weather event), or of human intervention implemented for reasons other than the research question (e.g., army draft lottery)					-In approximating randomisation, obviates the need for accurate measurement of confounders -Potential to overcome measurement error, reverse causation, and selection bias	-Rare to find truly random or as if-random exposure assignment
	Standard natural experiments	-Natural experiments where individuals are as-if/randomly assigned to exposure and control groups	✓	✓	✓	✓		-May be difficult to find a standard natural experiment that maps on well to the actual research question of interest
	Instrumental variable analysis	-Assesses the relationship between an as-if/randomly assigned *proxy* for the exposure of interest and the outcome -A valid instrumental variable must be associated with the exposure of interest, be independent of confounders of the exposure-outcome relationship, and should affect the outcome only via the exposure	✓	✓	✓	✓	-Useful when the exposure itself is difficult to manipulate or measure	-Difficult to find valid instrumental variables -Potential for weak instrument bias i.e., when the instrument explains a small amount of variance in exposure -Relies on assumption that the instrumental variable is not associated with exposure-outcome confounding
	Genetic instrumental variables	-Kind of instrumental variable analysis using genetic variants as proxies for exposure -The most prominent technique is Mendelian Randomisation	✓	✓	✓	✓	-Genes cannot be confounded by environment, cannot be subject to reverse causality, and are stable over time -Multiple variants can be combined to explain more variance in exposure, mitigating weak instrument bias	-Genetic instrumental variables are proxies for lifelong exposure, with limited ability to isolate the effects of exposure over a smaller period or to assess the impact of a time-varying exposure -Potential for violation of assumptions via weak instrument bias, pleiotropy (gene affecting more than one phenotype) and linkage disequilibrium

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
								
								(when genes are not inherited independently of one another) -Possible population stratification (there may be population subgroups with different distributions of genes)
Quasi-experiments (Dunning, 2012)		-Like natural experiments, exploit exogenous events to assess relationships between exposures and outcomes, but lack random or as-if random assignment	✓	✓		✓	-Same as for natural experiments	-Without random assignment, confounding is still possible i.e., cause of exposure may also contribute to outcome
Family-based designs (Gage et al., 2016; Richmond et al., 2014)	Twin studies; Sibling comparison	-By comparing genetically related participants discordant for the exposure or outcome of interest, accounts for confounding from genetic or shared environmental sources	✓				-Controls for unmeasured/unobservable confounding (for shared covariates) -Comparing monozygotic and dizygotic twins enhances understanding of genetic vs. environmental confounding	-Still need to observe and accurately measure non-shared environmental covariates to control for this kind of confounding -May be difficult to find family members discordant for exposure of interest
Negative controls (Gage et al., 2016; Lipsitch et al., 2010)	Negative control exposures; Negative control outcomes	-Have the same confounding structures as the exposure–outcome relationship of interest, but lack a plausible causal mechanism -If association is greater for the relationship of interest than for the negative control, a causal relationship is likely; if not, suggests confounding/other shared biases responsible -May take the form of a negative control exposure or a negative control outcome	✓		✓[b]		-Can identify when confounding (or assumed shared bias) is responsible for apparent causal effects -Does not require observation/measurement of covariates	-Relies on assumption of no plausible causal mechanism in the negative control relationship -Relies on assumption that same confounding structure shared by relationship of interest and negative control relationship

 = Confounding  = Reverse causation = Selection bias  = Measurement error

[a] Specifically, bias due to censoring.

[b] Specifically, immortal time bias.

2.2 Methods

The protocol for this review was registered with the PROSPERO International Prospective Register of Systematic Reviews of the University of York (CRD42020185861), and published (Visontay, Sunderland, et al., 2021).

2.2.1 Search strategy and study selection

Electronic databases (MEDLINE, PsycINFO, Embase, and Scopus) were searched in May 2020 for peer-reviewed, English-language journal articles, with no restrictions on date range. Relevant grey literature returned by Scopus was also considered. The search strategy was designed to capture all studies using observational data to assess the relationship between varying levels of alcohol exposure and any health outcome, and that employed at least one of the pre-specified approaches to improving causal inference.

Searches ran a combination of MeSH or other controlled vocabulary terms as well as free text words. Five groups of terms were used in the search strategy, relating to: 1) Alcohol; 2) Levels or patterns of drinking 3) Observational, longitudinal studies; 4) Approaches to improve causal inference that are used in conjunction with observational, longitudinal study designs ('statistically-based' methods); and 5) Approaches to improve causal inference that may be considered their own study designs ('design-based methods'). These concepts were combined with the following logic: 1 and 2 and ((3 and 4) or 5). Search terms were adapted from recent systematic reviews and from keywords/indexing of a set of key eligible papers known to the study authors. These terms were iteratively refined to improve sensitivity by cross-checking results against the set of key papers. The MEDLINE search terms are

provided in Appendix 1.1. Additionally, reference lists of eligible, retrieved studies were manually searched.

2.2.2 Eligibility

Only human research was eligible. The exposure of interest was level of alcohol consumption (volume over a given period), or level and pattern of consumption (incorporating frequency/HED). While studies were eligible regardless of their findings on functional form, their methods must have been capable of detecting non-linearity – were it to be present. For this reason, studies were only eligible if they categorised alcohol consumption, and subsequently performed comparisons between a chosen reference category and the other levels of consumption. This approach does not require assuming a functional form (unlike a single regression using a continuous predictor). Specifically, a non-drinking/light drinking reference was required in addition to at least two other levels of consumption (alternative methods of comparison allowing for the detection of non-linearity were permitted for IV/MR designs).

Any long-term health outcome was eligible; studies only reporting on acute/short-term conditions (e.g., injury) were excluded. Eligible studies needed to employ one of the pre-specified approaches to improving causal inference (i.e., those contained in Table 2.1). Incorporating feedback from academics with expertise in causal inference methods, these approaches were clearly defined *a priori* and have been increasingly recognised in recent years for their potential to enhance causal inference. Studies needed to be longitudinal cohort or case-control designs (excepting IV/MR designs). IV/MR studies must have performed

formal IV analysis or otherwise provided estimates in terms of predicted alcohol consumption. Reviews and interventional studies were excluded.

Retrieved titles and abstracts were screened by one reviewer (RV), with a second reviewer (JW) additionally screening a random 25%. Full-text articles were independently assessed by two reviewers (RV and JW). A third reviewer (LM) was consulted regarding unresolved discrepancies.

2.2.3 Data extraction

Extraction was performed independently by two reviewers (RV and JW) using pre-piloted forms. Extracted data included publication details (author/s, year), participant characteristics (sample size, setting, mean age, eligibility criteria, cohort name), exposure details (number and spread of measurement occasions, nature of discretised categories), study design and analysis methods, health outcome/s (how assessed, whether binary/continuous, interval to measurement), and results (relationship strength and form). Study authors were contacted if further information was required (see Appendix 1.2).

2.2.4 Quality assessment

Given the range of designs targeted by this review, two risk of bias assessment tools were used. Cohort studies were assessed using the relevant Newcastle-Ottawa Scale (NOS; Wells et al., n.d.), and a recently developed tool specific to MR (Mamluk et al., 2020) was employed. One reviewer (RV) applied the tools to all studies, with a second reviewer (JW) additionally assessing a random 25%. In line with other similar reviews, formal assessment of evidence quality was limited to risk of bias (Mamluk et al., 2017, 2020).

2.2.5 Synthesis and reporting

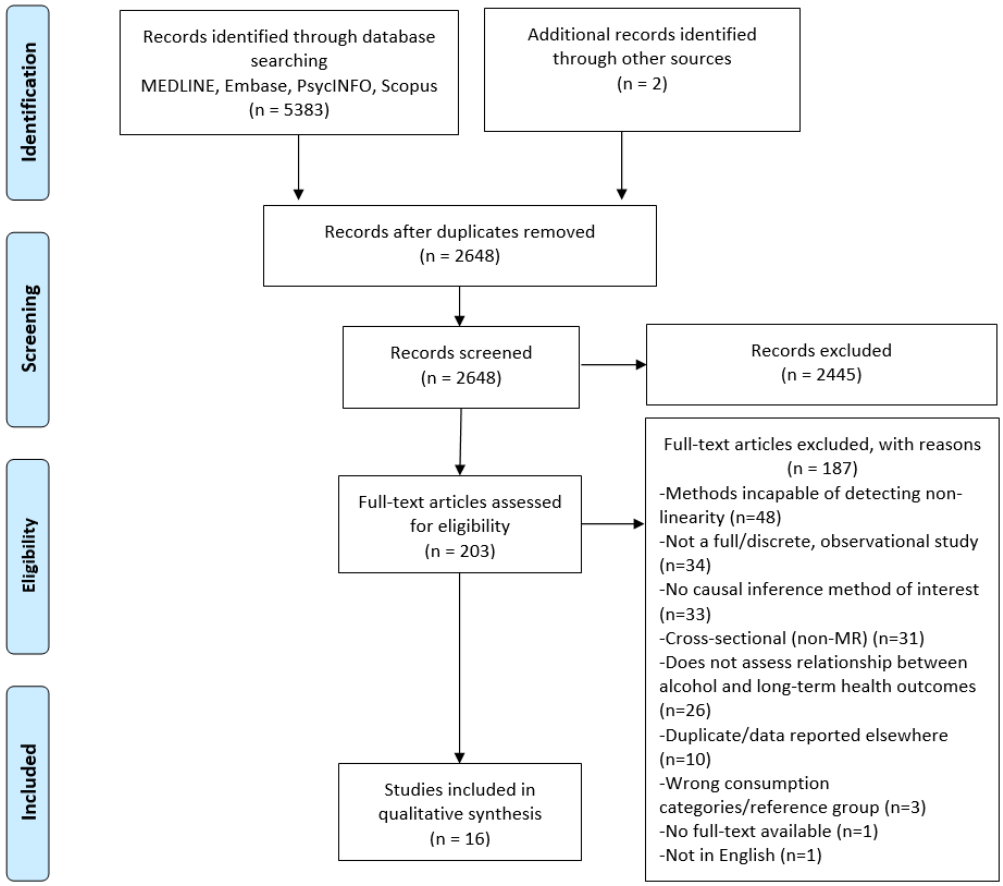
Data synthesis was limited to narrative description given the heterogeneity in health outcomes and methods employed by included studies. Reporting of this review complies with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA; Moher et al., 2009), the checklist for which can be found in Appendix 1.3.

2.3 Results

2.3.1 Characteristics of included studies

Sixteen articles met inclusion criteria (see Figure 2.1), comprising four MR studies, nine twin designs, and three prospective cohort studies employing MSMs. The studies reported on health outcomes broadly related to cancer, diabetes, dementia, mental health, cardiovascular health, mortality, HIV, and musculoskeletal health. Two cohorts provided all twin study data, with all but one non-MR study conducted with Swedish or Finnish populations. Study characteristics are summarised in Table 2.2, and exclusion reasons for key ineligible papers are provided in Appendix 1.4. The two reviewers were in agreement on title and abstract screening for 93.13% of cases, and all discrepancies at full-text screening were resolved by discussion between reviewers.

Figure 2.1. PRISMA flow chart.



2.3.2 Cancer

One twin study reported on two prostate cancer outcomes, prostate cancer and prostate cancer mortality, and results were consistent with reverse J- and J-shaped relationships, respectively, with light drinking at the nadir (Dickerman et al., 2016). Hazard ratios (HRs) for abstainers ranged from 2.85-2.98 (monozygotic [MZ] and combined twin analyses) compared to light drinkers, and for heavy drinkers compared to light drinkers ranged from 1.63-2.00. With the maximum sample (all twin analyses), the confidence interval (CI) for the abstainer comparison was large. Results were similar when restricting analyses to twins discordant for prostate cancer outcome.

2.3.3 Diabetes

Two studies reported on diabetes, including one twin study, and one MR study. Carlsson et al.'s (2003) twin-based findings on the risk for type 2 diabetes (T2D) resemble a J-shape but were critically underpowered, preventing interpretation. In Peng et al.'s (2019) MR study, local average treatment effects (LATEs) were employed to detect non-linearity – indicated by a LATE slope (effect of genetically-predicted alcohol consumption plotted against discretised observed alcohol consumption) significantly different to zero. Results did not support non-linearity for diabetes-related biomarkers. Substantive effects were only interpreted for men, with women used as a negative control due to their lack of alcohol consumption (effect of genetic instrument on diabetic markers should be observed in men but not women if alcohol consumption is the only causal pathway). In linear IV analyses, small positive linear relationships with narrow CIs were found for fasting blood glucose (FBG), 2-hour post-load plasma glucose (P2hBG) and insulin resistance (HOMA-IR), while there were no relationships for haemoglobin A1c (HbA1c) or beta-cell function (HOMA-beta).

2.3.4 Dementia

One twin study reported on dementia (Handing et al., 2015). Analyses were consistent with a J-shape, with HRs for abstainers compared to light drinkers between 1.37-1.39, and for moderate-to-very-heavy drinkers compared to light drinkers between 1.57-3.07. Analyses of dementia-concordant twins demonstrated that only very heavy alcohol consumers had a much earlier age of onset than their light-drinking co-twins (10.67-year discrepancy in diagnosis compared to 6.79 years when both twins were light drinkers).

2.3.5 Mental health

Two articles reported on mental health outcomes: a prospective cohort study of depression employing MSMs, and a twin study assessing disability pension due to mental health diagnoses (MHD). Gémes et al. (2019) employed DAG-informed MSMs incorporating inverse probability of exposure and attrition weights. Results support a U-shaped relationship (although there were few excessive consumers), with relative risks (RR) of 1.60 and 1.77 for abstainers and excessive drinkers respectively compared to light drinkers. Analyses excluding those with baseline depression found increased risk for excessive consumers such that the form approximated a J-shape, with RRs of 1.46 and 2.83 for abstainers and excessive drinkers respectively. Stratifying by gender, only abstainers were at increased risk for men, while both abstainers and excessive consumers remained at increased risk for women.

Samuelsson et al. (2013) discretised alcohol consumption into categories based on both volume and frequency. In analyses of outcome-discordant twins, abstainers were at increased risk for pension due to MHD compared to light frequent consumers (HRs of 1.93-2.17 for MZ and combined twin analyses), as were heavy infrequent consumers with an HR of 2.10 (disappeared in MZ-only analyses; HR of 1.09), and light infrequent consumers (HRs of 2.80 and 3.67 for MZ and combined twin analyses respectively). Heavy frequent consumers were at decreased risk, although there were too few pairs to conduct MZ-specific analyses.

2.3.6 Cardiovascular events/diagnoses

Four studies reported on cardiovascular events/diagnoses, including two twin studies, one MSM and one MR study. Ilomäki et al. (2012) reported on myocardial infarction (MI), comparing various crude, adjusted, and MSM models. The DAG-informed MSM

incorporated time-varying consumption and both time-varying and invariant covariates. Results were consistent with a J-shape, with RRs of 1.27 for the lowest group and 1.59 for the highest group respectively, both with fairly narrow CIs (although they included the null). Models 2, 3, and 4 (non-MSM but with assorted incorporation of time-varying exposures/confounders) were consistent with monotonically increasing, reverse-J/U, and J-shaped relationships respectively.

Kadlecová et al. (2015) examined the relationship between midlife alcohol consumption and later stroke/transient ischaemic attack (TIA) using MZ twins. In co-twin analyses, all groups had higher odds ratios (ORs) for stroke/TIA than very light consumers, with the highest estimate for abstainers (OR of 2.22; consistent with a reverse J-shape). In twins concordant for stroke/TIA, heavy consumers had shorter time to event (5.68 years), while all other groups had slightly longer time to event than the very light drinking group.

Milwood et al. (2019) also reported on stroke (ischaemic, intracerebral haemorrhage, and total stroke), in addition to acute myocardial infarction (AMI), and total coronary heart disease (CHD). As for Peng et al., women were negative controls. Comparing categories of genetically-predicted alcohol consumption, MR results were consistent with monotonically increasing relationships for stroke and subtypes, and with no causal relationships for AMI/CHD. Log RRs for those relationships with evidence of linearity ranged from 1.27-1.58 per 280 grams of alcohol per week consumed.

Finally, Ropponen & Svedberg (2013) et al. reported on disability pension due to circulatory system diagnoses using a twin design. For same-sex twins discordant for outcome, there appeared to be little clear relationship in MZ-only analyses, while the dizygotic (DZ)-only

analyses were consistent with a reverse J-shape (both moderate and heavy consumers had HRs <1 compared to abstainers).

2.3.7 Continuous cardiovascular measures

Three MR studies reported on lipids, with two of these also reporting on blood pressure and obesity anthropometrics. Peng et al., using the same methods as for diabetes outcomes, found no evidence of non-linear relationships for any lipids, blood pressure measures or obesity anthropometrics, but did find positive linear relationships for BMI, waist circumference, hip circumference, non-HDL-C, triglycerides (TG), total cholesterol (TC), systolic blood pressure (SBP), and diastolic blood pressure (DBP).

Silverwood et al. (2014) applied the LATE method to pooled data from 22 studies to examine cardiovascular and inflammatory measures, finding non-linearity (J-shapes) for SBP, non-HDL-C, BMI, WC, and C-reactive protein (CRP). Nadirs for these relationships corresponded to small volumes of alcohol, ranging from 1-3.5 units of alcohol/week, and the differences in biomarker outcomes at the nadir compared with abstinence were also small. For those outcomes with no evidence of non-linearity, standard IV analysis revealed a positive linear relationship between alcohol consumption and IL-6 (an inflammatory marker), but a lack of relationship with HDL-C and triglycerides.

Finally, Vu et al. (2016) discretised genetically-predicted alcohol consumption into quartiles, comparing lipids in each with the lowest quartile. Results provide evidence of non-linearity for TG, TC, HDL2-C, LDL-C, sdLDL-C, and apoB. For these outcomes, all quartiles had more favourable levels than quartile 1. Benefits peaked at quartile 3, equivalent to .5-1

genetically-predicted units per week. Results do not support causal relationships between alcohol and HDL-C overall, HDL3-C, or Lp(a).

2.3.8 Mortality

One twin study assessed all-cause mortality, finding the three heaviest alcohol consuming groups had greater mortality risk compared to their lighter consuming reference, with HRs ranging from 1.60-2.99 (Sipilä et al., 2016). This pattern replicated in the MZ-only sample, but with less precision and with CIs crossing the null. Abstainers were at decreased risk (HR of .43) in the MZ-only sample, but the CI included the null.

2.3.9 HIV seroconversion

One study employing MSMs reported on HIV seroconversion in men, incorporating both inverse probability of exposure and censoring weights (Sander et al., 2013). Results were consistent with a monotonically increasing risk function, with an RR of 1.61 for heavy drinkers compared with abstainers.

2.3.10 Musculoskeletal health (MSD)

Three twin studies reported on MSD – all using receipt of disability pension due to MSD conditions as the outcome. Pietkäinen et al. (2011) found a roughly monotonically increasing relationship for pension due to lower back disorders, with lower risk for abstainers the clearest effect (HR of .79), although all CIs included the null. When stratified by sex, the protective effect of abstinence was more pronounced in men (HR .45; CI .13,1.48), while the functional form in women changed such that there was also reduced risk for moderate consumers (HR .76; CI .45,2.32).

Using the same cohort, Ropponen et al. (2011) examined disability pension due to osteoarthritis and due to MSD more generally. Results were not consistent with a clear functional form but do support abstainers having the lowest risk for both outcomes.

Finally, in a different cohort, Ropponen & Svedberg evaluated risk for disability pension due to MSD. Again, outcome-discordant twin analyses did not reveal a clear functional form, with discrepant results between MZ and DZ samples.

2.3.11 Risk of bias

Cohort studies ranged in scores from 7-9 out of 9 on the NOS tool (see Appendix 1.5), with most losing marks for self-reported ascertainment of exposure. MR studies all had a combination of low and moderate risk across the five domains (see Appendix 1.6).

Table 2.2. Characteristics of included papers (n=16)

<table>
<tr><th>Study</th><th>Health outcome/s and interval to follow-up</th><th>Study design</th><th>Participant characteristics</th><th>Cohort/s name and sample size</th><th>Statistical methodology
(reference group bolded)</th><th>Explicit testing for non-linearity?</th><th>Main results
(statistically significant findings bolded)</th><th>Conclusion</th></tr>
<tr><td colspan="9">Cancer</td></tr>
<tr><td>(Dickerman et al., 2016)</td><td>Prostate cancer (PC) and prostate cancer mortality
-median 30 yrs of follow-up</td><td>Twin</td><td>Male twins; mean age 40.1 at baseline</td><td>Older Finnish Twin Cohort
11,372; 225 and 43 discordant twin pairs for PC and PC-mortality respectively</td><td>-Co-twin (discordant for both alcohol consumption level and either time to diagnosis among outcome-concordant pairs concordant or time to event vs death/end of follow-up among outcome-discordant pairs) and pooled cohort Cox analyses to examine risk for PC and PC-mortality
-Alcohol consumption measured twice (6 years apart) and averaged; categories: abstainers, light (0.01–3 drinks/wk), moderate (>3–14 drinks/wk), heavy (>14 drinks/wk); 1 drink considered = 12g of alcohol</td><td>×</td><td>-PC-risk:
<table><tr><th>HR (95% CI)</th><th>Cohort</th><th>MZ twins</th><th>DZ twins</th><th>All twins</th></tr><tr><td>Abstainer</td><td>1.27 (.94,1.71)</td><td>2.85 (.67,12.1)</td><td>3.80 (1.36,20.6)</td><td>2.98 (1.35,6.60)</td></tr><tr><td>Moderate</td><td>1.2 (.99,1.46)</td><td>1.28 (.6,2.74)</td><td>1.54 (0.92,2.57)</td><td>1.36 (.91,2.04)</td></tr><tr><td>Heavy</td><td>1.46 (1.12,1.91)</td><td>2.00(.62,6.45)</td><td>1.71 (.87,3.39)</td><td>1.63 (.92,2.88)</td></tr></table>-PC-mortality:
<table><tr><th>HR (95% CI)</th><th>Cohort</th><th>MZ twins</th><th>DZ twins</th><th>All twins</th></tr><tr><td>Abstainer</td><td>1.90 (1.04,3.47)</td><td>2.31 (.19, 27.4)</td><td>1.83 (.42,8.03)</td><td>1.37 (.44,4.28)</td></tr><tr><td>Moderate</td><td>1.22 (.76,1.97)</td><td>9.13 (.70,119)</td><td>1.43 (.37,5.62)</td><td>2.44 (.79, 7.52)</td></tr><tr><td>Heavy</td><td>1.32 (.66,2.62)</td><td>-</td><td>2.39 (.33,17.3)</td><td>7.31 (1.3, 41)</td></tr></table></td><td>Reverse J-shaped relationships, with light drinking at the nadir, were evident in all twin analyses of PC risk. J-shaped curves were evident in twin analyses of PC mortality. The evidence is consistent with a causal protective effect of light drinking, but twin analyses in general lacked power.</td></tr>
</table>

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Diabetes								
(Carlsson et al., 2003)	Type 2 diabetes (T2D) -20 yrs of follow-up	Twin	Same-sex twins without diabetes at baseline; mean age 34.3 (M) and 35.4 (F) at baseline	Finnish Twin Cohort 22,788; 27 discordant twin pairs analysed	-Co-twin (discordant for alcohol category) and pooled cohort Cox analyses to examine risk for T2D -Categories for pooled cohort (based on 3 questionnaires over 15 years): abstainers, **<5g/day**, 5-19.9g/day (women)/5-29.9g/day (men), ≥ 20g/day (women)/ ≥ 30g/day (men); for twin analyses (based on baseline questionnaire only): **low** (<5g/day), moderate (15-29.9g/day), high (≥ 30g/day)	✓	RR (95% CI) / Cohort Men / Cohort Women / OR (95% CI) / Twin Abstainer / 1.1 (.7,1.5) / 1.1 (.9,1.5) / Moderate / .5 (.2,1.3) 5-29.9g/day (m); 5-19.9g/day (f) / .8 (.6,1.1) / .7 (.4,1.1) / High / 1.2 (.4,3.9) ≥ 30g/day (m); ≥ 20g/day (f) / .9 (.6,1.4) / 1.6 (.8,3.5) / - / -	**Twin analyses were critically underpowered, preventing causal inference.**
(Peng et al., 2019) *See also cardiovascular outcomes section	Diabetes-related biomarkers (FBG, P2hBG, HbA1c, HOMA-IR, HOMA-beta) -cross-sectional	MR	Chinese adults living in the Yi-Ling district of Yichang; mean age 55 (SD 0.1) at baseline	One community from the Risk Evaluation of cAncers in Chinese diabeTic Individuals: a LONgitudinal (REACTION) study 4,536	-1-sample MR using a single variant and standard IV analysis (2SLS); local average treatment effects (LATEs) computed for subgroups of observed alcohol consumption (non-zero LATE slopes indicate non-linearity) -Only standard, linear IV analysis was performed for categorical diabetes risk, not allowing for detection of non-linearity	✓	-Non-linear analyses: the LATE slopes for all diabetes-related biomarkers were not sig. different to zero, indicating no non-linear relationships Unstandardised β (95% CI) / Using log-transformed alcohol intake FBG (log-transformed) / -.01 (-.04,.01) P2hBG (log-transformed) / .01 (-.04,.06) HbA1c (log-transformed) / -.01 (-.03,.00) HOMA-IR (log-transformed) / -.00 (-.09,.08) HOMA-beta (log-transformed) / .03 (-.04, .11)	**No evidence of non-linear relationships between genetically predicted alcohol consumption and diabetes-related biomarkers. Alcohol appears to raise FBG, P2hBG and HOMA-IR in a linear fashion in**

<table>
<tr><th>Study</th><th>Health outcome/s and interval to follow-up</th><th>Study design</th><th>Participant characteristics</th><th>Cohort/s name and sample size</th><th>Statistical methodology
(reference group bolded)</th><th>Explicit testing for non-linearity?</th><th>Main results
(statistically significant findings bolded)</th><th>Conclusion</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>-Analysis performed in women as a negative control due to their lack of alcohol consumption

-No conventional analyses conducted for comparison</td><td></td><td>-Standard IV analyses:<table><tr><th>Unstandardised β (95% CI)</th><th>Per 1-unit increase in log-transformed genetically-predicted alcohol consumption</th></tr><tr><td>FBG (log-transformed)</td><td>.04 (.02,.05)</td></tr><tr><td>P2hBG(log-transformed)</td><td>.07 (.04,.11)</td></tr><tr><td>HbA1c (log-transformed)</td><td>.00 (-.01,.01)</td></tr><tr><td>HOMA-IR(log-transformed)</td><td>.10 (.04, .17)</td></tr><tr><td>HOMA-beta (log-transformed)</td><td>.00 (-.05,.06)</td></tr></table></td><td>males, although these are small effects.</td></tr>
<tr><td>Dementia</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>(Handing et al., 2015)</td><td>Dementia

-43 yrs of follow-up</td><td>Twin</td><td>Twins without dementia prior to baseline; <=65 at baseline and >=60 at study end; mean age 54.2 (SD 5.9) at baseline; reported drinking <=100g/day of alcohol</td><td>Swedish Twin Registry

12,362; 576 dementia discordant pairs (177 MZ); 396 concordant pairs (160 MZ)</td><td>-Co-twin (discordant for dementia) logistic regressions and pooled cohort Cox models used to examine risk for dementia; co-twin (concordant for dementia) mixed-effects analyses used to examine age at onset

-Categories for pooled cohort + twin age of onset analyses: abstainers, light (>0 and ≤5g/d), moderate (>5 and ≤12g/d), heavy (>12 and ≤ 24g/d), and very heavy (>24g/d). For twin dementia risk analyses: abstainers, light (>0 and ≤5g/d), and moderate-to-very heavy (>5g/d)</td><td>✓</td><td>-Dementia risk:<table><tr><th>HR (95% CI)</th><th>Cohort</th><th>OR (95% CI)</th><th>MZ twins</th><th>All twins</th></tr><tr><td>Abstainer</td><td>1.05 (1.00,1.11)</td><td>Abstainer</td><td>1.37 (.6,3.16)</td><td>1.39 (.89,2.16)</td></tr><tr><td>Moderate</td><td>.98 (.92-1.04)</td><td>Moderate-to-very heavy</td><td>3.07 (1.37,6.86)</td><td>1.57 (1.04,2.37)</td></tr><tr><td>Heavy</td><td>1.1 (1.01,1.19)</td><td></td><td></td><td></td></tr><tr><td>Very heavy</td><td>1.18 (1.01,1.36)</td><td></td><td></td><td></td></tr></table></td><td>Results are consistent with a J-shaped curve. Twin analyses do support a causal role for increased risk of dementia for moderate-to-very heavy drinking, and faster dementia onset for very heavy drinkers.</td></tr>
</table>

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
							-Dementia age at onset: Mean difference in years between twin diagnosed first vs twin diagnosed later (*p* value) / MZ twins / All twins Abstainer: -5.37 (.83) / -5.49 (.68) Light: -6.28 (NA – reference group) / -6.79 (NA – reference group) Moderate: -5.41 (.13) / -7.00 (.98) Heavy: -6.33 (.99) / -6.73 (.32) Very heavy: -12.67 (.09) / **-10.67** (.02)	
Mental health								
(Gémes et al., 2019)	Depression -7-9 yrs of follow-up	MSM	General population aged 20-64 at recruitment; mean age 43.3 (SD 12.2) at baseline	Psykisk Hälsa–Arbete–Relationer (PART) cohort (Sweden) 5,087	-MSM (weighted logistic regression) + standard logistic regression for comparison -Alcohol consumption measured pre-baseline (for use as time-variant confounder) and baseline; categories: no consumption, **light** (>0-≤7 drinks/wk), moderate (>7-≤14 drinks/wk), and excessive (>14 drinks/wk); 1 drink=12g	✓	-Adjusting for baseline MDI score: RR (95% CI) / Non-MSM / MSM Abstainer: 1.10 (.69,1.74) / **1.60** (1.27,2.01) Moderate: .54 (.28,1.04) / 1.05 (.83,1.32) Excessive: .61 (.21,3.07) / **1.77** (1.13,2.78) -Excluding those with baseline depression (MDI>26): RR (95% CI) / Non-MSM / MSM Abstainer: 1.24 (.72,2.14) / **1.46** (1.10,1.95) Moderate: .63 (.29,1.41) / .75 (.55,1.03) Excessive: 2.72 (.61,12.19) / **2.83** (1.80,4.44)	**MSM results support a non-linear (U/J-shaped) relationship between alcohol consumption and depression.**
(Samuelsson et al., 2013)[a]	Disability pension (DP) due to mental health diagnoses (MHD) -median 10 yrs of follow-	Twin	Twins with data from a prior study, and at time of that study were living in Sweden, <65, and without DP/old age	Swedish Twin Study of Disability Pension and Sickness Absence (STODS), drawn from the Swedish Twin Registry	-Co-twin (discordant for dementia) Cox regressions + pooled cohort Cox regressions performed -Differentiated between frequent and infrequent drinkers (have/not consumed alcohol in previous two	✓	HR (95% CI) / Cohort / MZ twins / DZ twins / All twins Abstainer: **1.99** (1.57,2.54) / 1.93 (.54,6.96) / **2.63** (1.05, 6.62) / **2.17** (1.06,4.45) Moderate frequent: 1.07 (.78,1.49) / .46 (.11,1.87) / 2.70 (.81-9.02) / 1.24 (.53,2.91) Heavy frequent: .98 (.61,1.54) / - / .46 (.09,2.36) / .48 (.12,1.90)	**Results support a non-linear relationship between alcohol consumption and DP due to MHD, such that abstainers are at**

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
	up from prior study (baseline)		pension; mean age 52.9 (SD 5.6)	28,613; 229 DP-discordant twin pairs (95 MZ pairs)	months); categories for pooled cohort + twin analyses: abstainers, **light frequent** (12-84g/wk), moderate frequent (M: 85-168g/wk; F: 85-108g/wk) and ≤12g/d), heavy frequent (M: >168g/wk; F: >108g/wk), light infrequent (12g/occasion), moderate infrequent (M:13-48g/occasion; F: 13-36g/occasion), heavy infrequent (M: >48g/occasion; F:>36g/occasion)		Light infrequent 1.09 (.69,1.73) 2.80 (.24,32.6) 3.98 (.71,22.4) 3.67 (.91,14.8) Moderate infrequent 1.18 (.91,1.54) .52 (.18,1.49) 1.71 (.75,3.91) 1.03 (.55,1.93) Heavy infrequent 1.20 (.92,1.57) 1.09 (.33,3.53) **3.61** (1.28, 10.2) 2.10 (.99,4.46)	**increased risk compared to light frequent drinkers. Light and heavy infrequent drinkers are also at increased risk, but CIs are very wide for the former, and effect disappears for the latter in MZ-only twins.**
Cardiovascular events/diagnoses								
(Ilomäki et al., 2012)	Myocardial infarction (MI) -12-14 yrs of follow-up	MSM	General population males; mean age 52 (SD 6.7) at initial exam (4 yrs before 'baseline')	Kuopio Ischaemic Heart Disease Risk Factor Study (KIHD); Finland 1,030	-5 discrete-time hazard models run and compared: M1 (baseline alcohol consumption with no covariate adjustment or inclusion of t-4 consumption), M2 (baseline alcohol consumption adjusted for covariates and t-4 consumption), M3 (baseline and t+7 alcohol consumption with no covariate adjustment or inclusion of t-4 consumption), M4 (baseline and t+7 alcohol consumption adjusted for covariates and t-4 consumption), M5 (MSM with baseline and t+7 alcohol	✓	RR (95% CI) M1 M2 M3 M4 M5 (MSM) <12g/wk 1.20 (.86,1.67) 1.01 (.70,1.45) **1.55** (1.08, 2.21) 1.27 (.88,1.81) 1.27 (.88,1.83) 84-167g/wk 1.05 (.71,1.56) 1.13 (.75,1.72) 1.20 (.78,1.84) 1.27 (.81,2.00) 1.18 (.75,1.87) ≥168g/wk .98 (.60,1.58) 1.20 (.68,2.12) 1.40 (.91,2.18) **1.71** (1.03, 2.85) 1.59 (.93,2.72)	**Results are generally consistent with increased risk for MI for those drinking less than weekly and those drinking heavily. However, effect sizes and significance varied with model specifications.**

<table>
<tr><th>Study</th><th>Health outcome/s and interval to follow-up</th><th>Study design</th><th>Participant characteristics</th><th>Cohort/s name and sample size</th><th>Statistical methodology
(reference group bolded)</th><th>Explicit testing for non-linearity?</th><th>Main results
(statistically significant findings bolded)</th><th>Conclusion</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>consumption with stabilised IP weights at t and t+7)
-Baseline alcohol categories: <12g/wk, 12-83g/wk, 84-167g/wk, ≥168g/wk; additional alcohol measurements at t-4 and t+7 included in various models</td><td></td><td></td><td></td></tr>
<tr><td>(Kadlecová et al., 2015)</td><td>Stroke and transient ischaemic attack (TIA) events
-43 yrs of follow-up</td><td>Twin</td><td>Same-sex twins ≤60 with no history of stroke at baseline and with ≥5 yrs of follow-up; mean age 50.5 (SD 5.29) at baseline</td><td>Swedish Twin Registry
11,644; 370 stroke/TIA discordant pairs (all MZ); 167 stroke/TIA concordant pairs (all MZ)</td><td>-Co-twin (discordant for stroke) logistic regressions and pooled cohort Cox models used to examine risk for stroke/TIA; co-twin (concordant for stroke) mixed-effects analyses used to examine time to stroke/TIA
-Categories for pooled cohort + twin analyses: abstainers, very light (>0-5g/day), light (>5-12g/day), moderate (>12-24g/day), heavy (>24g/day)</td><td>✓</td><td>-Stroke/TIA risk:<table><tr><th>HR (95% CI)</th><th>Cohort</th><th>OR (p value)</th><th>Twins</th></tr><tr><td>Abstainer</td><td>1.11 (.98,1.23)</td><td>Abstainers</td><td>2.22 (.058)</td></tr><tr><td>Light</td><td>.98 (.85,1.15)</td><td>Light</td><td>1.56 (.17)</td></tr><tr><td>Moderate</td><td>.99 (.85,1.15)</td><td>Moderate</td><td>1.59 (.26)</td></tr><tr><td>Heavy</td><td>1.34 (1.04,1.70)</td><td>Heavy</td><td>1.23 (.78)</td></tr></table>-Time to stroke/TIA:<table><tr><th>Mean difference in yrs to stroke/TIA (p value)</th><th>Twins</th></tr><tr><td>Abstainer</td><td>2.07 (.11)</td></tr><tr><td>Light</td><td>.77 (.54)</td></tr><tr><td>Moderate</td><td>1.15 (.47)</td></tr><tr><td>Heavy</td><td>-5.68 (.029)</td></tr></table></td><td>Twin analyses are consistent with an increased risk of stroke/TIA for abstainers compared to very light drinkers, with somewhat increased risks for heavier drinking groups as well (reverse J-shape). Results support a causal role of heavy alcohol consumption in hastening time to event among those who</td></tr>
</table>

<table>
<tr><th>Study</th><th>Health outcome/s and interval to follow-up</th><th>Study design</th><th>Participant characteristics</th><th>Cohort/s name and sample size</th><th>Statistical methodology
(reference group bolded)</th><th>Explicit testing for non-linearity?</th><th>Main results
(statistically significant findings bolded)</th><th>Conclusion</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>experience stroke/TIA.</td></tr>
<tr><td>(Millwood et al., 2019)</td><td>Ischaemic stroke, intra-cerebral haemorrhage (ICH), total stroke, acute myocardial infarction (AMI), total coronary heart disease (CHD)

-roughly 10 yrs of follow-up</td><td>MR</td><td>Permanent residents from 10 Chinese regions aged roughly 35-74 and without major disabilities , (those with a history of CVD were excluded from analyses of disease incidence); mean age 52 (SD 11) at baseline</td><td>China Kadoorie Biobank

512,715; 161,498 of which had genotype data (male and female combined)</td><td>-1-sample MR using Cox models

-Instrument composed of 9 combinations of 2 SNPs (ALDH2 rs671 and ADH1B rs1229984) within each of 10 geographic areas, producing 90 combinations overall; then, based on mean ‘usual’ alcohol consumption within each of those combinations (incorporating repeat measurements to account for measurement error), six final categories were produced, aligned with increasing genetically-predicted alcohol consumption: Category1: 0-10g/wk, Category2: 10-25g/wk, Category 3: 25-50g/wk, Category 4: 50-100g/wk, Category 5 100-150g/wk, Category 6 >150g/wk

-Conventional analyses using observed alcohol consumption conducted for comparison using Cox models, with consumption categories (for</td><td>✓</td><td>-Ischaemic stroke:
<table><tr><th>Log RR (95% CI)</th><th>MR</th><th></th><th>Conventional</th></tr><tr><td>C2</td><td>1.00 (.91,1.10)</td><td>Ex-drinker</td><td>1.39 (1.32,1.46)</td></tr><tr><td>C3</td><td>1.03 (.96,1.11)</td><td>Non-drinker</td><td>1.21 (1.16,1.25)</td></tr><tr><td>C4</td><td>1.11 (1.02,1.20)</td><td>Occasional</td><td>1.00 (.97,1.03)</td></tr><tr><td>C5</td><td>1.23 (1.15,1.33)</td><td>140-279g/wk</td><td>1.13 (1.07,1.19)</td></tr><tr><td>C6</td><td>1.23 (1.12,1.35)</td><td>280-419g/wk</td><td>1.23 (1.15, 1.32)</td></tr><tr><td></td><td></td><td>≥420g/wk</td><td>1.31 (1.21,1.41)</td></tr><tr><td colspan="4">Per 280g/wk (assuming linearity)</td></tr><tr><td></td><td>1.27 (1.13,1.43)</td><td></td><td>1.28 (1.19,1.38)</td></tr></table>-ICH:
<table><tr><th>Log RR (95% CI)</th><th>MR</th><th></th><th>Conventional</th></tr><tr><td>C2</td><td>1.01 (.88,1.14)</td><td>Ex-drinker</td><td>1.52 (1.38, 1.68)</td></tr><tr><td>C3</td><td>1.02 (.88,1.14)</td><td>Non-drinker</td><td>1.33 (1.25,1.43)</td></tr><tr><td>C4</td><td>1.08 (.96,1.22)</td><td>Occasional</td><td>1.02 (.96,1.09)</td></tr><tr><td>C5</td><td>1.29 (1.15,1.44)</td><td>140-279g/wk</td><td>1.35 (1.21,1.52)</td></tr><tr><td>C6</td><td>1.54 (1.36,1.76)</td><td>280-419g/wk</td><td>1.52 (1.32,1.74)</td></tr><tr><td></td><td></td><td>≥420g/wk</td><td>1.73 (1.52,1.97)</td></tr><tr><td colspan="4">Per 280g/wk (assuming linearity)</td></tr><tr><td></td><td>1.58 (1.36,1.84)</td><td></td><td>1.59 (1.37,1.85)</td></tr></table></td><td>In contrast to conventional analyses, MR results are consistent with a monotonically increasing relationship between alcohol and stroke events. The results do not support a causal relationship between alcohol consumption and AMI or total CHD.</td></tr>
</table>

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
					men): ex-drinker, non-drinker, occasional drinker (less than weekly), <**140g/wk**, 140-279g/wk, 280-419g/wk, ≥420g/wk		-Total stroke: Log RR (95% CI) / MR / Ex-drinker / Conventional C2 / 1.02 (.95,1.10) / Ex-drinker / **1.40** (1.34,1.46) C3 / 1.05 (.95,1.10) / Non-drinker / **1.23** (1.19,1.27) C4 / **1.13** (1.06,2.10) / Occasional / 1.00 (.98, 1.03) C5 / **1.27** (1.19,1.34) / 140-279g/wk / **1.16** (1.10,1.21) C6 / **1.35** (1.26,1.45) / 280-419g/wk / **1.28** (1.20,1.36) ≥420g/wk / **1.39** (1.31,1.48) Per 280g/wk (assuming linearity) **1.38** (1.26,1.51) / **1.35** (1.27,1.44) -AMI: Log RR (95% CI) / MR / Ex-drinker / Conventional C2 / 1.02 (.87,1.19) / Ex-drinker / **1.66** (1.48,1.85) C3 / 1.05 (.93,1.19) / Non-drinker / **1.63** (1.51,1.76) C4 / .93 (.81,1.07) / Occasional / **1.23** (1.16,1.31) C5 / .94 9.81,1.09) / 140-279g/wk / 1.11 (.97,1.26) C6 / .97 (.83,1.15) / 280-419g/wk / **1.19** (1.01,1.41) ≥420g/wk / 1.14 (.95,1.37)	

<table>
<tr><th>Study</th><th>Health outcome/s and interval to follow-up</th><th>Study design</th><th>Participant characteristics</th><th>Cohort/s name and sample size</th><th>Statistical methodology
(reference group bolded)</th><th>Explicit testing for non-linearity?</th><th>Main results
(statistically significant findings bolded)</th><th>Conclusion</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-Total CHD:<table><tr><th>Log RR (95% CI)</th><th>MR</th><th></th><th>Conventional</th></tr><tr><td>C2</td><td>1.03 (.93,1.13)</td><td>Ex-drinker</td><td>1.45 (1.38,1.52)</td></tr><tr><td>C3</td><td>1.08 (1.01,1.15)</td><td>Non-drinker</td><td>1.31 (1.26,1.36)</td></tr><tr><td>C4</td><td>.94 (.87,1.02)</td><td>Occasional</td><td>1.07 (1.04,1.11)</td></tr><tr><td>C5</td><td>1.04 (.97,1.11)</td><td>140-279g/wk</td><td>1.03 (.97,1.09)</td></tr><tr><td>C6</td><td>1.06 (.98,1.15)</td><td>280-419g/wk</td><td>1.11 (1.03,1.19)</td></tr><tr><td></td><td></td><td>≥420g/wk</td><td>1.13 (1.04,1.22)</td></tr></table></td><td></td></tr>
<tr><td>(Ropponen & Svedberg, 2013)

*See also musculo-skeletal health section</td><td>DP due to circulatory system diagnoses

-5-10 years of follow-up</td><td>Twin</td><td>Twins with data from a prior study, and at time of that study were living in Sweden, <65, working and without DP/old age pension; mean age at baseline 53.7 (SD 5.7) a</td><td>Swedish Twin Registry

31,206; 216 DP due to circulatory system diagnoses-discordant pairs (of which 95 are MZ)</td><td>-Co-twin (discordant for DP due to MSD) Cox models and pooled cohort Cox models used to examine risk (stratified for sex in pooled model)

-Categories for pooled cohort + twin analyses: abstainers, light (≤3 drinks/wk), moderate (F: >3-≤7 drinks/wk; M: >3-≤14 drinks/wk), heavy (F: >7drinks/wk; M: >14 drinks/wk)</td><td>✓</td><td><table><tr><th>HR (95% CI)</th><th>Cohort</th><th>MZ twins</th><th>DZ twins</th><th>All twins</th></tr><tr><td>Abstainer</td><td>1.22 (.91,1.64)</td><td>.97 (.31,3.01)</td><td>1.55 (.70,3.44)</td><td>1.3 (.69,2.45)</td></tr><tr><td>Moderate</td><td>.85 (.63,1.15)</td><td>-</td><td>.86 (.39,1.91)</td><td>1.54 (.78,3.04)</td></tr><tr><td>Heavy</td><td>.79 (.63,.98)</td><td>.91 (.49,1.68)</td><td>.89 (.52,1.51)</td><td>.86 (.57,1.28)</td></tr></table></td><td>Results do not offer clear support for causal relationships between alcohol consumption and later receipt of DP due to circulatory system diagnoses.</td></tr>
</table>

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Cardiovascular disease biomarkers								
(Peng et al., 2019) *See also diabetic outcomes section	Lipids: HDL-C, non-HDL-C, tri-glycerides (TG), total cholesterol (TC); blood pressure: systolic (SBP), diastolic (DBP); obesity anthro-pometric measures: BMI, waist circum-ference (WC), hip circum-ference (HC), waist-to-hip ratio (WHR) -cross-sectional	MR	Chinese adults living in the Yi-Ling district of Yichang; mean age 55 (SD 0.1) at baseline	One community selected from the Risk Evaluation of cAncers in Chinese diabeTic Individuals: a LONgitudinal (REACTION) study 4,536	-1-sample MR using ALDH2 to instrument for alcohol consumption with standard IV analysis (2SLS); local average treatment effects (LATEs) computed for subgroups of observed alcohol consumption (non-zero LATE slopes indicate non-linearity) -Analysis performed in women as a negative control due to their lack of alcohol consumption -No conventional analyses conducted for comparison	✓	-Non-linear analyses: the LATE slopes for all lipids, blood pressure and obesity parameters in both sexes were not sig. different to zero, indicating no presence of non-linear relationships Unstandardised β (95% CI) Using log-transformed alcohol intake HDL-C -.01 (-.06,.04) Non-HDL-C -.01 (-.13,.10) TG (log-transformed) .03 (-.05,.10) TC -.03 (-.16,.08) SBP -.85 (-3.15,1.3) DBP -.70 (-2.32,.71) BMI -.15 (-.51,.21) WC -.02 (-1.09,1.09) HC .34 (-.37,1.07) WHR -.00 (-.01,.00) -Standard IV analyses: Unstandardised β (95% CI) Per 1-unit increase in log-transformed genetically-predicted alcohol consumption HDL-C .04 (-.00,.08) Non-HDL-C **.11** (.02,.20) TG (log-transformed) **.11** (.05,.16) TC **.15** (.06,.25) SBP **2.91** (1.06,4.76) DBP **3.03** (1.87,4.19) BMI **.57** (.28,.87) WC **2.37** (1.47,3.27) HC **1.01** (.39,1.62) WHR .02 (.00,.02)	**LATE results offer no evidence of non-linear relationships between genetically-predicted alcohol consumption and lipid markers, blood pressure or obesity parameters. Alcohol appears to raise BMI, WC, HC, non-HDL-C, TG, TC, SBP and DBP in a linear fashion in males, with little effect on HDL-C or WHR.**

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
(Silverwood et al., 2014)	Lipids: non-HDL-C, HDL-C, TG; blood pressure: SBP; obesity anthro-pometric measures: BMI, WC; inflamm-atory markers: CRP, interleukin 6 (IL-6) -cross-sectional	MR	Individuals of European descent from Europe and North America; mean age 56.75 (calculated from Holmes et al. [cite])	22 individual studies (18 cohorts, 2 nested case-control, 1 RCT, 1 case-control) from the Alcohol-ADH1B consortium 80,057 individuals total; 78,172 for SBP, 60,140 for non-HDL-C, 60,227 for HDL-C, 79,454 for BMI, 57,172 for WC, 63,367 for CRP, 23,535 for IL-6, 63,667 for TG	-1-sample MR using the rs1229984 polymorphism in ADH1B to instrument for alcohol consumption and standard IV analysis (2SLS); local average treatment effects (LATEs) computed for subgroups of observed alcohol consumption (non-zero LATE slopes indicate non-linearity) -Where non-linearity is present, the difference in outcome between no alcohol and median observed consumption in low (>0–7 units/wk), moderate (7–21 units/wk), heavy (21–70 units/wk) and very heavy (>70 units/wk) groups is predicted, as well as curve nadir, difference in outcome between nadir and abstinence, and level of consumption matching outcome for abstinence -No conventional analyses conducted for comparison	✓	-Non-linear analyses: the LATE slopes for SBP, non-HDL-C, BMI, WC and CRP were all sig. different to zero, indicating non-linearity Unstandardised β (95% CI) / Using log-transformed alcohol intake HDL-C: .00 (-.06,.06) Non-HDL-C: **.37** (.19,.55) TG (log-transformed): -.02 (-.10,.06) SBP: **3.30** (1.0,5.5) BMI: **.90** (.3,1.4) WC: **2.00** (.6,3.6) CRP (log-transformed): **.26** (.10,.43) IL-6 (log-transformed): -.13 (-.34,.29) -Predicted differences in outcome between category medians and abstinence (unstandardised): 3.04 units/wk / 12.15 units/wk / 31.90 units/wk / 84.52 units/wk Non-HDL-C: -.39 (-.79,.06) / -.15 (-.72,.47) / .40 (-.28,1.10) / **1.30** (.45,2.16) SBP: .1 (-5.5,6.1) / 5.2 (-2.6,13.9) / **12.4** (3.4,22.1) / **22.8** (12.2,34.6) BMI: -.6 (-2.2,.8) / .2 (-2.0,2.1) / 1.6 (-.8,3.8) / **3.9** (1.2,6.3) WC: -.6 (-4.7,3.5) / 1.9 (-3.9,7.8) / 5.7(-.6,12.5) / **11.5** (4.5,12.5) CRP (log-transformed): -.29 (-.68,.15) / -.15 (-.68,.5) / .22 (-.37,.95) / **.83** (.15,1.69)	**Results support non-linear relationships between alcohol consumption and SBP, non-HDL-C, BMI, WC and CRP, with nadirs for these outcomes falling in the low drinking range. These relationships are best characterised by J-shapes with shallow nadirs, followed by gentle, elongated inclines. Results are consistent with a positive linear relationship between alcohol and IL-6, and do not support any**

<table>
<tr><th>Study</th><th>Health outcome/s and interval to follow-up</th><th>Study design</th><th>Participant characteristics</th><th>Cohort/s name and sample size</th><th>Statistical methodology
(reference group bolded)</th><th>Explicit testing for non-linearity?</th><th>Main results
(statistically significant findings bolded)</th><th>Conclusion</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-Predicted curve features for those outcomes with evidence of non-linearity (unstandardised):
<table><tr><th></th><th>Nadir (units/wk; 95% CI)</th><th>Difference in outcome at nadir</th><th>Units/wk (95% CI) with outcome equivalent to abstinence</th></tr><tr><td>Non-HDL-C</td><td>3.2 (.7,6.0)</td><td>-.39 (-.85,-.03)</td><td>16.9 (2.1, 48.2)</td></tr><tr><td>SBP</td><td>1.00 (.0,3.6)</td><td>-.7(-5.4,.0)</td><td>2.8 (.0,19.6)</td></tr><tr><td>BMI</td><td>2.3 (.0,6.0)</td><td>-.6 (-2.3,.0)</td><td>10.1 (.0,48.4)</td></tr><tr><td>WC</td><td>1.5 (.0,5.4)</td><td>-.8 (-4.9,.0)</td><td>5.3 (.0,37.4)</td></tr><tr><td>CRP</td><td>3.5 (.0,7.2)</td><td>-.30 (-.75,.00)</td><td>19.4 (.0,66.0)</td></tr></table>-Standard IV analyses for those outcomes with no evidence of non-linearity:
<table><tr><th>β (95% CI)</th><th>Per 1-unit increase in log-transformed genetically-predicted alcohol consumption</th></tr><tr><td>HDL-C</td><td>-.02 (-.07,.03)</td></tr><tr><td>TG (log-transformed)</td><td>-.01 (-.06,.07)</td></tr><tr><td>IL-6 (log-transformed)</td><td>.30 (.16,.45)</td></tr></table></td><td>causal relationships between alcohol and HDL-C or TG.</td></tr>
<tr><td>(Vu et al., 2016)</td><td>Lipids: TG, total cholesterol, HDL-C, HDL2-C, HDL3-C, LDL-C, sdLDL-C, apoB, Lp(a)

-cross-sectional</td><td>MR</td><td>European Americans; mean age 54.3 (SD 5.7) at baseline</td><td>Atherosclerosis Risk in Communities (ARIC); USA

10,893 individuals total; 9,911 for TG, 9,751 for total cholesterol and LDL-C, 10,132 for HDL-C, 10,120 for HDL2-C and HDL3-C, 8,102 for sdLDL-C, 7,663</td><td>-1-sample MR using a genetic risk score composed of 5 SNPs (rs2066702, rs1693457, rs1789891, rs698, and rs1126671) and standard IV analysis (2SLS); model fitted separately for quartiles of genetically-predicted alcohol consumption (Q1=1.49-3.63g/wk, Q2=3.63-4.66g/wk, Q3=4.66-10.57g/wk, Q4=10.57-19.54g/wk) to determine presence of non-linearity</td><td>✓</td><td>-TG (log-transformed):
<table><tr><th>Unstandardised β (95% CI)</th><th>MR</th><th>Using log-transformed alcohol intake</th><th>Conventional</th></tr><tr><td>Q2</td><td>-.06 (-.09,-.02)</td><td>Former/ infrequent</td><td>-.08 (-.11,-.05)</td></tr><tr><td>Q3</td><td>-.13 (-.20,-.07)</td><td>Low-to-moderate</td><td>-.16 (-.19,-.13)</td></tr><tr><td>Q4</td><td>-.08 (-.17,.00)</td><td>Heavy</td><td>-.13 (-.17,-.09)</td></tr></table>-TC:
<table><tr><th>Unstandardised β (95% CI)</th><th>MR</th><th>Using log-transformed alcohol intake</th><th>Conventional</th></tr></table></td><td>Results support causal relationships between alcohol and TG, TC, HDL2-C, LDL-C, sdlDL-C and apoB, such that some level of consumption leads to more favourable levels than abstinence/very</td></tr>
</table>

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
				for apoB, 9,924 for Lp(a)	-Conventional analyses regressing outcomes on observed alcohol consumption were also conducted, using consumption categories: **never drinkers**, former/infrequent drinkers (<1 drink/wk), low-to-moderate current drinkers (M: ≤210g/wk, F: ≤105g/wk), heavy current drinkers (M: >210g/wk, F: >105g/wk)		Q2 **-5.54** (-8.23,-2.85) Former/infrequent **-2.75** (-4.99,-.51) Q3 **-7.71** (-13.26,-2.15) Low-to-moderate -.73 (-3.03,1.57) Q4 -4.56 (-11.36,2.25) Heavy **4.05** (.73,7.37) -HDL-C (log-transformed): Unstandardised β (95% CI) / MR / Using log-transformed alcohol intake / Conventional Q2 .01 (-.01,.03) Former/infrequent **.03** (.01,.04) Q3 .04 (.00,.07) Low-to-moderate **.14** (.12,.15) Q4 .03 (-.02,.07) Heavy **.26** (.23,.28) -HDL2-C (log-transformed): Unstandardised β (95% CI) / MR / Using log-transformed alcohol intake / Conventional Q2 **.04** (.00,.07) Former/infrequent **.07** (.04,.10) Q3 **.10** (.03,.17) Low-to-moderate **.17** (.14,.20) Q4 .06 (-.03,.15) Heavy **.30** (.26,.35) -HDL3-C: Unstandardised β (95% CI) / MR / Using log-transformed alcohol intake / Conventional Q2 -.19(-.87,.49) Former/infrequent .40 (-.14,.94) Q3 .08 (-1.23,1.38) Low-to-moderate **4.16** (3.60,4.73)	**low consumption does. Analyses suggest non-linearity, with benefits peaking for most outcomes at roughly .5-1 drinks per week, although effect sizes vary. Results do not support causal relationships between alcohol and HDL-C, HDL3-C and Lp(a).**

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion

Q4	.11 (-1.51,1.74)	Heavy	**8.70** (7.84,9.55)

-LDL-C:

Unstandardised β (95% CI)	MR	Using log-transformed alcohol intake	Conventional
Q2	**-4.60** (-7.18,-2.03)	Former/ infrequent	**-2.59** (-4.69,-.49)
Q3	**-6.87** (-12.24,-1.50)	Low-to-moderate	**-4.38** (-6.53,-2.23)
Q4	-4.57 (-11.11,1.96)	Heavy	**-7.48** (-10.70,-4.25)

-sdLDL-C (log-transformed):

Unstandardised β (95% CI)	MR	Using log-transformed alcohol intake	Conventional
Q2	**-.04** (-.08,-.01)	Former/ infrequent	**-.04** (-.07,-.01)
Q3	**-.08** (-.15,-.01)	Low-to-moderate	**-.05** (-.08,-.01)
Q4	-.08 (-.17,.01)	Heavy	.01 (-.04,.06)

-apoB (log-transformed):

Unstandardised β (95% CI)	MR	Using log-transformed alcohol intake	Conventional
Q2	**-.03** (-.04,-.01)	Former/ infrequent	-.01 (-.02,.01)
Q3	-.04 (-.07,.00)	Low-to-moderate	-.01 (-.03,.01)
Q4	-.04 (-.08,.01)	Heavy	-.02 (-.04,.01)

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
							-Lp(a) (log-transformed): Unstandardised β (95% CI): -.02 (-.09,.06); -.05 (-.20,.10); -.01 (-.20,.18) MR: Former/ infrequent -.03 (-.10,.03); Low-to-moderate .01 (-.06,.08); Heavy -.06 (-.16,.03) Using log-transformed alcohol intake: Conventional	
Mortality								
(Sipilä et al., 2016)	All-cause mortality -median 30.2 yrs of follow-up	Twin	Same-sex twins aged 18-54 at baseline and free of chronic disease 6 years post-baseline; mean age 35.9 at baseline	Older Finnish Twin Cohort 14,787; 3,389 drinking-discordant pairs (of which 926 pairs are MZ)[a]	-Co-twin (discordant for alcohol consumption) Cox models and pooled cohort Cox models used to examine risk for mortality -Categories for pooled cohort + twin analyses based on average of two measurements 6 yrs apart: abstainers, **1-69g/mnth**, 70-139g/mnth, 140-209g/mnth, 210-419g/mnth, 420-839g/mnth, 840-1199g/mnth, ≥1200g/mnth	×	HR (95% CI): Cohort / MZ twins / All twins 0 g/mnth: 1.02 (.85,1.22) / .43 (.17,1.11) / .96 (.63,1.45) 70-139 g/mnth: .95 (.81,1.10) / .64 (.29,1.40) / .96 (.68,1.36) 140-209 g/mnth: 1.08 (.91,1.29) / 1.12 (.47,2.65) / .85 (.57,1.26) 210-419 g/mth: **1.29** (1.09,1.53) / 1.55 (.65,3.71) / 1.40 (.94,2.09) 420-839 g/mnth: **1.56** (1.31,1.85) / 1.70 (.69,4.22) / **1.60** (1.06,2.43) 840-1,199 g/mnth: **2.17** (1.74,2.70) / 1.65 (.54,5.08) / **2.65** (1.50,4.69) ≥1,200 g/mnth: **2.81** (2.26,3.50) / 3.18 (.81,12.43) / **2.99** (1.60,5.59)	**Results are consistent with a monotonically increasing risk function. Twin analyses support a causal role for increased risk of all-cause mortality for moderate-to-very heavy drinking.**
HIV seroconversion								
(Sander et al., 2013)	HIV sero-conversion -median 10.5 yrs of follow-up	MSM	Men who have sex with men and who were sexually active and HIV-seronegative at	Multicenter AIDS Cohort Study (MACS); USA 3,752	-MSM (Cox models) + standard Cox models for comparison -Categories for analyses based on average of two measurements 1 yr apart:	×	RR (95% CI): Non-MSM / MSM Moderate: .91 (.65,1.27) / 1.10 (.78,1.54) Heavy: 1.19 (.83,1.70) / **1.61** (1.12,2.29)	**Results are consistent with a monotonically increasing risk function. Abstainers and**

Study	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name and sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
			baseline; median age 33.4 at baseline		**abstainers**, moderate (1-14 drinks/wk), heavy (>14 drinks/wk); 1 drink considered = 14 mL			**moderate drinkers appear to have similar risk for HIV seroconversion, while heavy drinkers are at increased risk.**
Musculoskeletal health								
(Pietikäinen et al., 2011)	DP due to low back disorders (LBD) -29 years of follow-up	Twin	Same-sex twins aged 18-64 and not receiving pension at baseline; mean age 33.2 (SD 12) at baseline	Finnish Twin Cohort 24,043; 504 (284 M) pairs discordant for DP due to LBD	-Co-twin (discordant for DP due to LBD) Cox models and pooled cohort Cox models used to examine risk for mortality -Categories for pooled cohort + twin analyses: abstainers, **light (≤3 drinks/wk),** moderate (F: >3-≤7 drinks/wk; M: >3-≤14 drinks/wk), heavy (F: >7drinks/wk; M: >14 drinks/wk)	✖	HR (95% CI) Cohort: Abstainer .85 (.61,1.20); Moderate 1.07 (.79,1.43); Heavy 1.08 (.80,1.46). All twins: Abstainer .79 (.49,1.27); Moderate .94 (.62,1.42); Heavy 1.07 (.69,1.66)	**Results suggest a roughly monotonically increasing relationship between alcohol and later receipt of DP due to LBP, with reduced risk for abstainers the clearest feature.**
(Ropponen & Svedberg, 2013) *see also CVD section	DP due to Musculo-skeletal diagnoses (MSD) -5-10 years of follow-up	Twin	Twins with data from a prior study, and at time of that study were living in Sweden, <65, working and without DP/old age pension;	Swedish Twin Registry 31,206; 922 DP due to MSD-discordant pairs (of which 357 are MZ)	-Co-twin (discordant for DP due to MSD) Cox models and pooled cohort Cox models used to examine risk (stratified for sex in pooled model) -Categories for pooled cohort + twin analyses: abstainers, **light** (≤3 drinks/wk),	✓	HR (95% CI) Cohort / MZ twins / DZ twins / All twins: Abstainer .93 (.80,1.07) / 1.49 (.98,2.26) / .8 (.54,1.19) / 1.07 (.81,1.42); Moderate **.8** (.69,.93) / **2.33** (1.39,3.91) / **.67** (.48,.95) / 1.02 (.78,1.34); Heavy **.73** (.67,.81) / 1.11 (.82,1.51) / **.77** (.60,.98) / .88 (.73,1.07)	**Results do not support a clear relationship and later receipt of DP due to MSD, particularly**

<table>
<tr><th>Study</th><th>Health outcome/s and interval to follow-up</th><th>Study design</th><th>Participant characteristics</th><th>Cohort/s name and sample size</th><th>Statistical methodology
(reference group bolded)</th><th>Explicit testing for non-linearity?</th><th>Main results
(statistically significant findings bolded)</th><th>Conclusion</th></tr>
<tr><td></td><td></td><td></td><td>mean age at baseline 53.7 (SD 5.7) [b]</td><td></td><td>moderate (F: >3-≤7 drinks/wk; M: >3-≤14 drinks/wk), heavy (F: >7drinks/wk; M: >14 drinks/wk)</td><td></td><td></td><td>given the discrepancy in direction of effects between MZ and DZ twins.</td></tr>
<tr><td>(Ropponen et al., 2011)</td><td>DP due to Musculo-skeletal disorder (MSD) and osteo-arthritis specifically (OA)
-29 years of follow-up</td><td>Twin</td><td>MZ and same-sex DZ twins aged ≥18 and working at baseline; mean age 33.2 (SD 12) at baseline</td><td>Finnish Twin Cohort
24,043; 1,317 pairs discordant for DP due to MSD, 461 pairs discordant for DP due to OA</td><td>-Co-twin (discordant for DP due to MSD/OA) Cox models and pooled cohort Cox models used to examine risk for MSD and OA in men and women separately
-Categories for pooled cohort + twin analyses: abstainers, light (≤3 drinks/wk), moderate (F: >3-≤7 drinks/wk; M: >3-≤14 drinks/wk), heavy (F: >7drinks/wk; M: >14 drinks/wk)</td><td>×</td><td>-For DP due to MSD:
<table>
<tr><th>HR (95% CI)</th><th>Cohort (M)</th><th>Cohort (F)</th><th>All twins (M)</th><th>All twins (F)</th></tr>
<tr><td>Light</td><td>.91 (.59,1.40)</td><td>1.15 (.93,1.42)</td><td>2.04 (1.09,3.82)</td><td>1.07 (.81,1.41)</td></tr>
<tr><td>Moderate</td><td>.95 (.70,1.30)</td><td>1.23 (1.00,1.51)</td><td>1.50 (.94,2.38)</td><td>.97 (.74,1.27)</td></tr>
<tr><td>Heavy</td><td>1.01 (.73,1.39)</td><td>1.08 (.84,1.39)</td><td>1.58 (.99,2.53)</td><td>1.37 (.98,1.91)</td></tr>
</table>
-For DP due to OA specifically:
<table>
<tr><th>HR (95% CI)</th><th>Cohort (M)</th><th>Cohort (F)</th><th>All twins (M)</th><th>All twins (F)</th></tr>
<tr><td>Light</td><td>.99 (.50,1.96)</td><td>1.19 (.86,1.63)</td><td>4.07 (1.15,14.36)</td><td>1.33 (.82,2.14)</td></tr>
<tr><td>Moderate</td><td>.78 (.48,1.28)</td><td>.96 (.69, 1.32)</td><td>2.01 (.82,4.93)</td><td>1.12 (.70,1.82)</td></tr>
<tr><td>Heavy</td><td>1.04 (.64,1.71)</td><td>.86 (.60,1.25)</td><td>2.39 (.97,5.89)</td><td>2.32 (1.21,4.47)</td></tr>
</table></td><td>Results do not offer enough support for a clear functional form, but are consistent with abstinence representing the lowest risk for DP due to MSD and OA specifically. Heavy drinkers were also at increased risk for both outcomes and sexes.</td></tr>
</table>

[a] Ropponen & Svedberg (2013) also used this cohort to examine the relationship between alcohol and disability pension due to mental health diagnoses – not reported on here as Samuelsson et al. used the more informative exposure categorisation.

[b] From correspondence with authors.

Note: Where multiple models are reported in the original paper, the most adjusted analyses are reflected in the above table (excluding MR studies).

2.4 Discussion

This review found that improved causal inference methods have been applied minimally to research on alcohol–long-term health relationships. Non-linearity was apparent for several outcomes: prostate cancer and related mortality (reverse J-shaped and J-shaped respectively), dementia risk (J-shaped; although age of onset better characterised by monotonically increasing relationship), mental health (U/J-shaped for depression; increased risk for abstainers for disability pension due to MHD), and certain lipids (LDL-C; reverse J-shaped, sdLDL-C and apoB; monotonically decreasing, and HDL-2C; inverted reverse J-shaped). However, many of the individual comparisons from which these overall forms were derived were of small effect size or were imprecise. While the level of consumption coinciding with lowest risk varied between outcomes, it tended to fall in the light range – as little as .5-1 units/week (Vu et al., 2016). Positive linear/monotonically increasing relationships were found for DBP, hip circumference, IL-6, all-cause mortality, and HIV seroconversion (although it is not possible to partition short-term pathways via risky sexual behaviour and longer-term effects on immune function). No relationships were found between alcohol and HDL-3C, Lp(a), or waist-to-hip ratio.

Where multiple studies reported on an outcome, findings were inconsistent. For diabetes-related biomarkers, there was a positive linear relationship, but the one study reporting on T2D itself lacked power to support a clear functional form. For cardiovascular events/diagnoses, one twin study found preliminary evidence for a J-shaped relationship with myocardial infarction, while MR failed to find any relationship. For stroke, one twin study found a reverse J-shape (monotonically increasing for time to stroke), while MR found

monotonically increasing relationships. For cardiovascular disease more generally, no clear causal relationship emerged. The results of Millwood et al. imply that broad outcomes (in this case, total CHD) mask various discrepant sub-functional forms (as is likely for all-cause mortality; Rehm, 2019). For cardiovascular biomarkers, all three studies evaluating HDL-C were consistent with a lack of causal relationship. There was little consistency across other cardiovascular biomarkers, lipids, and obesity anthropometric measures, with conflicting functional forms found for non-HDL-C, triglycerides, total cholesterol, SBP, BMI, and waist circumference. This was the case even when the same MR method was used, which may reflect the impact of using different ethnic populations and genetic instruments. Finally, for musculoskeletal health, results varied between a monotonically increasing form, no clear functional form, and no clear functional form with nadir at abstinence.

2.4.1 Triangulation with the observational literature

Some of these findings are roughly consistent with the conclusions on functional form made by recent reviews of the broader observational literature, but there were also outcomes where the present findings do *not* concord with the broader literature. Evidence has been triangulated with the broader observational research in Table 2.3. This is an essential step, as evidence for most health outcomes was only available from one or two studies in the present review, and thus not definitive.

Table 2.3. Triangulation of included studies with reviews of the broader observational literature

Outcome	Study/ies first author (year)	Simplified summary of findings	Consistent with broader literature?	Details
Prostate cancer	Dickerman (2018)	-Reverse J-shaped relationship	No	A 2016 meta-analysis found a monotonically increasing relationship (J. Zhao et al., 2016)
All-cause mortality	Sipilä (2016)	-Monotonically increasing	Yes	A 2016 meta-analysis found a monotonically increasing relationship (Stockwell et al., 2016)
Diabetes biomarkers: **–FBG**	Peng (2019)	-Positive linear relationship	Yes	A 2017 review found results consistent with a positive linear relationship (Taghdiri, 2017)
–Glucose tolerance, insulin sensitivity and glycated haemoglobin	Peng (2019)	-Positive linear relationships for P2hBG (glucose tolerance) and HOMA-IR (insulin sensitivity) -Lack of relationship for HbA1c (glycated haemoglobin) and HOMA-beta (insulin sensitivity)	N/A	A 2017 review found that, for the other diabetes biomarkers included in this review, there was not enough/consistent evidence from which to draw conclusions (Taghdiri, 2017)
CVD biomarkers: **–Blood pressure**	Silverwood (2014)	-Non-linear relationship (small protective effect of light drinking) for SBP	No	A 2018 meta-analysis found no protective effect for hypertension (Roerecke et al., 2018)
	Peng (2019)	-Positive linear relationships for both SBP and DBP	Yes	
–Obesity anthropometrics	Silverwood (2014)	-Non-linear relationship (small protective effect of light drinking) for BMI and waist circumference	Mixed	A 2011 review found mixed evidence on the relationship between alcohol and weight/measures of abdominal adiposity, with heavy drinking generally associated with increased values, but moderate consumption either associated with lower values or not associated at all (suggested that discrepancies may depend on type of alcohol consumed; Sayon-Orea et al., 2011)
	Peng (2019)	-Positive linear relationships for BMI, waist circumference, and hip circumference, with no relationship for waist-to-hip ratio	Mixed	

Outcome	Study/ies first author (year)	Simplified summary of findings	Consistent with broader literature?	Details
–**Inflammatory markers**	Silverwood (2014)	-Non-linear relationship (small protective effect of light drinking) for CRP but a positive linear relationship for IL-6	N/A	A 2020 review of studies from the previous decade found no relevant longitudinal observational studies that were capable of detecting non-linear relationships (Minzer et al., 2020)
–**Lipids: HDL-C**	Silverwood (2014)	-No relationship for HDL-C	No	A 2021 review of studies from the previous decade found just one longitudinal observational study, which reported a reverse J-shaped relationship in decreases in HDL-C from baseline (Minzer et al., 2020)
	Peng (2019)	-No relationship for HDL-C	No	
	Vu (2016)	-No relationship for HDL-C, or HDL-3C, but an inverted reverse J-shaped relationship for HDL-2C	Mixed	
–**Lipids: other**	Silverwood (2014)	-No relationship for TG, but a non-linear relationship (small protective effect of light drinking) for non-HDL-C	N/A	A 2021 review of studies from the previous decade found no relevant longitudinal observational studies (Minzer et al., 2020)
	Peng (2019)	-Positive linear relationships for TG, non-HDL-C, and TC		
	Vu (2016	-Reverse J-shaped relationships for LDL-C, TG and TC, no relationship for Lp(a), and monotonically decreasing relationships for sdLDL-C and apoB		
CVD events/diagnoses: -Myocardial infarction	Ilomäki (2011)	-J-shaped relationship	N/A	A 2017 review found alcohol was detrimental for myocarditis (Rehm, Gmel Sr, et al., 2017), but no reviews on myocardial infarction were found
	Millwood (2019)	-No relationship	N/A	
–**Stroke**	Kadlecová (2015)	-Reverse J-shaped relationship for stroke and TIA	Mixed	A 2016 meta-analysis found J-shaped relationships for ischaemic stroke, but monotonically increasing relationships for intracerebral haemorrhage and subarachnoid haemorrhage (Larsson et al., 2016)

Outcome	Study/ies first author (year)	Simplified summary of findings	Consistent with broader literature?	Details
	Millwood (2019)	-Monotonically increasing relationships for ischaemic stroke, intracerebral haemorrhage and total stroke	Mixed	
–**Total coronary heart disease**	Millwood (2019)	-No relationship	No	A 2016 meta-analysis found a U-shaped relationship (Yang et al., 2016)
–**Circulatory system diagnoses**	Ropponen (2013)	-No clear functional form	N/A	Most reviews evaluate CVD sub-conditions, finding divergent functional forms
Dementia	Handing (2015)	-J-shaped relationship	Yes	A 2021 review found most recent evidence is consistent with a J-shaped relationship (Visontay, Rao, et al., 2021)
Mental health	Gémes (2019)	-U-/J-shaped relationship	Yes	A 2020 meta-analysis found a J-shaped relationship with depressive symptoms (Li et al., 2020)
	Samuelsson (2013)	-Non-linear relationship with abstainers at increased risk over light frequent drinkers	Yes	
HIV seroconversion	Sander (2013)	-Monotonically increasing relationship	N/A	There is a lack of dose-response information on alcohol's relationship with HIV (Rehm, Probst, et al., 2017; Shuper et al., 2010)
Musculoskeletal disorders	Pietkäinen (2011)	-Monotonically increasing relationship	N/A	No reviews on the relationship between alcohol and musculoskeletal disorders were found
	Ropponen (2013)	-No clear functional form	N/A	
	Ropponen (2011)	-No clear functional form, but abstainers have lowest risk	N/A	

Note: Risk for type 2 diabetes (Carlsson et al.) is excluded from this table as the analyses were critically underpowered, and thus the results are not appropriate for inclusion in the triangulation process. Findings from RCTs were not integrated into triangulation as they are almost exclusively conducted in short-term time frames, which is not of interest to this review.

Importantly though, where included studies performed conventional analyses for comparison with modern methods to address confounding, results were often discrepant. Most starkly, Millwood et al. found typical J/U/reverse-J-shaped relationships via conventional analyses, but monotonically increasing (or no) relationships when using MR. Even when functional forms roughly replicated across methods, the strength of individual comparisons differed. With pooled cohort analysis, Dickerman et al. found abstainers had slightly increased risk for prostate cancer, compared to much greater risk when utilising discordant twins. The application of improved causal inference methods is therefore essential to accurately characterise alcohol–long-term health relationships.

Also of note, there were several methods of interest for which no eligible studies were identified. It may be that certain methods are ill-suited to address this research question – for example, it may be difficult to find a non-genetic IV to proxy for multiple levels of alcohol consumption, while others, such as G-estimation, may not yet have gained traction in research more generally (Watkins, 2019). Negative controls, while not identified as a primary design approach, were incorporated into two of the MR studies. Of the methods that were represented, there were fewer eligible studies than expected – particularly for MR, where many articles were excluded for only performing linear IV analyses, or for providing estimates in terms of the effects of genetic variants (e.g., ADH1B A-allele carriers vs non-carriers) rather than genetically-predicted alcohol consumption. Consistent with other reviews of the literature (Wallach et al., 2020), covariates controlled for across studies varied considerably (see Appendix 1.7).

2.4.2 Strengths

This review applied a novel framework to examining alcohol–long-term health relationships, identifying and synthesising information from those observational studies that best mitigate confounding and thus promote causal inference. The search strategy included terms for a broad range of analytical and design-based approaches informed by the literature and consultation with experts. As many included studies performed both conventional analyses and causal inference approaches, this review was able to highlight the difference that such methods make. A further strength was that all long-term health outcomes were eligible, providing a comprehensive picture of the state of the evidence base, and importantly, on the large gaps in the literature where the aforementioned methods require application.

2.4.3 Limitations

While the included studies mitigate confounding, other methodological limitations (not necessarily captured by the NOS) may be present. For example, the prospective cohort studies likely suffered from sick quitter bias in failing to separate ex-drinkers from lifelong abstainers – exacerbated in those studies where baseline mean age was over 50. Misclassification was likely in many of the twin studies as most based drinking categories on a single measurement, and most of these focused on shared confounding, without additionally controlling for measured covariates. While MSMs can account for consumption and covariates at multiple timepoints, these studies were still vulnerable to residual confounding, with Gémes et al. (2019) noting that unmeasured social confounders may partially underpin their findings. As approaches to minimise these biases consist largely of literature-informed, considered researcher decisions and sensitivity analyses, they are not suited to systematic

database searching, and were not the focus of this review. While MR is largely immune to both misclassification and confounding, it suffers from its own idiosyncratic limitations, with controversy over its application to alcohol–long-term health research specifically (Davey Smith et al., 2020; Ellison et al., 2021; Mukamal et al., 2020). For example, two of the included MR studies discretised genetically-predicted alcohol consumption in a way that resulted in the lowest categories aligning with occasional consumption – not strictly comparable with abstinence.

Additionally, despite evidence of the importance of accounting for pattern of consumption (Roerecke & Rehm, 2014), MR studies are limited in their ability to do so (Ellison et al., 2021), and only one cohort study (Samuelsson et al., 2013) used drinking pattern as the exposure, rather than volume alone (or volume and frequency separately). Finally, several of the included studies evaluated alcohol's relationship with condition-specific disability pension, rather than the condition itself. This is an imperfect proxy, with receipt of pension also reflecting the interference of the disease with one's ability to work, as well as incentive to apply (Pietikäinen et al., 2011). Given that all included studies evaluating musculoskeletal health used disability pension as a proxy, findings with respect to these outcomes should be interpreted with caution.

2.4.4 Future Directions

This review has identified clear gaps in alcohol–long-term health research, demonstrating great potential for further application of enhanced causal inference methods. Only three of the causal methods of interest identified *a priori* have been applied thus far: twin studies, marginal structural models, and MR, indicating that these may be the most appropriate and

feasible for answering research questions in this space. This is also in keeping with trends in causal inference/epidemiological literature more broadly, where G-methods and MR appear to be the key statistical and design-based approaches respectively.

Analysis and design-based approaches that mitigate confounding should be combined with sensitivity analyses such as multiverse analyses (to quantify robustness to data processing/analysis decisions; Steegen et al., 2016), as well as bias analyses such as E-value generation (to quantify robustness to unmeasured confounding; VanderWeele & Ding, 2017). Given the unique advantages and limitations of each analytical and design-based approach, triangulation of findings across observational evidence is crucial. Combining data across studies through data harmonisation techniques should also be considered for mitigating power issues, which were evident in several included studies with rare exposure–outcome combinations. Finally, while acknowledging the limitations of the included studies, the identification of some evidence consistent with causal protective effects of light-to-moderate alcohol consumption for several health outcomes justifies further exploration of the biological mechanisms that could underpin these (potential) effects. This is particularly true of those outcomes for which findings were concordant with the broader observational literature (see Table 2.3).

2.4.5 Conclusions

This novel review found that, when enhanced causal inference approaches are applied, a variety of functional forms – including linear, J-shaped, and no relationship – are found between alcohol consumption and various long-term health outcomes. However, few studies have employed these methods, with covariate-adjusted, conventional cohort analyses

remaining dominant, preventing a conclusive picture of the nature of these relationships from emerging. Given that associations found between moderate alcohol consumption and good health impact safe drinking guidelines and public health policy (Sherk et al., 2019; Stockwell & Room, 2012), further research employing methods to mitigate confounding and other biases is urgently required to establish whether such findings are truly causal.

Chapter 3

Is the longitudinal association between alcohol consumption and inflammation robust?: A multiverse analysis

Preface

Chapter 2 found a dearth of studies employing causal inference approaches when examining alcohol–long-term health relationships. This is therefore an area ripe with opportunity for the application of such methods. But there are necessary conditions for causation that must be met *even before* these methods can be applied. An important but oft-overlooked criterion required for assigning causation to an association is that the association is a robust, consistent one. This aligns roughly with Bradford-Hill's consistency criterion for causation, as described in **Chapter 1**. This chapter reports on a study that aimed to assess this criterion for the relationship between alcohol consumption and subsequent levels of inflammation by applying enhanced sensitivity methods.

Multiverse analyses are applied to comparisons of low-to-moderate consumption and consumption above various international drinking guidelines with an 'abstinent' reference, with variation in definitions of drinking and reference groups, alcohol consumption measurement year, outcome variable transformation, and breadth of covariate adjustment. The association between low-to-moderate drinking and lower high-sensitivity C-reactive

protein (hsCRP) levels is found to be largely robust to common variations in researcher-defined parameters, warranting further research to establish whether this relationship is causal. The association between above-guidelines drinking and hsCRP levels seems less definitive.

A slightly modified version of this chapter was published as Visontay, R., Mewton, L., Sunderland, M., Bell, S., Britton, A., Osman, B., North, H., Mathew, N., & Slade, T. (2023). A comprehensive evaluation of the longitudinal association between alcohol consumption and a measure of inflammation: Multiverse and vibration of effects analyses. *Drug and Alcohol Dependence, 247*, 109886, and is available in Appendix C. Supplementary materials are available in Appendix 2.

3.1 Introduction

Acute inflammation is a vital and adaptive response to pathogens and tissue injury. However, low-grade chronic inflammation, i.e., inflammation that fails to resolve or is continuously triggered, can negatively affect tissue and organs over time. This pathology ranges in severity, on the higher end associated with inflammatory diseases such as asthma, atherosclerosis, inflammatory bowel disease, and rheumatoid arthritis (Calder et al., 2013; Furman et al., 2019; Ridker, 2004). Inflammation is also increasingly implicated in a variety of other serious conditions including psychiatric disorders, cancer, type 2 diabetes, and cardiovascular disease (Ansar and Ghosh, 2013; Avan et al., 2018; Furman et al., 2019; Réus et al., 2015).

Of the many putative lifestyle influences on inflammation, there is strong evidence for the role of alcohol consumption. Alcohol use disorder is associated with increased levels of pro-inflammatory cytokines (Adams et al., 2020) as well as neuroinflammation specifically (de Timary et al., 2017; He & Crews, 2008). Harmful drinking is responsible for substantial disease burden globally (Bryazka et al., 2022), and there is evidence that inflammation may mediate the increased risk heavy drinking brings for a range of health outcomes (González-Reimers et al., 2014).

However, as for many conditions to which alcohol consumption is thought to contribute pathophysiologically (Visontay et al., 2022), some research suggests that compared to abstaining, lower levels of drinking may be beneficial when it comes to inflammation (Albert et al., 2003; Bell et al., 2017; Pai et al., 2006; Paulson et al., 2018; Volpato et al., 2004). This is typically indexed via inflammatory markers such as hsCRP – widely used by the clinical

and research communities due to its stable expression in response to elevated pro-inflammatory cytokines (Coventry et al., 2009). Reduced inflammation has also been suggested as a partial mediator of the lower risk moderate drinking seemingly affords against conditions such as cardiovascular disease (Pai et al., 2006) and depression (Paulson et al., 2018).

However, there are major inconsistencies in the evidence base. For example, short-term alcohol administration tends to demonstrate no relationship with hsCRP (Brien et al., 2011). Several cohort studies have also failed to find a relationship between drinking and inflammation (Menezes et al., 2019; van't Klooster et al., 2020). A possible explanation for these conflicting findings is that data processing and analysis choices, such as which covariates are controlled for, vary considerably between studies. A growing literature is beginning to uncover the impact these alternative specifications in research parameters, or 'researcher degrees of freedom' (Simmons et al., 2011), can have on effect estimates, challenging implied assumptions that these decisions are innocuous.

To this end, three similar analytic frameworks have emerged to quantify sensitivity (or robustness) to alternative specifications: multiverse analysis (Steegen et al., 2016), specification curve analysis (Simonsohn et al., 2020), and vibration of effects (VoE; C. J. Patel et al., 2015). These frameworks are an extension of standard sensitivity analyses, providing methods to comprehensively analyse and summarise the sensitivity of exposure–outcome associations to alternative specifications.

It is well documented that these alternative specifications can influence effect sizes, o values, and even effect direction for a variety of epidemiological associations. This is becoming

increasingly clear in alcohol–long-term health associations specifically. While regressing a health outcome on drinking level is ostensibly straightforward, this process involves numerous decisions on the part of the researcher. Even when using the same dataset, and therefore holding study design and available data constant, different researchers could, and do, arrive at very different analyses based on data processing and analysis choices. Such parameters include how the abstaining reference group is defined (Stockwell et al., 2016), with increasing calls (but limited uptake) for occasional drinkers to be used as the reference given their greater prevalence and more normative risk factor profiles (T. Naimi et al., 2022). How alcohol intake is categorised, which covariates are adjusted for, and choice of modelling approach have also been identified as factors that may substantially impact findings in this space (Bell et al., 2017; Chu et al., 2020; Stockwell et al., 2016). It is thus concerning that there remains great variation in these parameters in the literature, and that their potential impact is rarely acknowledged in individual studies, where typically only one set of specifications are tested.

The present study aims to address this by systematically applying methods from enhanced sensitivity frameworks to test the alcohol–inflammation association. Using the 1970 British Birth Cohort Study (BCS70) – a source of rich, representative data – we assessed the relationship between alcohol consumption in early/mid-adulthood and hsCRP at age 46. The research parameters of interest explored here comprise: composition of reference group; classification of drinking groups according to various national guidelines ranging from conservative through liberal; measurement year(s) for alcohol consumption; transformation of the hsCRP outcome; and breadth of covariate adjustment. Each unique combination of specifications within these parameters constitutes a possible 'universe' within the analytic

multiverse; here alcohol–hsCRP analyses are iterated over each universe and results interpreted on aggregate.

3.2 Methods

Reporting for this study adheres to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (von Elm et al., 2007).

3.2.1 Data source and participants

BCS70 is a prospective cohort study targeting all individuals born in a single week in 1970 in Great Britain. It has had 11 sweeps to date, collecting diverse information on health, education, lifestyle behaviours, and SES factors (Elliott & Shepherd, 2006). Individuals possessing valid alcohol data for at least one of ages 34 and 42, and an hsCRP measurement at age 46, were considered eligible for the present study.

3.2.2 Exposure

Self-reported alcohol consumption data at ages 34 (2004) and 42 (2012) were used to generate the exposure variables. In the present study, data on frequency and volume of alcohol consumption were converted to grams of alcohol per week, and then used to categorise individuals as either lifetime abstainers, former drinkers, occasional drinkers (less than weekly), low-to-moderate drinkers (at least once per week), or those consuming above various international drinking guidelines at each of the waves. Additional exposure variables were created re-categorising current drinkers according to their average weekly volume across the two ages, as well as their maximum weekly volume. Four separate versions were created for each of these four variables, where the threshold separating low-to-moderate from above-guidelines drinkers varied according to the national drinking guidelines of four chosen countries (see Table 3.1).

More specifically, when classifying participants based on their age 34 data, those responding 'Has never had an alcoholic drink' to the frequency of consumption question were classified as lifetime abstainers, those responding 'Never nowadays' were classified as former drinkers, and those responding '2 to 3 times a month' or 'Less often or only on special occasions' were classified as occasional drinkers. Those responding 'On most days', '2 to 3 days a week', or 'Once a week', were classified as current, more-than-occasional drinkers, and then further classified based on information provided on units consumed per week. Specifically, a derived units per week variable (converting answers about units of particular beverages in the past seven days, based on the formula: beer units x 2 + spirits units + wine units + sherry units + alcopops units; Dodgeon, 2020) was used to classify these individuals as either low-to-moderate drinkers or above-guidelines drinkers separately according to each of the four national guidelines' thresholds.

When classifying participants based on their age 42 data, frequency of alcohol consumption was used again. As there were no variables going to lifetime abstention at this wave, those responding 'Never' and who were classified as lifetime abstainers at age 34 were also classified as lifetime abstainers at age 42, while those responding 'Never' but were classified as something other than a lifetime abstainer at age 34 were classified as former drinkers. Those responding 'Never' and missing age 34 drinking status classification were classified as an abstainer not otherwise defined, and their data was only able to contribute to those universes where the reference group was composed of all abstainers. Those responding 'Monthly or less' or '2-4 times a month' were classified as occasional drinkers, those responding '2-3 times a week' or '4 or more times a week' were classified as current, more-than-occasional drinkers, and then further classified based on information provided on units

consumed per week. Participants answered a question about the number of alcoholic drinks consumed on a typical drinking day, with options 1-2, 3-4, 5-6, 7-8, and 10+. To create a units per week variable, participants were assumed to drink the average number of days of their frequency response (i.e., 2.5 times per week or 5.5 times per week), and the average number of drinks of their units per drinking day response (i.e., 1.5 drinks, 3.5 drinks, 5.5 drinks, or – according to precedent for adding three-quarters of the range of the second-highest drinking category to the upper bound for open-ended upper categories; Stockwell et al., 2016 – 10.75 drinks). Frequency and units were then multiplied to produce a units per week variable, and used to classify these individuals as either low-to-moderate drinkers or above-guidelines drinkers separately according to each of the four national guidelines' thresholds.

When classifying participants based on their maximum drinking category of the two waves, those with valid data for only one of the waves were assigned that value as their maximum. When classifying participants based on their average drinking category, those with valid data for only one of the waves were assigned that value as their average. Those abstaining at both waves were classified as either lifetime abstainers or former drinkers, while those drinking occasionally at both waves were classified as occasional drinkers. For those with valid data for both waves where at least one wave indicated more-than-occasional drinking, an average of their derived units per week variables at the two waves was calculated (substituting 0 units per week for abstention or occasional drinking at a given wave where required). This average was then used to classify these individuals as either low-to-moderate drinkers or above-guidelines drinkers separately according to each of the four national guidelines' thresholds.

3.2.3 Outcome

A non-fasting blood sample was taken at age 46 (2016), with hsCRP concentrations assessed via immunoturbidimetry (Hamer et al., 2020). Prior to analysis, 110 individuals with hsCRP concentrations > 10 mg L^{-1} were excluded, given these concentrations suggest acute inflammation or infection (Giollabhui et al., 2020). Chi-Square tests of independence revealed no relationship between drinking category at age 34 and likelihood of hsCRP being > 10 mg L^{-1}, although there was some evidence of a relationship between drinking category at age 42 and likelihood of hsCRP being > 10 mg L^{-1} (Appendix 2.1). The hsCRP distribution was positively skewed, so a natural log-transformed version was also created which led to an approximately normal distribution, and universes with both untransformed and log-transformed hsCRP were tested. Normality was assessed via visual inspection.

3.2.4 Covariates

Covariates were chosen *a priori* based on their relevance to the alcohol–inflammation relationship and were derived from data collected at either birth (for time-invariant variables) or age 30 (this was the nearest measurement occasion preceding age 34 alcohol consumption measurement, chosen to ensure covariates temporally preceded alcohol measurement). Covariates of interest were divided into five groups: demographic/SES (sex, ethnicity, mother's age at birth, parents' marital status at birth, years of mother's education, father's job, mother's social status, employment status, financial situation, homelessness, marital status), physical health (delivery method at birth, self-rated health, disability, doctor visitation, BMI, history of various inflammatory/other conditions), lifestyle behaviours (smoking status, shift work, sleep, fruit/vegetable/sugar consumption, exercise), social (social

support, religious and community participation), and mental health (malaise inventory score, history of eating disorders, mental health practitioner visitations). See Appendix 2.2 for further detail.

3.2.5 Parameters of interest and their specifications

As detailed in Table 3.1, the research parameters for which alternative specifications were explored were: 1) composition of 'abstinence' reference group (including using occasional drinkers instead of abstainers); 2) the cut-point between low-to-moderate and above-guidelines drinking (based on four different national drinking guidelines ranging from conservative to liberal); 3) measurement year(s) of the alcohol consumption predictor; 4) hsCRP outcome transformation; and 5) breadth of covariate adjustment. This produced a total of 1,248 universes. Note that specification alternatives often resulted in sample size variation (reflected in the bottom section of each specification curve plot; Figures 3.1 and 3.3, and Appendices 2.6, 2.8, 2.10, and 2.11).

Table 3.1. Parameters and specifications tested

Parameter	**Specifications**	***N* of specifications**
Composition of 'abstinence' reference group	(i) Lifetime abstainers only (ii) Lifetime abstainers and former drinkers (iii) Occasional drinkers only	3
Cut-point for above-guidelines drinking	(i) Dutch guidelines (>70g/wk) (Health Council of the Netherlands, 2015) (ii) UK guidelines (>112g/wk) (Department of Health, 2016) (iii) US guidelines (F: >98g/wk; M: >196g/wk) (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020) (iv) Former Spanish guidelines[a] (F: >170g/wk; M: >280g/wk) (Spanish Ministry of Health, 2020)	4
Year of alcohol consumption measurement	(i) 2004 (ii) 2012 (iii) Maximum of 2004 and 2012 (iv) Average of 2004 and 2012	4
Transformation of hsCRP outcome	(i) Leave untransformed (ii) Natural log-transformed	2
Breadth of covariate adjustment	(i) No covariates adjusted for (ii) Demographics (iii) Demographics + Physical health (iv) Demographics + Mental health (v) Demographics + Lifestyle (vi) Demographics + Social (vii) Demographics + Physical health + Mental health (viii) Demographics + Physical health + Lifestyle (ix) Demographics + Mental health + Lifestyle (x) Demographics + Mental health + Social (xi) Demographics + Physical health + Social (xii) Demographics + Lifestyle + Social (xiii) All covariates adjusted for	13
Total universes		1,248

The total number of universes is equal to the product of the five parameters' specification *N*s.

[a] These guidelines are no longer in effect as of 2020 but are indicative of liberal thresholds for 'moderate consumption' adopted by various policymakers and researchers.

3.2.6 Statistical analysis

A linear regression of continuous hsCRP levels on categorical alcohol consumption was performed for all 1,248 universes, producing two comparisons: 'abstinence' vs. low-to-moderate alcohol consumption, and 'abstinence' vs. above-guidelines alcohol consumption. Effect sizes were standardised.

Various approaches from the sensitivity frameworks introduced earlier were employed to interpret results. Specification curve plots allowed for visual inspection of variation in effect sizes and *p* values across universes. In these plots, universes are arranged in order of effect size, with an accompanying panel depicting the corresponding specifications. Results were also graphically depicted using volcano plots, which promote evaluation of (in)consistency of effect direction (Klau et al., 2021). Metrics developed in the VoE framework were also calculated: the range of betas and range of *p* values. These are simple figures calculated as the range of standardised betas and *p* values between the 1^{st} and 99^{th} percentiles, where larger ranges indicate greater variability across universes (C. J. Patel et al., 2015). Additionally, variance in effect estimates were decomposed by parameter, and each specification plotted against effect estimates in box and whisker plots. All analyses were conducted using R version 4.1.3 with packages *multiverse* (Sarma, 2021) and *specr* (Masur & Scharkow, 2020).

3.3 Results

3.3.1 Descriptive characteristics

In total, 11,123 individuals provided valid age 34 and/or age 42 alcohol data, with 3,211 of these also having age 46 hsCRP data. The small size of the latter figure owed to attrition and low health test completion rates, although drinking status and covariates were largely unassociated with providing hsCRP data (Appendix 2.3). After removing those with elevated hsCRP, the final sample was 3,101.

Compared to other categories, low-to-moderate drinkers were more likely to rate their health as excellent and had lower mean Malaise Inventory and hsCRP values (Table 3.2 and Appendix 2.4). There was considerable movement amongst drinking categories between ages 34 and 42, particularly a decrease in individuals drinking above-guidelines and an increase in those drinking occasionally.

Table 3.2. Cohort characteristics for selected covariates and hsCRP by age 34 alcohol consumption categories based on UK guidelines

	Lifetime abstainer (N=52)	**Former drinker (N=108)**	**Occasional drinker (N=621)**	**Low-to-moderate drinker (N=1160)**	**Above-guidelines drinker (N=684)**	**Total (N=2625)**
UK standard drinks per week						
Mean (SD)	NA (NA)	NA (NA)	NA (NA)	6.99 (3.68)	29.8 (18.6)	15.5 (16.1)
Median [Min, Max]	NA [NA, NA]	NA [NA, NA]	NA [NA, NA]	6.00 [1.00, 14.0]	24.0 [15.0, 280]	10.0 [1.00, 280]
Missing	52 (100%)	108 (100%)	621 (100%)	0 (0%)	0 (0%)	781 (29.8%)
Sex						
Female	28 (53.8%)	71 (65.7%)	406 (65.4%)	686 (59.1%)	154 (22.5%)	1345 (51.2%)
Male	24 (46.2%)	37 (34.3%)	215 (34.6%)	474 (40.9%)	530 (77.5%)	1280 (48.8%)
Mother's age at birth						
Mean (SD)	28.1 (5.78)	25.1 (4.72)	25.9 (5.23)	26.4 (5.14)	26.1 (5.22)	26.2 (5.19)
Median [Min, Max]	28.0 [17.0, 41.0]	25.0 [16.0, 38.0]	25.0 [15.0, 44.0]	26.0 [15.0, 43.0]	25.0 [16.0, 49.0]	25.0 [15.0, 49.0]
Missing	12 (23.1%)	11 (10.2%)	35 (5.6%)	77 (6.6%)	48 (7.0%)	183 (7.0%)
Self-rated health (Age 30)						
Excellent	16 (30.8%)	25 (23.1%)	185 (29.8%)	411 (35.4%)	200 (29.2%)	837 (31.9%)
Good	25 (48.1%)	60 (55.6%)	322 (51.9%)	590 (50.9%)	355 (51.9%)	1352 (51.5%)
Fair	6 (11.5%)	14 (13.0%)	79 (12.7%)	96 (8.3%)	87 (12.7%)	282 (10.7%)
Poor	2 (3.8%)	5 (4.6%)	13 (2.1%)	6 (0.5%)	5 (0.7%)	31 (1.2%)

	Lifetime abstainer (N=52)	Former drinker (N=108)	Occasional drinker (N=621)	Low-to-moderate drinker (N=1160)	Above-guidelines drinker (N=684)	Total (N=2625)
Missing	3 (5.8%)	4 (3.7%)	22 (3.5%)	57 (4.9%)	37 (5.4%)	123 (4.7%)
BMI (Age 30)						
Mean (SD)	24.2 (4.27)	24.8 (5.48)	25.2 (4.81)	24.2 (4.09)	25.0 (3.44)	24.7 (4.21)
Median [Min, Max]	24.0 [18.1, 40.2]	23.6 [16.9, 45.2]	24.1 [16.4, 57.2]	23.4 [11.1, 61.2]	24.9 [9.98, 44.9]	24.0 [9.98, 61.2]
Missing	5 (9.6%)	9 (8.3%)	43 (6.9%)	83 (7.2%)	47 (6.9%)	187 (7.1%)
Smoking status (Age 30)						
Never smoker	37 (71.2%)	47 (43.5%)	304 (49.0%)	548 (47.2%)	241 (35.2%)	1177 (44.8%)
Former smoker	3 (5.8%)	18 (16.7%)	89 (14.3%)	253 (21.8%)	144 (21.1%)	507 (19.3%)
Occasional smoker	0 (0%)	9 (8.3%)	34 (5.5%)	94 (8.1%)	69 (10.1%)	206 (7.8%)
Daily smoker	9 (17.3%)	30 (27.8%)	172 (27.7%)	208 (17.9%)	193 (28.2%)	612 (23.3%)
Missing	3 (5.8%)	4 (3.7%)	22 (3.5%)	57 (4.9%)	37 (5.4%)	123 (4.7%)
Weekly exercise (Age 30)						
No	14 (26.9%)	29 (26.9%)	213 (34.3%)	295 (25.4%)	172 (25.1%)	723 (27.5%)
Yes	35 (67.3%)	74 (68.5%)	386 (62.2%)	808 (69.7%)	475 (69.4%)	1778 (67.7%)
Missing	3 (5.8%)	5 (4.6%)	22 (3.5%)	57 (4.9%)	37 (5.4%)	124 (4.7%)
Malaise scale total (Age 30)						
Mean (SD)	3.65 (3.73)	4.23 (3.69)	3.61 (3.48)	2.93 (3.06)	3.30 (3.05)	3.26 (3.22)

	Lifetime abstainer (N=52)	**Former drinker (N=108)**	**Occasional drinker (N=621)**	**Low-to-moderate drinker (N=1160)**	**Above-guidelines drinker (N=684)**	**Total (N=2625)**
Median [Min, Max]	3.00 [0, 15.0]	3.00 [0, 16.0]	3.00 [0, 18.0]	2.00 [0, 21.0]	2.00 [0, 22.0]	2.00 [0, 22.0]
Missing	3 (5.8%)	5 (4.6%)	25 (4.0%)	67 (5.8%)	41 (6.0%)	141 (5.4%)
Ever member of communal organisation (Age 30)						
No	34 (65.4%)	86 (79.6%)	461 (74.2%)	838 (72.2%)	553 (80.8%)	1972 (75.1%)
Yes	15 (28.8%)	17 (15.7%)	138 (22.2%)	265 (22.8%)	94 (13.7%)	529 (20.2%)
Missing	3 (5.8%)	5 (4.6%)	22 (3.5%)	57 (4.9%)	37 (5.4%)	124 (4.7%)
hsCRP (log-transformed) (Age 46)						
Mean (SD)	-0.122 (3.04)	-0.0338 (2.27)	0.0187 (2.12)	-0.312 (2.80)	-0.0637 (1.67)	-0.154 (2.38)
Median [Min, Max]	0.0953 [-20.7, 2.16]	0.182 [-20.7, 2.30]	0.0953 [-20.7, 2.24]	-0.105 [-20.7, 2.29]	-0.105 [-20.7, 2.30]	0 [-20.7, 2.30]

Abbreviations: SD, standard deviation; BMI, body mass index; hsCRP, high-sensitivity C-reactive protein.

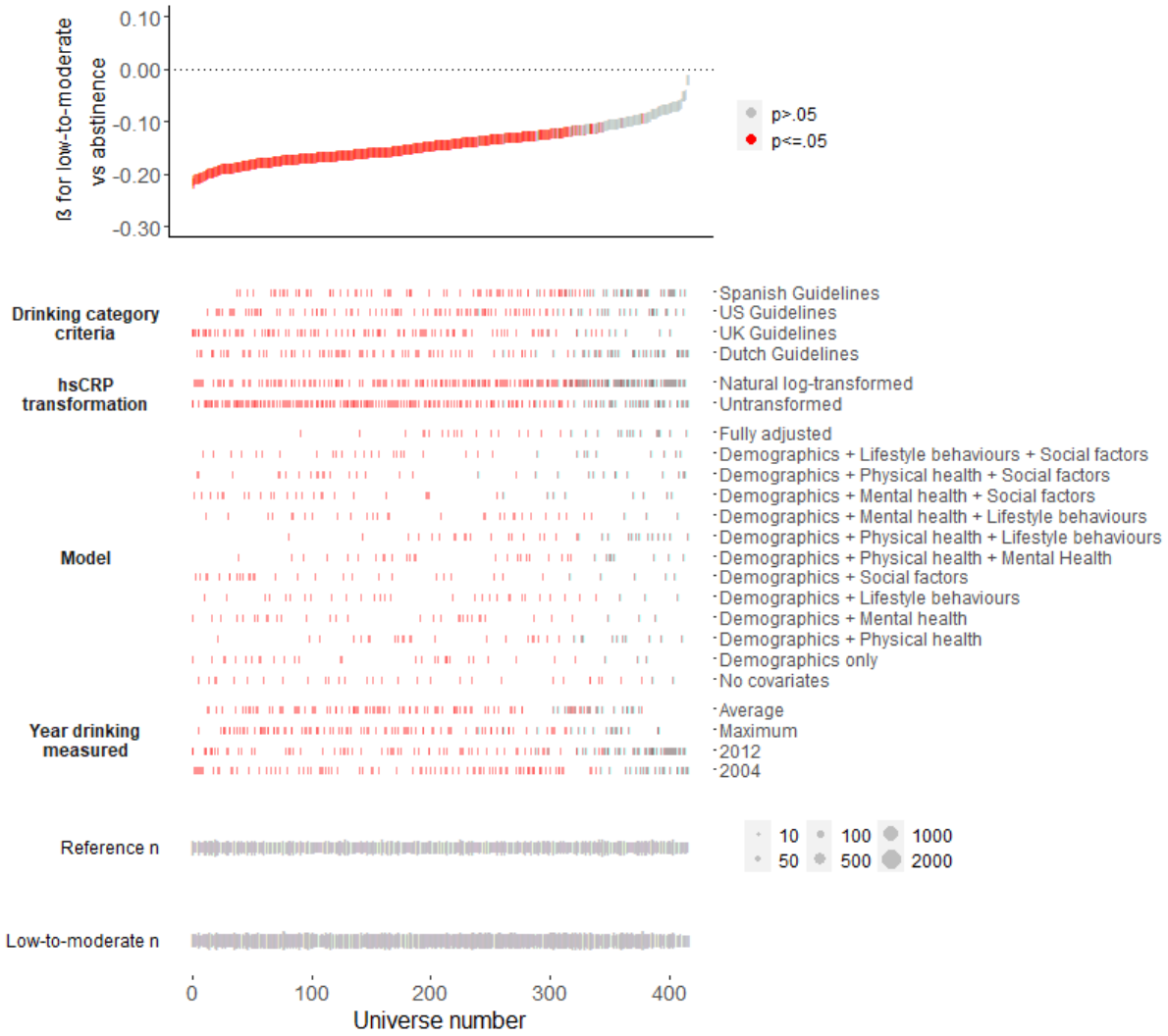
See Appendix 2.4 for characteristics by age 42 drinking categories.

3.3.2 Summary of effects and their heterogeneity

Despite a considerable range of effects (range of betas=0.46; range of *p* values=4.03), results favoured lower levels of hsCRP in low-to-moderate drinkers compared to the 'abstainer' reference, with effects at both the 1st (-0.50) and 99th (-0.05) percentiles in this direction. None of the 1,248 estimates were consistent with increased hsCRP levels (Appendices 2.6 and 2.7). Estimates comparing above-guidelines drinking with 'abstainers' were less definitive (range of betas=0.86; range of *p* values=2.29; Appendices 2.8 and 2.9), with 136/1,248 estimates consistent with increased hsCRP levels, and the effects at the 1st (-0.47) and 99th (0.39) percentiles being in opposite directions. Appendices 2.6 and 2.8 reveal considerable variation in effects according to reference group composition (lifetime abstainers and former drinkers; lifetime abstainers only; occasional drinkers), making it difficult to interpret other parameters.

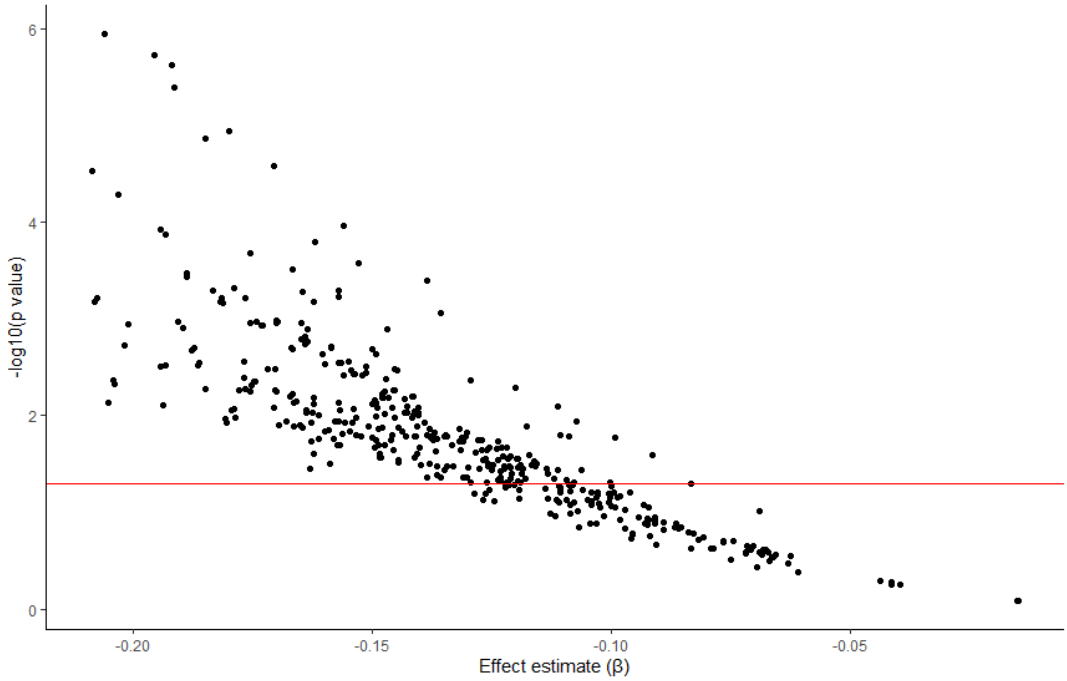
Occasional drinkers proved the only appropriate reference given the very small sample size resulting from the other two reference specifications. Analyses with this specification were thereafter considered the primary analyses, to mitigate confounding by sample size and aid interpretability (Figures 3.1-3.4). The range of effects in these analyses was narrower, with Ranges of Betas of 0.16 (range of *p* values= 4.65) and 0.69 (range of *p* values= 2.87) for the low-to-moderate (0/416 estimates consistent with increased hsCRP; 1st percentile effect: -0.21; 99th percentile effect: -0.04) and above-guidelines (85/416 estimates consistent with increased hsCRP; 1st percentile effect: -0.26; 99th percentile effect: 0.43) comparisons respectively.

Figure 3.1. Specification curve plot of estimates comparing low-to-moderate drinkers with a reference group composed of occasional drinkers.



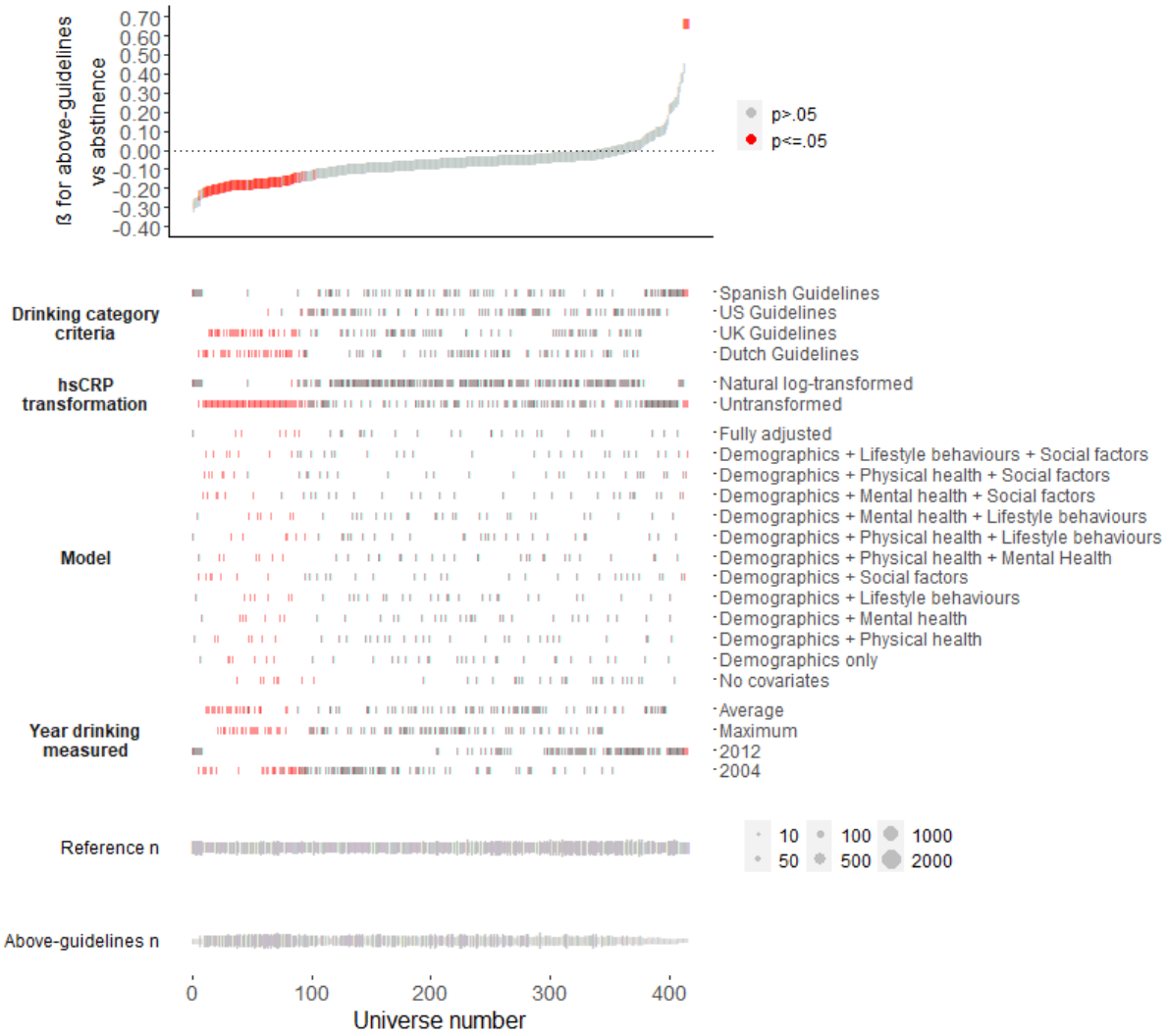
The top section displays the results associated with each universe, ordered by effect size. Negative values indicate drinking is associated with lower hsCRP. The middle section shows which specifications correspond to each result in the top panel. The bottom section displays the number of individuals in the reference and comparator groups in each universe.

Figure 3.2. Vibration of effects volcano plot depicting effects of low-to-moderate drinking and their *p* values compared with a reference group composed only of occasional drinkers.



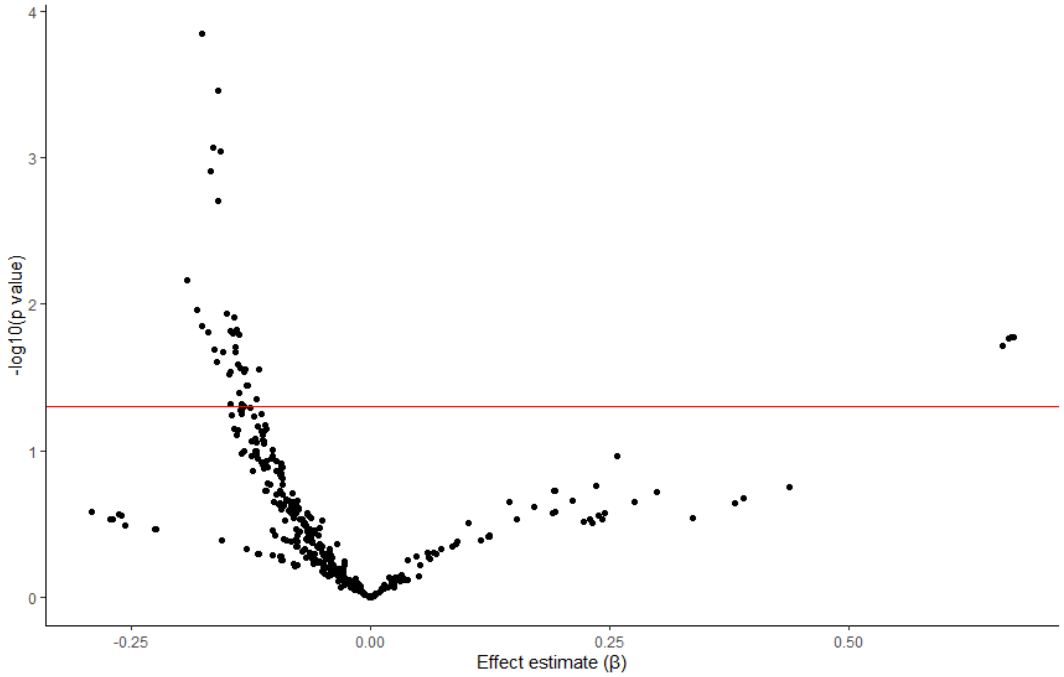
The red line indicates the log10-transformed number corresponding to a *p* value of .05.

Figure 3.3. Specification curve plot of estimates comparing above-guidelines drinkers with a reference group composed of occasional drinkers.



The top section displays the results associated with each universe, ordered by effect size. Negative values indicate drinking is associated with lower hsCRP. The middle section shows which specifications correspond to each result in the top panel. The bottom section displays the number of individuals in the reference and comparator groups in each universe.

Figure 3.4. Vibration of effects volcano plot depicting effects and their *p* values of above-guidelines drinking compared with a reference group composed only of occasional drinkers.



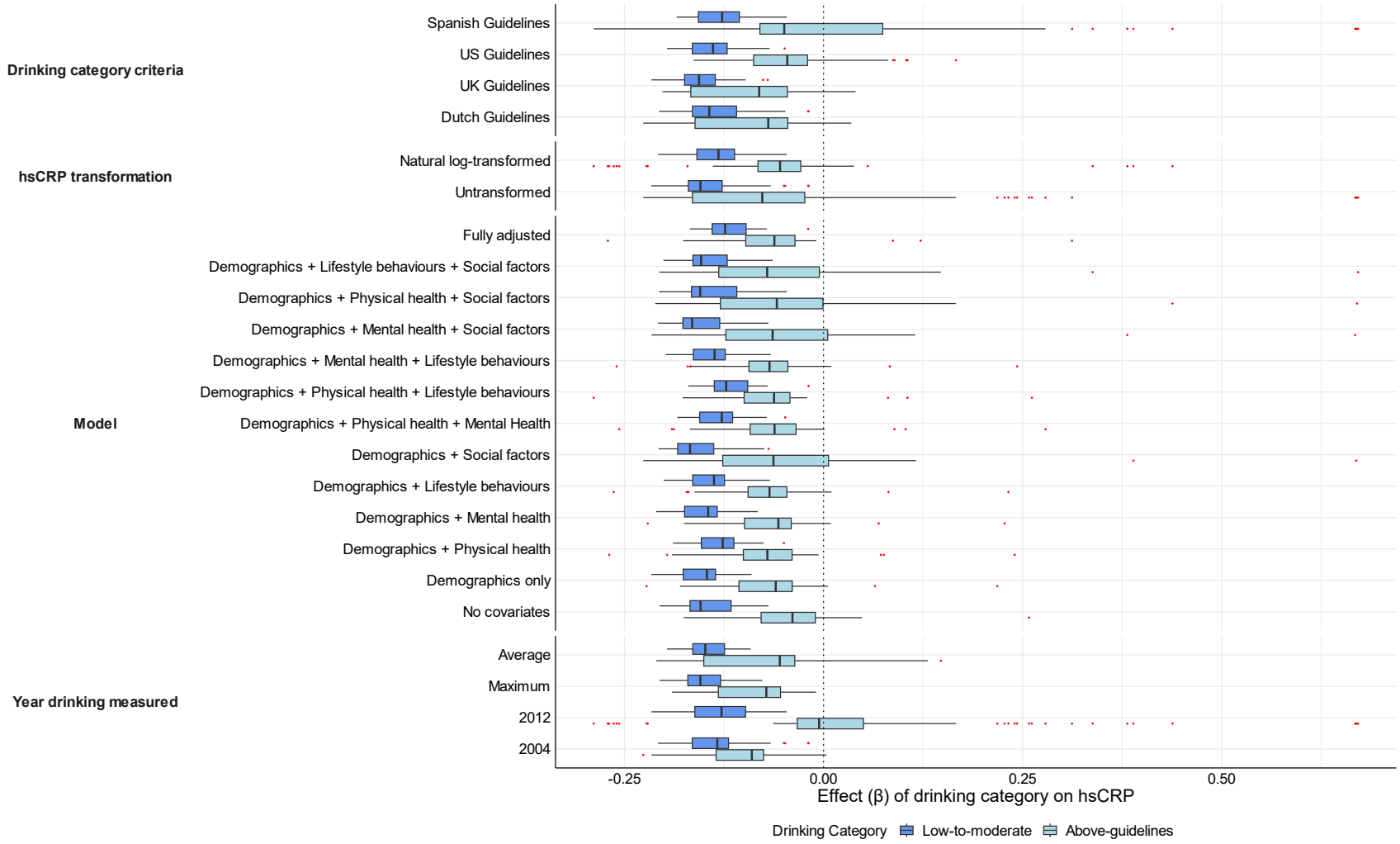
The red line indicates the log10-transformed number corresponding to a *p* value of .05.

3.3.3 Variance decomposition and impact of individual specifications

Using a multi-level model with random effects for each parameter of interest, the highest intraclass correlation coefficient (ICC) for the low-to-moderate drinkers vs occasional drinkers comparison belonged to breadth of covariate control (0.20), followed by which national guidelines were used to define drinking categories (0.13). The highest ICC for the above-guidelines vs occasional drinkers comparison belonged to measurement year that drinking category was based on (0.23), followed by which national guidelines were used to define drinking categories (0.14).

Box and whisker plots (Figure 3.5) offer insights into the impact of particular specifications. Mean protective effects were greater when the ceiling for low-to-moderate drinking was lower. Conversely, the Spanish-based categorisation of above-guidelines drinking – i.e., when this group was restricted to those drinking most heavily – resulted in the smallest mean protective effect. Figure 3.5 also shows that protective effects of above-guidelines consumption appear when considering drinking at age 34 only (12 years before hsCRP measurement) but not drinking at age 42 only (four years before hsCRP measurement).

Figure 3.5. Box and whisker plot depicting impact of specifications on comparisons between low-to-moderate/above-guidelines alcohol consumption and occasional drinking.



Red dots represent outliers.

3.3.4 Post-hoc sensitivity analyses

The multiverse analyses were re-run on a reduced sample, restricted to individuals with no missing covariate data (n=1,358). Results were consistent with the main analyses (Appendices 2.10-2.12).

3.4 Discussion

As inconsistencies in the existing evidence base would suggest, this analysis revealed substantial heterogeneity, or effect vibration, in estimates of the alcohol–hsCRP relationship according to variations in research parameter specifications. Composition of the 'abstinence' reference group had the clearest impact on effect estimates. When analyses were restricted to universes with occasional drinkers as reference, additional impacts of the breadth of covariate adjustment (for the low-to-moderate drinking comparison), as well as measurement year and national guidelines used to classify drinking groups (for the above-guidelines drinking comparison), were apparent.

However, despite heterogeneity in effect and p value size, the direction of results from comparisons of low-to-moderate alcohol consumption with 'abstinence' was robust to parameter specifications. Results were consistent with lower inflammation in the low-to-moderate alcohol consumption group at follow-up. While this does not substantiate a causal role of moderate drinking in reducing inflammation, it fulfils a prerequisite. That is, before determining whether an observed association is causal, that association must be confirmed as

a robust one that holds across methodological variations – what has been labelled the 'consistency' criteria for causation (Hill, 1965).

Plausible biological mechanisms that may account for this association have been suggested, largely derived from short-term administration studies in animal and human models, and *in vitro* work. These centre on low-dose ethanol's ability to inhibit pro-inflammatory cytokines and chemokines (Muralidharan et al., 2014; Szabo, 2007). Additionally, given the established impact of chronic stress on inflammation (Rohleder, 2019), moderate drinking may exert a protective effect mediated by stress reduction.

Interestingly, previous research has shown that excessive alcohol use begins to exert opposite effects on these mechanisms, promoting pro-inflammatory responses (Szabo, 2007). As such, it is ostensibly surprising that the above-guidelines comparisons were less definitive. In fact, those effects with smaller p values belonged to the analyses in which above-guidelines drinking were associated with lower levels of inflammation compared to the reference. Importantly though, the above-guidelines category did not reflect disordered/excessive drinking specifically, which is difficult to capture in general population samples. Given most individuals in this category were not drinking substantially over guidelines (Appendix 2.5), they may still have been drinking at levels below those at which ethanol promotes inflammation. Indeed, specifying above-guidelines drinking cut-points based on Spanish guidelines – i.e., those drinking most heavily – resulted in a bimodal split in effect direction (Figure 3.3), and the only interquartile effect range *not* situated entirely in the direction of reduced inflammation (Figure 3.5).

3.4.1 Strengths and limitations

To the authors' knowledge, this is the first study to examine the longitudinal association between alcohol consumption and a measure of inflammation via an enhanced sensitivity analysis framework. Indeed, it appears to be the first study to have employed enhanced sensitivity frameworks in assessing *any* alcohol–long-term health relationship within a single cohort. Previously, sensitivity considerations had only been explored in limited, disparate ways. The focus of the present study was on key parameters that researchers must consider carefully when engaging in study design and analysis – those which existing evidence indicates stand to impact results and for which diverging specifications are typical in the literature. The present findings can be used to inform both the evidence base on the alcohol–inflammation relationship as well as future research methodology.

A further strength of this study was the large sample size and the rich covariate selection available, including BMI and pre-existing inflammatory conditions. Age, a key determinant of inflammation, was held constant by design of the cohort. Sensitivity analyses of individuals with full covariate data were consistent with the main analyses, increasing confidence that findings were not confounded by sample variation between universes.

However, a limitation of this work was that the clear effects of reference group composition (Appendices 2.6 and 2.8) could not be confirmed to be a result of the characteristics of the groups themselves, as opposed to the small number of lifetime abstainers and former drinkers. Additionally, because BCS70 tested only a single inflammatory marker, the present analyses were restricted to hsCRP as the lone proxy for inflammation. This limits our ability to generalise results to inflammation more broadly; it is possible that other measures may

have shown discrepant results (e.g., differences may be evident between alcohol–CRP and alcohol–interleukin-6 [IL-6] relationships; Silverwood et al, 2014).

Measurement error may also have impacted the findings; while we removed cases with indications of acute inflammation, promoting a focus on chronic low-grade inflammation, hsCRP was only captured at a single time point. Additionally, the effect of certain covariates on hsCRP levels may have been larger had there been a shorter interval between their relative assessments. Also, most data were collected via self-report, which is problematic for the kinds of health covariates employed here (although it is the chief measurement modality in the literature). Further measurement bias was possible given the considerable movement between drinking categories from age 34 to age 42 – possibly explained by UK policy changes beginning in 2003 (Nicholls, 2012) as well as natural 'maturing out' of drinking – suggesting hidden 'drinker drift' was likely in the interim between ages 42 and hsCRP measurement.

Finally, exploring additional parameters of interest, e.g., predominant beverage consumed, proved difficult within the multiverse framework. Although given suggested protective mechanisms operate via ethanol itself (Krenz & Korthuis, 2012) – common to all alcoholic drinks – this should not have considerably impacted results.

3.4.2 Implications

Given the present evidence of a robust association between low-to-moderate drinking and reduced hsCRP, future work should investigate whether this relationship is causal. One such avenue is non-linear Mendelian randomisation; to date one study has found a small protective effect for low-level drinking against CRP, but a positive linear relationship between alcohol and IL-6 (Silverwood et al., 2014). Causal mediation analyses should also be

undertaken to test the hypothesis that inflammation may mediate the relationship between moderate drinking and reduced cardiovascular disease risk. Were moderate alcohol consumption found to have a causal role in lowering inflammation, this would be important information for clinicians and policymakers, including for the generation of low-risk drinking guidelines. However, these findings would have to be balanced against the known harms to other organ systems/disease processes that accrue at even low-level drinking, such as increased cancer risk (Rehm et al., 2019).

The present findings lend support to using occasional/very low volume drinkers as the reference group in future work, especially when the number of true abstainers is very small. While this requires the assumption that such infrequent alcohol consumption would be unlikely to have a biological effect, the present study demonstrates clear benefits in the way of increased statistical power and more normative covariate profiles compared with lifetime abstainers and former drinkers. Results derived using occasional drinker reference groups are also less likely to be contaminated by sick quitters (those for whom illness precipitates abstinence), former drinkers misidentified as lifetime abstainers, or hidden illness belonging to those excluded from the reference group (T. Naimi et al., 2022).

3.4.3 Conclusions

The present study supports a small but robust association between low-to-moderate alcohol consumption and lower levels of the inflammatory marker hsCRP, while the association between drinking at above-guidelines levels and hsCRP demonstrated greater variability. Further research using methodologies that promote causal inference is required. Common variations in research parameters in this field – such as composition of reference group, how

drinking groups are defined, and which covariates are controlled for – are *not* innocuous, and must be considered carefully.

Chapter 4

Does moderate alcohol consumption play a causal role in protecting against depression?: A marginal structural model approach

Preface

Once a stable longitudinal association between an exposure and outcome is established, efforts can be made to ascertain whether there is also causation. Introduced in **Chapter 1**, the marginal structural model (MSM) is an increasingly popular and promising statistical tool for accomplishing this task. In this chapter, a robust MSM approach is applied to the relationship between drinking in early-to-middle adulthood and depression at age 50.

The study focusses on incorporating dynamic, complex variables while producing simple, interpretable estimates for use in health practice and policy. Results resemble J-shaped relationships, contributing preliminary evidence that associations between moderate alcohol consumption and reduced risk for depression may be genuine causal effects.

A slightly modified version of this chapter was published as Visontay, R., Mewton, L., M., Slade, T., Aris, I. M., & Sunderland, M. (2023). Moderate alcohol consumption and depression: A marginal structural model approach promoting causal inference. *American Journal of Psychiatry, 180*(3), 209-17, and is available in Appendix D. Supplementary materials are available in Appendix 3.

4.1 Introduction

Depression contributes substantially to disease burden, particularly in those aged 25-49, where depressive disorders constitute the sixth most burdensome condition globally (Abbafati et al., 2020). This burden is likely to be exacerbated as incidence increases, with cases rising by roughly 50% worldwide from 1990 to 2017 (Q. Liu et al., 2020). On an individual level, depression impairs quality of life, heightens suicide risk (V. Patel et al., 2016), and increases risk for later-life conditions such as cardiovascular disease and associated morbidity and mortality (Seligman & Nemeroff, 2015). Prevention of depressive symptoms and disorders is therefore a public health priority. However, this relies on the accurate identification of modifiable risk factors for the condition – an area where current understanding is limited (Choi et al., 2020).

Alcohol consumption is one candidate worthy of further study, given the well-established comorbidity between alcohol use disorders and depression (Boden & Fergusson, 2011; Brière et al., 2014). Specifically, alcohol use disorders precede major depression (Boden & Fergusson, 2011; Fergusson et al., 2009; Li et al., 2020), with some additional evidence for bidirectionality (Pacek et al., 2013; i.e., depression also being associated with subsequent alcohol use disorders) consistent with the 'self-medication' hypothesis (Khantzian, 1985; Turner et al., 2018).

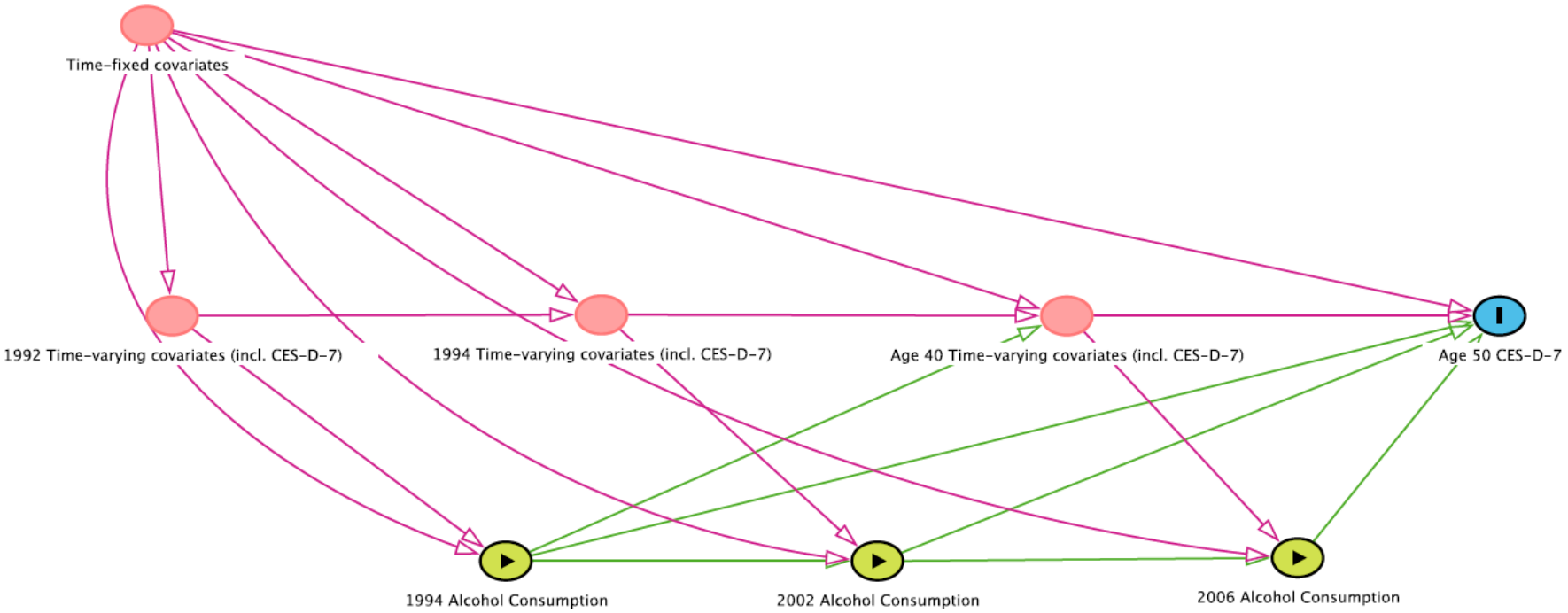
However, for non-disordered drinking at a population level, the nature of the alcohol–depression relationship is less clear. There are indications that light-to-moderate consumption is associated with lower risk for depressive conditions when compared to abstinence, with risk increasing again for heavier drinkers, resulting in a J-shaped

relationship – a finding supported by a recent meta-analysis (Li et al., 2020). However, the extent to which these protective effects are genuine causal relationships, as opposed to biased associations driven by methodological limitations, has not been established.

One such methodological bias may be confounding, given the possibility that background environmental (e.g., SES) factors may explain demonstrated alcohol–depression associations (Keyes et al., 2019). Most studies employ a 'conditional' approach when controlling for confounding i.e., adjusting for covariates in a regression model. To be effective, this approach requires accurate model specification, including the shape of the relationship of each covariate with the outcome (e.g., linear, exponential, etc; Benedetto et al., 2018; Brookhart et al., 2013; Thoemmes & Ong, 2016). Even then, this approach is limited when there are few individuals for certain variable combinations (Greenland et al., 2016; Melberg, 2006; Thoemmes & Ong, 2016), and may be biased when there exists a time-varying factor that can act as both a confounder and mediator over time (Thoemmes & Ong, 2016).

For example, income at some time during the exposure period may be a common cause of later alcohol consumption and depression and may itself be affected by past alcohol consumption. Conditional approaches that control for income at various waves of a study will give biased estimates by removing (i.e., adjusting away) the effects of prior alcohol consumption mediated through income. On the other hand, simply not adjusting for income would lead to results biased by confounding. See Figure 4.1 for a visualisation of these complex longitudinal relationships in the present study.

Figure 4.1. DAG of assumed longitudinal relationships between variables over time



DAG made with the DAGitty web application (Textor et al., 2017). Pink bubbles indicate ancestors of exposure and outcome, green bubbles indicate exposures, and the blue bubble indicates outcome. Pink lines indicate biasing pathways while green lines indicate causal pathways.

In addition to confounding, it is also difficult to isolate effects in a single direction, with reverse causation often present in the form of sick quitters – those for whom illness precipitates abstinence from or a reduction in alcohol consumption (Shaper et al., 1988). Additionally, selection biases may be at play, such as healthier drinkers being over-represented in recruited cohorts (T. S. Naimi et al., 2017), and/or those with poorer outcomes being more likely to drop-out over the follow-up period. Finally, categorisation of alcohol consumption is often based on a single baseline measurement, which ignores changes in consumption ('drinker drift') during the period to outcome measurement (Chikritzhs et al., 2009).

4.1.1 Advantages and assumptions of marginal structural models

Modern statistical methods for causal inference, such as marginal structural models (MSMs), have the potential to overcome the above limitations. MSMs attempt to fully adjust for measured confounders (both time-invariant and time-varying) through a 'marginal' approach i.e., balancing confounders across exposure (alcohol consumption level) groups before regression modelling, accomplished by assigning weights to each observation. These weights reflect the extent to which each observation is under- or over-represented in the sample, with respect to a target population in which potential confounders are balanced across exposure groups. Applying these weights to the sample creates a 'pseudopopulation' in which the distribution of confounders between exposure groups at each individual wave is balanced, while still allowing for longitudinal mediation via these confounders (Robins et al., 2000).

Drawing valid causal inference from MSMs rests on four assumptions: exchangeability, positivity, consistency, and correct specification of propensity score models used in weight

generation (Cole & Hernán, 2008). Exchangeability requires that the risk of an outcome in a particular alcohol consumption group X would have been the same for another consumption group Z, had those in X in fact consumed the same amount as those in Z (Hernán & Robins, 2020). This means there must be no unmeasured confounders and no residual confounding remaining after weighting. Whether exchangeability is met is not strictly testable (Cole & Hernán, 2008). Positivity requires that for any given combination of covariate values, there must be a positive probability of belonging to any of the alcohol consumption categories (as is the case in RCTs, where it is possible for any individual to be assigned to any condition; Hernán & Robins, 2020). This can be checked by inspecting descriptive cross-tabulations of exposure groups and categorical/categorised versions of continuous covariates (Zhong et al., 2018). Consistency requires that the observed outcome under a given exposure is equal to the potential outcome under that same observed exposure (Hernán & Robins, 2020). To satisfy the consistency assumption, there must not be multiple versions of a given exposure group (which may reasonably be questioned in the case of alcohol regarding, for example, different drinking patterns or different beverages consumed). Finally, correct model specification of both the exposure model (for inverse probability of treatment weights; IPTWs) and the censoring model (for inverse probability of censoring weights; IPCWs) is assumed, which can be assessed by inspecting the distribution of stabilised weights (Cole & Hernán, 2008).

Relying on these assumptions, the pseudopopulation should be free of confounding and can be used to approximate RCTs – the 'gold standard' for assessing causality – in contexts where their implementation is not feasible (Vandecandelaere et al., 2016). This approach can address many of the limitations of conventional cohort studies, as detailed in Table 4.1.

Table 4.1. Ways in which MSMs can improve causal inference

Common limitations of past research	MSMs:
Traditional confounder adjustment relies on making accurate modelling assumptions about confounder-outcome relationships (Benedetto et al., 2018; Brookhart et al., 2013; Thoemmes & Ong, 2016)	Instead require modelling confounder-treatment relationships, with the ability to include multiple functional forms
Adjusting only for baseline covariates ignores time-varying confounders, while standard control for time-varying covariates will block mediated effects and/or induce collider bias (Vandecandelaere et al., 2016; Williamson & Ravani, 2017)	Can incorporate time-varying covariates that act as both mediators and confounders
Difficulty in isolating effects in the direction of interest (i.e., alcohol's effect on depression isolated from depression's effect on alcohol consumption)	Can incorporate multiple measurements of both the exposure and outcome over time, accounting for time-varying exposures being affected by past outcome levels
Including only baseline alcohol consumption induces exposure misclassification, ignoring variability in consumption over follow-up (Chikritzhs et al., 2009)	Can incorporate multiple measurements of exposure over time
Selection biases are common (T. S. Naimi et al., 2017), including over-representation of healthy drinkers and differential attrition	Can incorporate both survey weights (making the baseline sample representative of the population) and censoring weights (making the final sample representative of baseline)

4.1.2 MSMs and the alcohol–depression relationship

Chapter 2 identified a single study applying MSMs to the alcohol–depression relationship (Visontay et al., 2022). In a sample aged 20-64, Gémes et al. found a U -shaped relationship with major depression after 7-9 years of follow-up (Gémes et al., 2019). However, to determine whether the Gémes et al. findings are truly robust to the methodological biases outlined above, an extension of the application of MSMs is required.

First, to utilise the full potential of MSMs and acknowledge drinker drift, multiple waves of exposure and covariate information should be included. Additionally, it is vital that a substantial portion of participants' drinking (or abstinence) histories are incorporated, lest the complex cumulative and/or delayed effects of alcohol use (Kerr et al., 2019) or the effects of sick quitters adulterate estimates of the targeted exposure's effect. Similarly, given that the nature of alcohol–depression relationships may change over the lifespan (Alati et al., 2005), estimates should be derived from a sample of fairly narrow age range so as not to mask underlying heterogeneity. Finally, exposure weights should only incorporate time-varying covariates that are known to temporally precede exposure.

This study was designed with the 'target trial' approach in mind (Hernán & Robins, 2016). This approach holds that valid causal inference from observational research necessitates conceiving of a target trial – an ideal, hypothetical RCT probing the question of interest (Hernán & Robins, 2020). In this framework, the observational analysis – including choice of eligibility criteria, treatment definition, follow-up period, outcome definition, and causal contrasts of interest – should be conducted according to the nature of the target trial the researcher is trying to emulate (Hernán & Robins, 2016). Among other considerations for a target trial of the alcohol–depression relationship, the intervention conditions in an ideal RCT would comprise *consistent* levels of consumption. Questions about the protective effects of moderate consumption tend to imply a stable level of consumption (e.g., 'Does moderate consumption confer benefits over abstinence?'), allowing for simple, interpretable public health messages.

With a secondary analysis of the National Longitudinal Survey of Youth (NLSY79) – a cohort offering rich, longitudinal data, the present study aimed to employ a robust MSM

approach in addressing the following research question: compared with consistent abstinence from alcohol, what is the effect of consistent occasional, moderate, and above-guidelines alcohol consumption throughout early-to-middle adulthood on depression at age 50? Given evidence that methodological biases are responsible for previously-identified J-shaped relationships, it was hypothesised that the use of a sophisticated MSM approach – reducing the impact of these biases – would result in a positive, linear relationship.

4.2 Methods

4.2.1 Study design and participants

The US-based NLSY79 is a representative cohort of 12,686 individuals born between 1957 and 1964 (the sampling procedure for which has been described elsewhere; Rothstein et al., 2019). Participants were first interviewed in 1979, when aged between 14 and 22. Data on depressive symptoms (as indexed by the Centre for Epidemiological Studies-Depression Scale short form; CES-D-SF) were first collected in 1992. Given the need to control for pre-baseline depressive symptoms, baseline for the current study was set at the following measurement occasion (1994). Eligible individuals were those possessing valid baseline alcohol data, and both time-fixed and pre-baseline (1992) time-varying covariate data (including CES-D-SF values), resulting in 5,667 participants. All NLSY79 participants provided informed consent.

Guided by the timing of CES-D-SF re-administration, which changed after 1994 from year-based measurement to age intervals, the MSM incorporated alcohol consumption at 1994,

2002 and 2006, covariates/CES-D-SF values in 1992, 1994 and at age 40, and a final CES-D-SF outcome measured at age 50 (see Appendix 3.1 and Figure 4.1 for more detail and a visualisation, respectively, of what data was used from which assessment points). Thus, the alcohol consumption variables span ages 29-37 through 41-49, corresponding with early-to-middle adulthood. In addition to those who dropped out entirely, individuals missing any alcohol consumption, covariate, or CES-D-SF values at 2002 or 2006, or CES-D-SF values at age 50, were considered censored at the earliest of those assessments.

4.2.2 Variable derivation

4.2.2.1 Alcohol consumption

Participants were asked about their current drinking frequency (number of drinking days in previous month), volume (number of drinks per drinking day, with a drink explicitly defined as "the equivalent of a can of beer, a glass of wine, or a shot glass of hard liquor"), and HED (defined as at least 6 drinks per occasion by the NLSY79 questionnaire). For this study, alcohol consumption was separately categorised for 1994, 2002, and 2006, with categorisation based on methods (Calvo et al., 2021) incorporating frequency, volume, and HED, modified such that volume criteria concorded with current US guidelines (see Table 4.2 for category criteria; US Department of Agriculture and US Department of Health and Human Services, 2020). For every pre-baseline occasion with adequate information, individuals were similarly categorised, and wave-specific categorisations condensed into a single historical variable (see detail related to derivation of baseline time-fixed variables in Appendix 3.1).

Table 4.2. Criteria for alcohol consumption categories

Category	**Criteria**
Abstinence	No alcohol consumption
Occasional consumption	Average frequency <1 day/week; no HED
Moderate consumption	Average frequency >= 1 day/week; average weekly drinks <=7 for females/<=14 for males; no HED
Above-guidelines consumption	Average frequency >= 1 day/week; average weekly drinks >=7 for females/>=14 for males; and/or HED

HED = heavy episodic drinking.

4.2.2.2 Covariates and CES-D-SF scores

The CES-D was designed to measure depressive symptoms in the general population and asks respondents to indicate the frequency with which they experienced a range of symptoms in the previous week, with higher scores suggesting greater symptoms (Radloff, 1977). The CES-D-SF is a reduced, seven-item version of the CES-D, and has demonstrated strong psychometric properties (Levine, 2013). Two versions of the analyses were performed: one in which CES-D-SF scores were analysed in their continuous form and another in which the scale was dichotomised into non- (<8) or probable (>=8) depression based on suggested cut-offs for probable depression (Levine, 2013). The distribution of the outcome variable was not sufficiently skewed to warrant transformation (Kim, 2013; see Appendix 3.1).

Covariate selection was guided by a recent meta-analysis evaluating the relationship between alcohol consumption and depressive symptoms (Li et al., 2020). Covariates controlled for in the meta-analysis' included papers were extracted and tallied (see Appendix 3.2). Twelve of the 16 variables most frequently controlled for were available in/could be derived from the present dataset (the four others were physical activity, diet, childhood adversities, and familial history of alcohol/mental health problems). Given other variables may help in predicting exposure group membership, several additional variables were also incorporated.

The baseline time-fixed variables included were: historical alcohol consumption, sex, ever smoked, ever used illicit drugs, pre-baseline self-esteem, pre-baseline frequent religious service attendance, race, educational attainment, and average parental educational attainment. The time-varying variables incorporated were: CES-D-SF scores, age, self-reported health limitations, marital status, smoking status (N/A at age 40), illicit drug use (N/A at age 40),

BMI, employment status, health insurance status, household size, urban/rural-dwelling, welfare receipt, and income. Note that the incorporation of CES-D-SF scores prior to the final outcome measurement means that the model specifically isolates the effect of alcohol consumption on age 50 depression. Further detail on covariate selection and derivation is presented in Appendix 3.1.

4.2.3 Statistical analysis

4.2.3.1 MSMs

The first step in constructing MSMs is to sequentially estimate IPTWs for each wave – i.e., the probability of an individual being in a given treatment (or exposure) group given how their covariate profile compares to the typical profile for that group. Individuals whose covariate profiles are atypical of their exposure category are assigned larger weights than those whose covariate profiles align more closely with their group's typical profile. Inverse probability of censoring weights (IPCWs) are calculated at each wave too, such that those whose covariate profiles are not typical of those who remained in the study are upweighted. Combined with the use of a survey weight, IPCWs allow MSMs to be representative of the target population. The final weight for each individual is the product of all time-specific IPTWs, IPCWs, and the survey weight.

In the second step, the weighted pseudopopulation is statistically analysed, comparing hypothetical joint interventions (i.e., exposure trajectories of interest) in their effect on an outcome. Such an analysis provides marginal estimates that apply to the whole population because these estimates are not conditional on other variables (i.e., they are not based on models that hold covariates constant; Williamson & Ravani, 2017).

4.2.3.2 Weight generation

Stabilised IPTWs were created for alcohol consumption at the 1994, 2002, and 2006 waves, as were IPCWs (relating to attrition status at 2002, 2006, and age 50). Survey weights for all individuals were obtained from the NLSY79 website (https://www.nlsinfo.org/weights/nlsy79), and were then normalised.

The purpose of the IPTWs was to reweight the sample at each timepoint such that covariates were no longer associated with exposure category. Residuals from generalised linear models regressing alcohol consumption on covariates indicated evidence of statistically significant non-linearity between household size and alcohol category, as well as depressive symptoms and alcohol category (not at all waves for either), so quadratic terms were added in the IPTW models for these variables. Individuals' normalised survey weights were incorporated in the PS model (Dugoff et al., 2014; Ridgeway et al., 2015). IPTWs were stabilised by incorporating prior time-varying alcohol exposure values in the numerator, for each person not yet censored.

The below gives the formula for the product of all IPTWs for an individual, where R is exposure category, C is attrition/censoring, $\bar{R}$ is exposure history, L is time-varying covariates, and X is baseline covariates:

$$W_i^{\bar{r}} = \prod_{t=1}^{t=3} \left\{ \frac{P(R_{i,t} | \bar{C}_t = 0,\ \bar{R}_{t-1} = \bar{R}_{i,t-1},)}{P(R_{i,t} | \bar{C}_t = 0,\ \bar{R}_{t-1} = \bar{R}_{i,t-1},\ \bar{L}_{t-1} = \bar{L}_{i,t-1},\ X = x_i\ )} \right\}$$

Stabilised IPCWs were also created for each of these waves, in order to reweight the sample at each timepoint so that covariates were no longer related to attrition status at the

following wave (i.e., 2002, 2006, age 50). The below gives the formula for the product of all IPCWs for an individual (same notation as above, and where $\bar{C}$ is censoring history):

$$W_i^c = \prod_{t=1}^{t=3} \left\{ \frac{P(\bar{C}_{t+1}=0 | \bar{C}_t=0, \bar{R}_t=\bar{R}_{i,t,})}{P(\bar{C}_{t+1}=0 | \bar{C}_t=0,\ \bar{R}_t=\bar{R}_{i,t},\ \bar{L}_t=\bar{L}_{i,t},\ X=x_i\ )} \right\}$$

The final weight for each individual was the product of their three IPTWs, three IPCWs, and survey weight. These final weights were trimmed at the 98th percentile to mitigate the effect of extreme weights (Aris et al., 2021), resulting in a set of weights in the final analysed sample ranging from 0.07 to 6.19, with a mean of 1.16 and standard deviation of 1.27. As illustrated in Appendix 3.3, the final weight greatly improved covariate balance between alcohol consumption groups at each timepoint, with mean differences between groups for most covariates falling within the conservative .1 rule-of-thumb after weighting.

4.2.3.3 Linear regression and contrasts

Using these final weights, weighted multivariable regression models regressed CES-D-SF scores (linear model)/probable depression (logistic model) at age 50 on 1994, 2002, and 2006 alcohol consumption. The first set of analyses, using a continuous CES-D-SF outcome, provide predicted mean depressive symptoms, while the second, using a binary outcome, provide predicted probabilities of probable depression. In results focussing on stable trajectories of alcohol consumption, drinking level 'values' were substituted into a model using the parameters derived from the whole sample to generate predictions about *hypothetical* drinking trajectories of interest. This involved using linear contrasts to estimate the predicted difference in outcome between hypothetical trajectories of consistent consumption.

Specifically, consistent occasional, consistent moderate, and consistent above-guidelines drinking were compared against a consistent abstinence reference category. We selected these stable drinking categories because they are particularly informative in terms of formulating national alcohol guidelines. For the continuous outcome, this is equivalent to summing the relevant wave-specific effect estimates for each category of alcohol consumption from the initial regression, and then taking the difference with the reference group. Results were bootstrapped using 500 bootstrap replicates to account for the additional sources of uncertainty introduced by weighting.

4.2.3.4 Sensitivity analyses

Despite incorporating a wide range of potential confounders, E-values were calculated for the contrast estimates to assess robustness to unmeasured confounding. E-values represent the effect size that an excluded confounder must have on both exposure and outcome to account for the observed association (VanderWeele & Ding, 2017). The main analyses were additionally performed in men and women separately.

All analyses were conducted in R version 4.1.0 (see Appendix 3.1 for packages and code used). $p<0.05$ was considered statistically significant, but results are also interpreted with respect to strength of the effect estimates and confidence intervals.

4.3 Results

4.3.1 Baseline sample characteristics

The sample was 50.5% female, and in 1992 had a mean age of 30.81. At baseline, 1,947 abstained from alcohol, 1,259 consumed occasionally, 660 consumed moderately, and 1,801 consumed above guidelines. The mean CES-D-SF score in 1992 was 4.16 (SD 4.01), with baseline abstainers having the highest mean (4.55; SD 4.36) and baseline moderate drinkers the lowest (3.17; SD 3.33). Compared to abstainers and above-guidelines drinkers, occasional and moderate drinkers had higher mean income at baseline, were less likely to be Black, and more likely to be in employment and to be married. Further baseline sample characteristics are presented in Table 4.3 (note that IPTW weighting effectively nullified baseline differences between exposure groups; see Appendix 3.3).

Table 4.3. Selected Baseline Characteristics of Eligible Participants, by Alcohol Consumption at Baseline

Characteristic[a]	**Abstainer (N=1,947)**		**Occasional (N=1,259)**		**Moderate (N=660)**		**Above-guidelines (N=1,801)**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CES-D-SF	4.55	4.36	3.83	3.75	3.17	3.33	4.32	3.93
Age	30.81	2.24	30.90	2.22	31.03	2.30	30.67	2.21
Income (nearest US$)	16,005	12,938	20,745	13,213	24,084	15,567	17,541	13,408
BMI	26.48	5.49	25.65	5.03	24.72	4.32	26.18	4.66
	N	%	N	%	N	%	N	%
Sex (Female)	1187	61.0	778	61.8	305	46.2	592	32.9
Historical alcohol consumption								
Lifetime abstainer	80	4.1	1	0.1	0	0.0	0	0.0
Past occasional drinker	502	25.8	238	18.9	30	4.5	53	2.9
Past moderate drinker	196	10.1	201	16.0	134	20.3	59	3.3
Past above-guidelines drinker	1169	60.0	819	65.1	496	75.2	1689	93.8
Race								
Black	617	31.7	298	23.7	144	21.8	476	26.4
Hispanic	376	19.3	256	20.3	77	11.7	324	18.0
Non-Black and Non-Hispanic	954	49.0	705	56.0	439	66.5	1001	55.6
Current smoking (Yes)	615	31.6	357	28.4	178	27.0	885	49.1

Employment status (Unemployed)	579	29.7	241	19.1	81	12.3	367	20.4
Marital status								
Married	1077	55.3	778	61.8	396	60.0	826	45.9
Separated/divorced/widowed	340	17.5	162	12.9	97	14.7	334	18.5
Never married	530	27.2	319	25.3	167	25.3	641	35.6

[a] Baseline alcohol categorisation is based on 1994 values; time-varying characteristics are based on 1992 values.

4.3.2 Primary analyses

3,593 individuals provided valid outcome data at age 50. The results of the regression models are presented in Table 4.4. For both the continuous and binary outcomes, the largest protective effects for both occasional (b=-0.61, bootstrap CI= -1.19, -0.01; OR=0.70, bootstrap CI=0.46, 1.04) and moderate (b=-0.82, bootstrap CI=-1.46, -0.13; OR=0.58, bootstrap CI=0.29, 1.00) drinking corresponded to 1994 consumption, while the greatest detrimental effects for above-guidelines drinking corresponded to 2006 consumption (b=0.55, bootstrap CI=-0.98, 0.40; OR=1.17, bootstrap CI=0.72, 1.96).

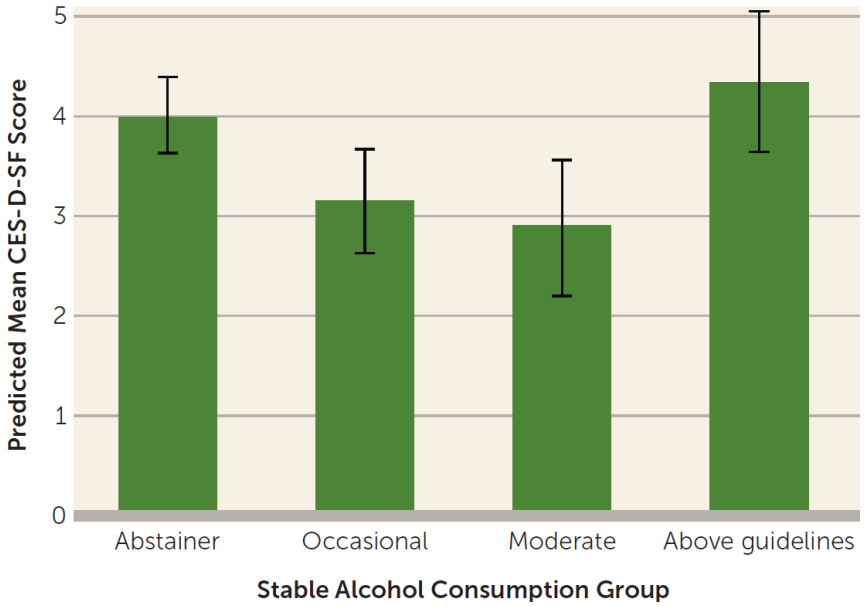
Table 4.4. Output from outcome regressions of Age 50 CES-D-SF on alcohol category predictors

	Mean CES-D-SF score analyses				**Probable depression analyses**			
Predictors	**Estimates** (b)[a]	**CI**	***P***	**Bootstrapped CI**	**Estimates** (OR)[a]	**CI**	***P***	**Bootstrapped CI**
Intercept	3.99	3.70, 4.28	<0.001	3.63, 4.39	0.24	0.20, 0.28	<0.001	0.18, 0.30
1994_occasional	-0.61	-1.01, -0.21	0.003	**-1.19, -0.01**	0.70	0.55, 0.89	0.004	0.46, 1.04
1994_moderate	-0.82	-1.33, -0.32	0.001	**-1.46, -0.13**	0.58	0.42, 0.79	0.001	0.29, 1.00
1994_above-guidelines	-0.11	-0.52, 0.30	0.596	-0.71, 0.50	0.95	0.75, 1.19	0.630	0.66, 1.32
2002_occasional	-0.01	-0.40, 0.39	0.977	-0.61, 0.62	0.99	0.78, 1.26	0.957	0.64, 1.43
2002_moderate	-0.01	-0.49, 0.47	0.974	-0.67, 0.76	1.34	1.01, 1.76	0.042	0.80, 2.15
2002_above-guidelines	-0.09	-0.59, 0.41	0.718	-0.88, 0.80	0.96	0.71, 1.29	0.788	0.55, 1.57
2006_occasional	-0.22	-0.61, 0.16	0.259	-0.73, 0.34	0.83	0.65, 1.04	0.111	0.55, 1.25
2006_moderate	-0.25	-0.71, 0.21	0.290	-0.98, 0.40	0.77	0.58, 1.01	0.065	0.45, 1.29
2006_above-guidelines	0.55	0.04, 1.06	0.036	-0.24, 1.43	1.17	0.87, 1.55	0.296	0.72, 1.96
Observations	3593				3593			
R2 / R2 adjusted	0.011 / 0.009				-			

[a] Reference group for each wave is abstinence.

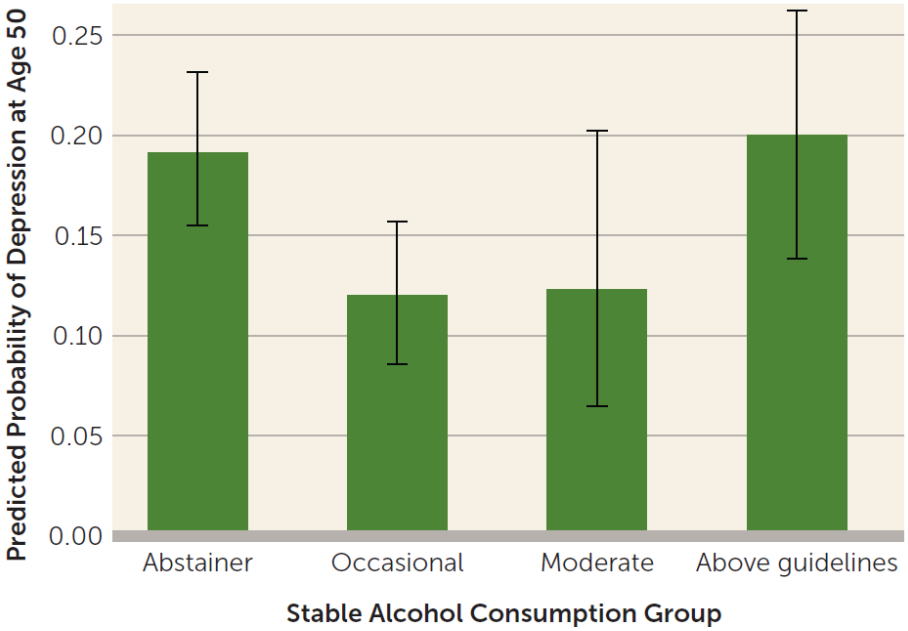
Results from the contrast analyses approximated a J-shape (see Figures 4.2 and 4.3): after bootstrapping, consistent occasional (b=-0.84, CI= -1.47, -0.11) and consistent moderate (b=-1.08, CI=-1.88, -0.20) drinkers were both predicted to have statistically significantly lower mean CES-D-SF values at age 50 than consistent abstainers. Consistent above-guidelines drinkers had greater predicted CES-D-SF values than consistent abstainers, but the difference was not statistically significant (b=0.34, CI= -0.47, 1.25). Similar results were obtained for the binary outcome analyses, with consistent occasional drinkers predicted to have significantly lower odds of depression than consistent abstainers at age 50 (OR=0.58, CI=0.36, 0.88), and consistent above-guidelines drinkers having marginally higher odds (OR=1.06, CI=0.66, 1.72). Consistent moderate drinkers had reduced odds of developing probable depression comparable to that of occasional drinkers, but this was not significant after bootstrapping (OR=0.59, CI=0.26, 1.13).

Figure 4.2. Predicted mean CES-D-SF scores at age 50 for individuals consistently abstaining or drinking occasionally, moderately, or above guidelines over the 1994, 2002, and 2006 measurement waves in a representative U.S. sample.



Groups do not possess a specific number of subjects as they represent hypothetical trajectories. Error bars indicate bootstrapped 95% confidence intervals. CES-D-SF=Center for Epidemiologic Studies Depression Scale–Short Form. To view predicted means for all hypothetical trajectories, please visit the R Shiny applications that accompany this chapter (males: https://rachelvisontay.shinyapps.io/Men_continuous/; females: https://rachelvisontay.shinyapps.io/Women_continuous/).

Figure 4.3. Predicted probability of probable depression at age 50 for individuals consistently abstaining or drinking occasionally, moderately, or above guidelines over the 1994, 2002, and 2006 measurement waves in a representative U.S. sample.



Groups do not possess a specific number of subjects as they represent hypothetical trajectories. Error bars indicate bootstrapped 95% confidence intervals. CES-D-SF=Center for Epidemiologic Studies Depression Scale–Short Form. To view predicted means for all hypothetical trajectories, please visit the R Shiny applications that accompany this chapter (males: https://rachelvisontay.shinyapps.io/Men_probable_depression/; females: https://rachelvisontay.shinyapps.io/Women_probable_depression/).

See Appendices 3.4 and 3.5 for full output.

4.3.3 Sensitivity analyses

E-values for the consistent abstinence vs. consistent moderate consumption contrast in the continuous outcome analyses were 1.83 (for a standardised β of -0.25) and 1.30 (for the confidence interval). This indicates that to fully account for the observed association, an unmeasured confounder(s) would need to almost double both the probability of an individual belonging to the moderate exposure group and the probability of being high (compared to low) on CES-D-SF scores (with a smaller effect required to alter the confidence interval such

that it includes the null). The E-value was similar for the binary outcome analyses (1.93). For the consistent abstinence vs. consistent occasional consumption contrast, the E-values for mean depression score were 1.68 and 1.19 for a standardised β of -0.20 and confidence interval respectively, and for the binary outcome were 1.95 for the effect and 1.33 for the confidence interval.

Women were less likely to be above-guidelines drinkers than men (see Table 4.3) and had higher depressive symptoms overall: at age 50 the mean for women was 4.20 (SD 4.48) compared to 3.05 (SD 3.91) for men. However, in stratified analyses, the pattern of results from the overall sample replicated for both men and women (see Appendix 3.1 for methods detail; Appendices 3.6 and 3.7 for plots of results). This was true of both the continuous and binary outcome analyses, although for the latter, consistent moderate drinking resulted in the lowest predicted probability for males rather than being equivalent with consistent occasional drinking as it was for females.

4.4 Discussion

This MSM analysis of the relationship between alcohol consumption in early-to-middle adulthood and depression at age 50 replicated the classic J-shaped relationship reported in the conventional epidemiological literature. In the present study, consistent occasional or moderate drinking was predicted to result in lower CES-D-SF mean scores and reduced odds of probable depression than consistent abstinence.

This finding was contrary to the hypothesis that an MSM using methods that address potential biases would result in a positive, linear relationship between alcohol use and

depression. That protective effects of low-to-moderate consumption were found even after accounting for common biases offers preliminary evidence that these effects may be causal. Unlike the harmful effects of above-guidelines consumption, which were largest at the wave just prior to outcome measurement, alcohol consumption at baseline contributed most strongly to the protective effects of occasional and moderate drinking (while moderate drinking at 2002 was associated with *increased* risk in the binary outcome analyses, this estimate had wide confidence intervals after bootstrapping). This suggests the benefits of low-level drinking accrue over time.

Various mechanisms have been suggested for these benefits, including possible biological mediation via both GABAergic (Bellos et al., 2016; Steffensen et al., 2009) and dopaminergic effects (Tizabi et al., 2018). Moderate alcohol consumption is also associated with increased levels of brain-derived neurotrophic factor (BDNF), as well as lower levels of inflammatory biomarkers (Bell et al., 2017; Imhof et al., 2001; Tizabi et al., 2018), which are both implicated in depression (Milaneschi et al., 2021). However, given moderate drinking is known to reflect/facilitate healthy social interaction (Peele & Brodsky, 2000) – an established protective factor for depression (Cruwys et al., 2013) – this should also be acknowledged as a key pathway for observed benefits (Gémes et al., 2019).

Analyses using continuous CES-D-SF scores maximised statistical power and offer population-level interpretability. While applying effect size heuristics to psychological research is imperfect given very small/small effects are typical (Owens et al., 2021), the protective effects found in the continuous outcome analyses were small (Cohen, 2013). At first glance, these may not represent a clinically meaningful benefit, but given evidence that increased depressive symptom severity is associated with risk for subsequent physical health

conditions (Seligman & Nemeroff, 2015), changes in mean scores – even if below clinical thresholds – are important to study.

Further, these effects are consequential at the population level, particularly when contributing to modelling that informs national drinking guidelines, where the incorporation of evidence of alcohol's protective effects can substantially alter recommendations (National Health and Medical Research Council, 2020). Indeed, the present findings offer support for the current US drinking guidelines (US Department of Agriculture and US Department of Health and Human Services, 2020), at least with respect to depression.

The second set of analyses, using a binary outcome, provided predicted probabilities of probable depression and maximise clinical interpretability. These analyses also supported protective effects for stable occasional and moderate consumption (although due to reduced statistical power, the moderate–abstainer contrast was no longer statistically significant). However, any evidence for protective effects must be balanced against that demonstrating increasing dose-response relationships between alcohol consumption and other health outcomes, including several cancers (Rehm et al., 2019).

While consistent above-guidelines drinking was associated with increased depressive symptoms and risk for probable depression, this increase was modest and not statistically significant. At higher levels, alcohol begins to exert opposite effects on the same pathways via which lower consumption may offer protection (Tizabi et al., 2018). As such, it is ostensibly surprising that consistent above-guidelines drinkers did not have greater depression risk. However, recent meta-analytic evidence has suggested that the increased risk of depression among heavy drinkers may be largely driven by confounding (Li et al., 2020).

The findings from the current study are consistent with this; only modest increases in depression were found among the above-guideline drinkers after rigorously controlling for available confounders. Importantly, evidence still indicates that *disordered* drinking more specifically is associated with increased depression risk (Li et al., 2020).

That the results of the primary analyses replicated when the sample was stratified by sex is consistent with previous findings that, despite sex-based differences in the epidemiology of alcohol consumption and depression more generally, the nature of the alcohol–depression relationship is similar (Gémes et al., 2019; Liang et al., 2021).

4.4.1 Strengths

The application of causal inference methods in this area is a nascent field – the present study contributes preliminary evidence that associations between moderate alcohol consumption and reduced depression may be genuine causal effects. The key strength of MSMs in this context is in generating effect estimates which account for the dynamic nature of alcohol consumption and confounders over time. Studies that ignore drinker drift are likely to produce biased estimates, and while using latent class growth models (LCGM) to characterise trajectories of consumption avoids this problem, it relies on modelled classes rather than observed patterns in the actual data and cannot answer questions about stable levels of exposure. This also applies to emerging methods combining LCGM and MSMs (Diop et al., 2021).

The present study controlled for a wide range of variables commonly considered to confound the alcohol–depression relationship. This included pre-baseline alcohol consumption, which was key to mitigating sick quitter bias given 60% of baseline abstainers were previous above-

guidelines drinkers. Robust pre-baseline data, as well as the incorporation of multiple waves of alcohol consumption over follow-up, represent considerable improvements over the only other extant MSM analysis in this space.

4.4.2 Limitations

MSMs, like any statistical model, depend on the quality and breadth of confounder measurement. The breakdown of exposure group by covariates in Table 4.3 illustrates that individual covariates – excepting prior alcohol consumption – tended to only modestly associate with baseline drinker category. This may indicate that the identified protective effects are unlikely to be solely a result of unmeasured confounding. That said, those trends that were evident, such as occasional and moderate drinkers having higher income and employment levels, constitute important social determinants of health. It is plausible that one of/a combination of similar unmeasured covariates may have effects on exposure and outcome exceeding the generated E-values.

In particular, the present study lacked data on parental alcohol use disorder, as well as sophisticated measures of social support and interaction, which may confound (and mediate) the alcohol–depression relationship in a longitudinal context. Residual confounding may also be possible given that weighting for the 2002 and 2006 exposure groups did not achieve balance for all covariates within the conservative 0.1 standardised mean difference heuristic (see Appendix 3.3). Moreover, the desired interpretation of the weighted estimates rests on strong assumptions that, at best, approximately hold.

4.4.3 Future directions

More research applying novel approaches to the alcohol–depression relationship, such as flexible, non-linear Mendelian Randomisation (which can eschew confounding entirely), is needed to answer remaining questions around causality. Alongside this, triangulation of evidence across various observational designs and statistical methods is key. Due consideration must also be given to the manifold data processing and analysis decisions researchers make, the impacts of which can be quantified using multiverse analysis (Steegen et al., 2016).

4.4.4 Conclusions

In a representative US cohort, occasional and moderate drinking throughout early-to-middle adulthood predicted a small reduction in depression at age 50 when compared with abstinence. Given the importance of identifying modifiable risk and protective factors for depression, more research is required to establish that the nature of this relationship is a causal one.

Chapter 5

Does low-level drinking protect against type 2 diabetes?: Applying sophisticated genetic-based methods

Preface

Marginal structural models and other G-methods represent some of the most promising statistical approaches to improve causal inference from observational data. However, as discussed in **Chapter 4,** even these have limitations – most notably, vulnerability to unmeasured confounding. As introduced in **Chapter 2**, design-based methods for promoting causal inference can often avoid confounding entirely. This chapter reports on the application of one such method, non-linear Mendelian Randomisation (MR), to the analysis of the alcohol–type 2 diabetes relationship.

Results suggest that low-level alcohol consumption may have causal benefits for average glucose levels, but a protective effect against risk for developing diabetes is more ambiguous and at best is confined to women consuming small volumes. It is also apparent that observational evidence substantially overestimates the benefits, and underestimates the harms, of alcohol for diabetes risk.

This chapter is in preparation for submission to *JAMA Network Open*. Supplementary materials are available in Appendix 4.

5.1 Introduction

Type 2 diabetes (T2D) is a chronic health condition characterised by insulin resistance, deficiencies in insulin production, and elevated blood glucose. It is associated with substantial and rapidly increasing disease burden on a global scale (Tinajero & Malik, 2021). Despite the strong causal contribution of poor diet to T2D risk, observational evidence tends to suggest a non-linear relationship between alcohol consumption and the condition, such that low-level drinking is associated with reduced risk compared with abstinence (Griswold et al., 2018; Knott et al., 2015; Rehm, Gmel Sr, et al., 2017).

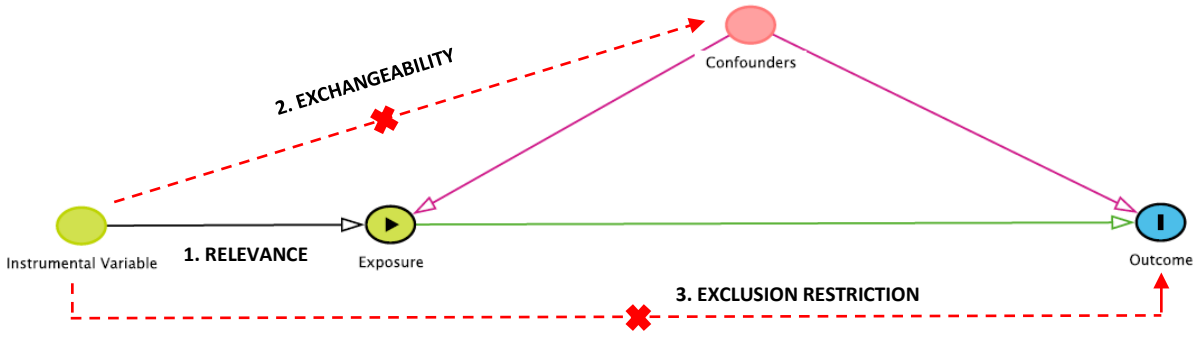
These findings are incorporated into estimates of alcohol's disease burden (Griswold et al., 2018) and inform public health policy and low-risk drinking guidelines, e.g., those of Australia's National Health and Medical Research Council (2020), often tempering evidence of harms for other health outcomes that accrue from low-level drinking. However, as for a host of other outcomes where protective effects of alcohol have been observed, it is unclear whether these findings reflect true causal relationships, or merely artefacts of confounding and reverse causation (Knott et al., 2015).

RCTs offer evidence that, in the acute-to-short term, low-level ethanol has positive effects on biomarkers associated with glucose metabolism, e.g., postprandial glucose response (Brand-Miller et al., 2007) and oxidative stress (Ma et al., 2022), glycated haemoglobin (HbA1C; used also for the diagnosis of diabetes), and fasting insulin (Schrieks et al., 2015). There are also indications from intervention trials that alcohol may increase adiponectin, which regulates insulin sensitivity, and promote high-density lipoprotein cholesterol (Beulens et al., 2008; Brien et al., 2011) – both of which are implicated in T2D. However, not all evidence is

consistent, and whether these benefits persist in the long-term is unknown (Ma et al., 2022). Indeed, RCTs of alcohol's effects on long-term metabolic changes, and more importantly risk for T2D itself, are all but non-existent (restricted to a single two-year study in diabetics offering mixed findings; Gepner et al., 2015).

Given the ethical and logistical difficulties of conducting long-term RCTs,, methods to promote causal inference from observational cohorts are critical. Of particular promise are instrumental variable (IV) designs. An IV is a proxy for an exposure of interest, but theoretically avoids any confounding or reverse causation that the exposure could be subject to. As captured in Figure 5.1, IV designs rely on three assumptions (Burgess & Thompson, 2021): assumption 1 is that the IV is associated with the exposure ('relevance'), assumption 2 is that the IV is *not* associated with confounders of the exposure–outcome relationship ('exchangeability'), and assumption 3 is that the IV affects the outcome *only* via its effect on the exposure ('exclusion restriction').

Figure 5.1. DAG of an instrumental variable design.



5.1.1 An overview of the Mendelian Randomisation approach

The technique of Mendelian Randomisation (MR) is a kind of IV design that uses genes (specifically, single-nucleotide polymorphisms; SNPs) as the instrument/s. While single SNPs are sometimes sufficiently strong predictors of a given exposure to be the sole constituent of a genetic IV, IVs in MR studies are usually comprised of several/many SNPs, reflecting polygenic contributions to complex exposures/phenotypes. As noted in **Chapter 1**, because genes are assigned at birth, are distributed more-or-less randomly in the population with respect to covariates, and can't be affected by reporting bias, their use in place of an observed exposure largely avoids the problems of reverse causality, confounding, and measurement error, mimicking an RCT (Burgess & Thompson, 2021; Ebrahim & Davey Smith, 2008). When estimating a given exposure–outcome relationship, instead of using data on observed levels of that exposure, MR uses data on *genetically-predicted* levels of exposure. If genes that increase levels of an exposure are also associated with increased levels of or risk for an outcome, that is evidence of a causal relationship.

Initially, MR analyses were performed with data from a single sample ('one-sample' MR), using individual-level data. This involves using each individual's data on genes, exposure, and outcome to estimate the IV–exposure and IV–outcome relationships. However, as large-scale genetic consortia have increasingly made their findings publicly available in the form of summary-level data, 'two-sample' MR – which involves combining the IV–exposure and IV–outcome estimates from different datasets – has become the dominant method. This summary-level data is generated from genome-wide association studies (GWASs), where associations are calculated between a phenotype of interest (e.g., alcohol consumption) and

every SNP in the genome (or as many SNPs as have been genotyped in a given cohort). Two-sample methods allow researchers without access to individual-level genetic data to perform MR analyses, are beneficial in cases where the exposure of interest and outcome of interest are not present in the same dataset, and greatly increase statistical power. However, for certain epidemiological questions – *including whether a relationship is non-linear* – two-sample, summary-level data is not sufficient, necessitating a reliance on individual-level data.

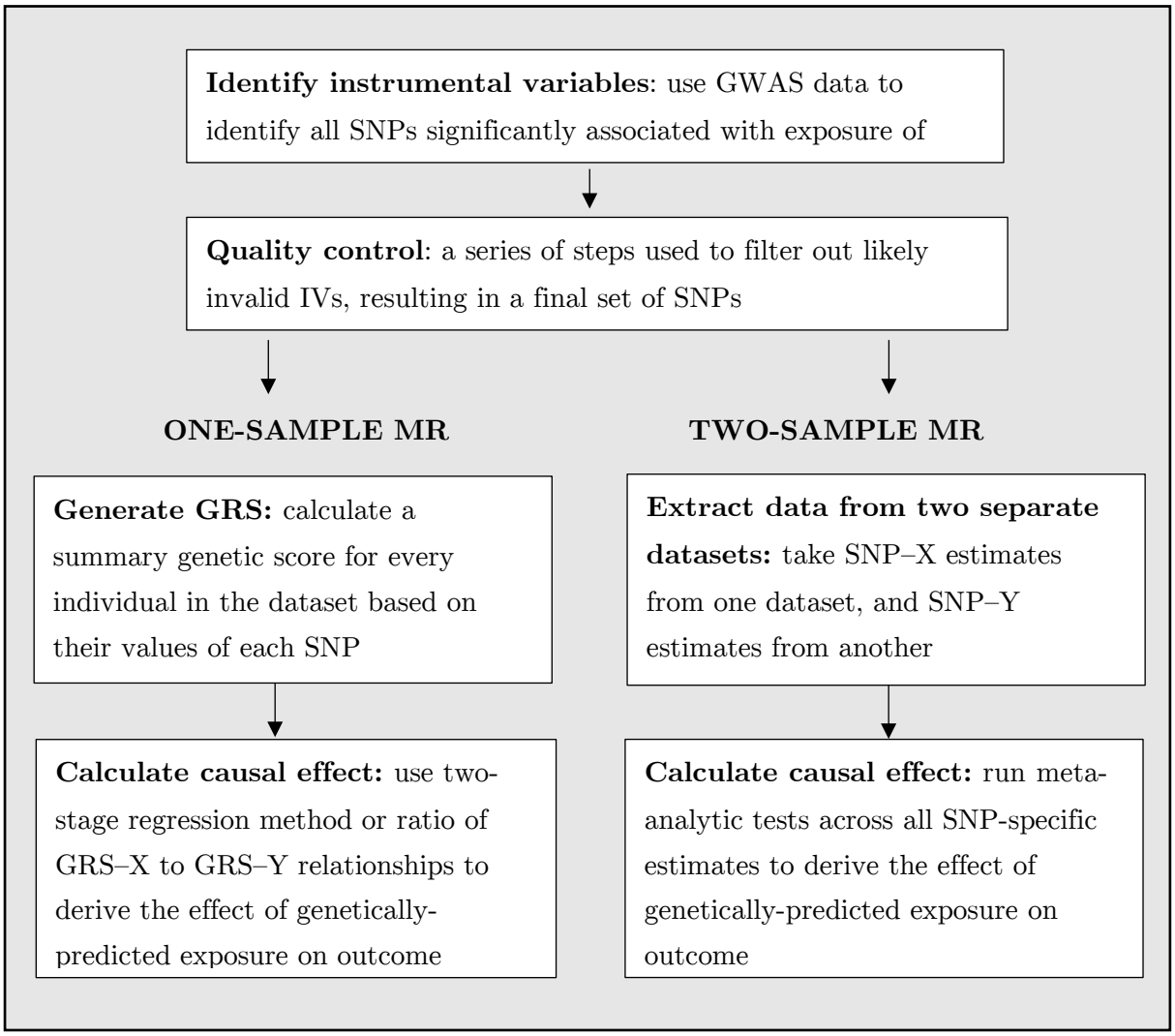
The basic steps in MR analyses are depicted in Figure 5.2. For both one- and two-sample MR, the first step is performing 'SNP discovery': identifying a set of SNPs that are associated with the exposure of interest at a genome-wide level (i.e., corrected for very large-scale multiple testing; generally considered to be $p < 5 \times 10^{-8}$). SNP discovery is usually performed using the publicly available GWAS data described above. Next, there is filtering and quality control of this set of SNPs, resulting in a final set of SNPs.

For two-sample MR, two separate GWASs are located: one with data on the exposure of interest, and the other with data on the outcome of interest. The two datasets are harmonised so that the associations between a given SNP and exposure and the same SNP and outcome are expressed per copy of the same allele. The betas and standard errors of the SNP–exposure and SNP–outcome associations are extracted for each SNP in the final set described above. Finally, various meta-analytic methods are applied across all SNPs to assess whether, on the whole, genes that increase levels of exposure are also associated with increased levels of or risk for an outcome, i.e., whether there is a causal effect.

For one-sample MR, a single cohort with genetic, exposure, and outcome data is sought, and the IV–exposure and IV–outcome relationships both calculated within that dataset. Because

genetic IVs typically predict just a small amount of variability in exposure, very large cohorts are required to achieve adequate statistical power. In one-sample MR, an IV is typically comprised of a genetic risk score (GRS; a weighted sum of SNP values) in place of many independent SNPs. This makes assumption checking simpler and more efficient, and mitigates bias associated with weak IVs (Burgess & Thompson, 2021). In this case, as there is a single IV, simpler (non-meta-analytic) methods are used to relate the IV–exposure relationship to the IV–outcome relationship in determining the presence of a causal effect (i.e., two-stage regression, or taking the ratio of the two relationships). For both one- and two-sample MR, results can then be interpreted in terms of the genetically-predicted exposure's effect on the outcome.

Figure 5.2. Steps in an MR study.



Abbreviations: GWAS, genome-wide association study; SNP, single-nucleotide polymorphism; IV, instrumental variable; GRS, genetic risk score

5.1.2 MR and alcohol–T2D research

Only recently have formal methods for non-linear MR – capable of identifying non-linear relationships between an exposure and outcome – been developed and refined (Burgess et al., 2014; Staley & Burgess, 2017). **Chapter 2** found no studies implementing these methods to examine the alcohol–T2D relationship (Visontay et al., 2022). A single non-linear MR analysis of alcohol and HbA1C found no evidence of either a linear or non-linear relationship

(Peng et al., 2019), although this was conducted in a very small, underpowered sample (n<5,000). Existing linear MR analyses of the alcohol–T2D relationship include two studies in East Asian populations finding increasing risk associated with alcohol (Cho et al., 2015; Peng et al., 2019), and two in European populations finding no robust results in either direction (Holmes et al., 2014; Lankester et al., 2021).

However, if the true relationship is non-linear, these results can only tell us about population-averaged causal effects resulting from shifts in exposure distribution (Burgess et al., 2014; Sun et al., 2019). These would average out over the various positive and negative effects for individuals sitting at different points along the exposure distribution (Burgess & Thompson, 2021). That is, they shed no light on whether an increase in alcohol consumption may reduce risk for non-drinkers but heighten risk for those already drinking substantially, as indicated by the observational literature. As such, there have been recent calls for non-linear MR in studies of risk factors for diabetes, and the alcohol–diabetes relationship specifically (Bryazka et al., 2022; Yuan et al., 2023).

Observational evidence also indicates that the alcohol–T2D relationship varies between men and women – specifically, that a protective effect is at least more pronounced for, if not exclusively present in, the latter (Knott et al., 2015). A linear MR analysis in UK Biobank found that women had greater alcohol-induced reductions in HbA1C, and lower alcohol-induced risk for T2D, but none of these differences were statistically significant (Lankester et al., 2021). Again however, these results are unable to tell us whether the *shape* of the relationship differs between sexes.

The present study performs the most robust causal analysis of the alcohol–T2D relationship to date. It aimed to apply formal non-linear MR analysis to determine the functional form of this relationship, contrast the relationship across the sexes, comprehensively assess sensitivity findings, and compare results to linear MR and observational evidence.

5.2 Methods

5.2.1 Data sources, participants, and inclusion criteria

The UK Biobank is a prospective cohort of over 500,000 participants aged 40-69 at baseline and recruited between 2006 and 2010 (Sudlow et al., 2015). Designed to identify risk factors for disease in middle and old age, it gathered comprehensive data from participants at baseline via questionnaires and nurse interviews (probing SES, early life exposures, psychosocial and environmental factors, lifestyle, health status, and cognitive function), anthropometrics/physical testing (e.g., body size and composition, grip strength, blood pressure), biochemical assays, and genotyping. Follow-up of the cohort via health record linkage is ongoing.

For inclusion in MR and observational analyses, individuals of European ancestry with complete genetic, alcohol consumption, covariate, and outcome data were eligible. Related individuals (kinship coefficient >0.0884) were randomly removed (Hanscombe et al., 2019). Individuals with either baseline prevalent or incident type 1 diabetes (as ascertained via UK Biobank derived 'first occurrence' variables, or self-reported insulin use at baseline), were removed for all analyses. This included those with a first diagnosis of T2D and subsequent

type 1 diabetes diagnosis, on the assumption that this reflects an initial misdiagnosis. All those with baseline prevalent diabetes of any kind (excluding gestational) for whom T2D was *not* the first recorded kind of diabetes were also excluded. Where survival analyses were used (i.e., for the primary non-linear MR and observational analyses), those with baseline prevalent diabetes of any kind (except gestational) were additionally removed. This was ascertained via UK Biobank derived 'first occurrence' variables, HbA1C ≥ 48 mmol/mol and/or glucose ≥11.1 mmol/l (Eastwood et al., 2016), and self-reported medication use at baseline (Mason, 2021).

SNP discovery was made using results in those of European descent from GSCAN, a meta-analysis of GWAS studies focused on drinking and smoking behaviours (N=941,280; M. Liu et al., 2019). SNP–exposure summary statistics (beta coefficients and standard errors) were also taken from GSCAN, but from the version without UK Biobank participants (N=226,223). This was done to avoid 'winner's curse' bias caused by overlapping samples in the SNP discovery and effect estimates datasets in the one-sample analysis (Jiang et al., 2022), and sample overlap in the two-sample analysis. Summary statistics for SNP–outcome relationships were taken from two large GWASs in European samples, based on 659,316 (T2D; 62,892 cases; Xue et al., 2018) and 146,806 (HbA1c; those with diabetes excluded; Chen et al., 2021) individuals respectively. Each of the three GWASs adjusted for age, sex, and ancestry.

5.2.2 Exposure variables

An overall usual drinks per week variable was derived from a range of questions answered by participants at baseline about drinking/non-drinking status, frequency of consumption, and

volume of beverage-specific consumption. Beverage-specific responses were converted to drinks per week using the following conversions: 1 UK standard drink=8 grams of alcohol; drinks reported per month were considered as per 4.3 weeks (Topiwala et al., 2022); 1 pint of beer/cider=16g (2 drinks); 1 wine=16.8g (2.1 drinks); 1 fortified wine=14.8g (1.76 drinks); 1 shot of spirits=8g (1 drink); 1 other alcoholic drink=12g (1.5 drinks; Biddinger et al., 2022). Current drinkers missing answers to any of the beverage-specific volume questions were excluded from analyses (except the questions pertaining to 'other' alcoholic drinks). Those drinking less than weekly were only asked about usual volume consumed if their baseline assessment was from 2008 onwards, so less-than-weekly drinkers enrolled before this time were excluded from analyses as a drinks per week variable could not be calculated. Non-drinkers were categorised as either lifetime abstainers or former drinkers.

5.2.3 Outcome variables

In 2022, the UK Biobank's derived 'first occurrence' variables were updated to thoroughly incorporate diabetic conditions. Utilising a combination of self-report, primary care, hospital inpatient, and death register data, they provide information for each participant on whether a particular condition was ever recorded, and if so, when first recorded and the source/s used to ascertain presence of that condition. These variables served as the basis for T2D outcome data in all analyses. Glycated haemoglobin (HbA1C), indicative of average blood glucose for the previous ~120 days (Burgess et al., 2021), was measured at baseline. It was used in all analyses as a continuous variable, with extreme values winsorised (i.e., any values greater than three standard deviations from the mean were replaced with the mean plus three

standard deviations; any values below three standard deviations from the mean were replaced with the mean minus three standard deviations).

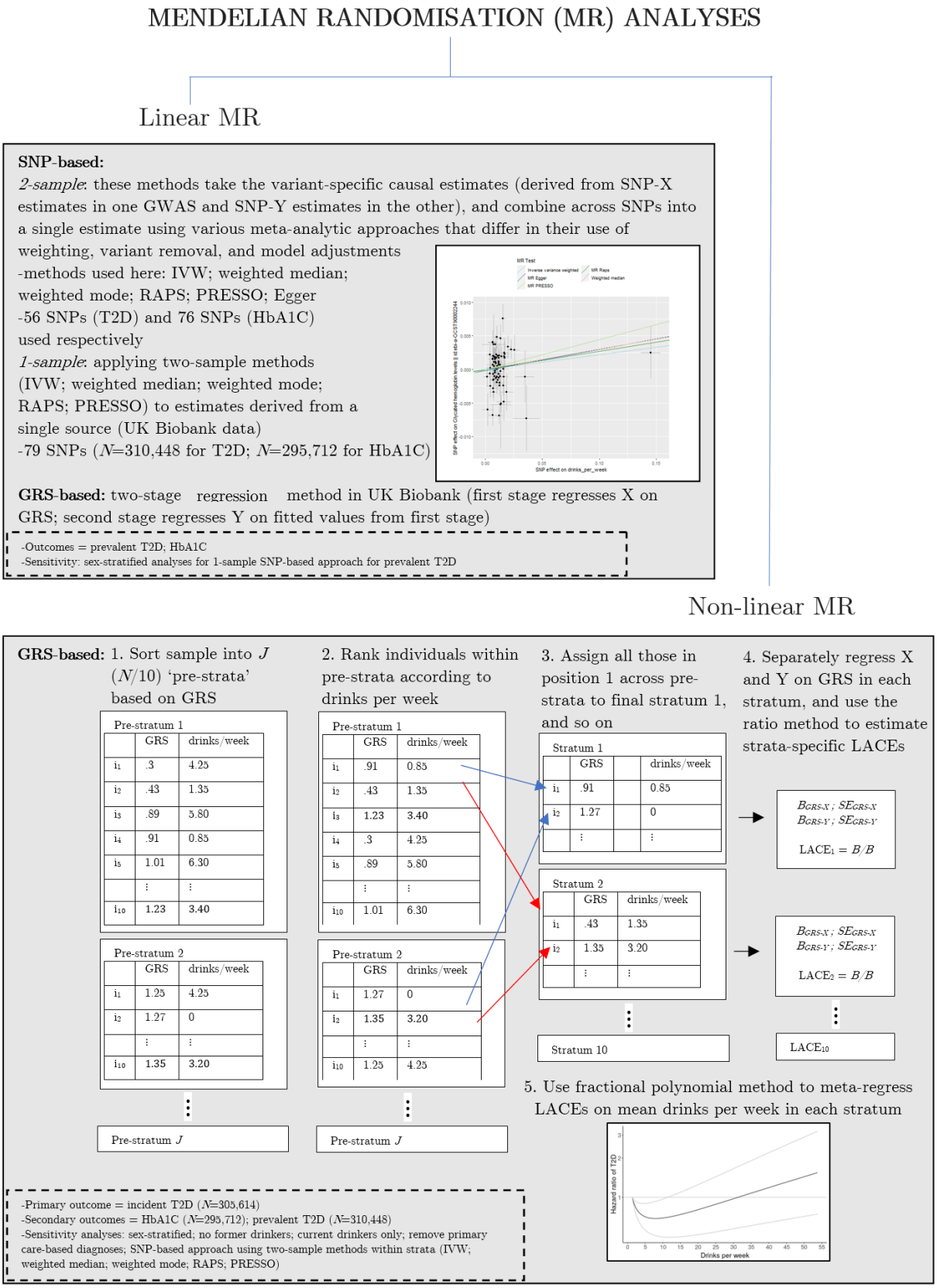
5.2.4 Genetic instruments

GSCAN identified 93 SNPs significantly associated with drinks per week at a genome-wide level in those of European descent. Several quality control measures were taken to increase confidence that the assumptions of an IV design, described above, would hold, excluding SNPs where necessary (see Appendix 4.1 for further detail). This resulted in 79 SNPs for the individual-level MR and 77 SNPs for the two-sample MR (see Appendix 4.2). To produce GRSs for the one-sample analyses, dosages (number of effect alleles) for each of the 79 SNPs were weighted by the relevant beta coefficients identified in GSCAN (version without UK Biobank participants) and summed.

5.2.5 Statistical analysis

Figure 5.3 illustrates in detail the steps involved in both the linear and non-linear MR analyses conducted. While the non-linear analyses were those of primary interest, linear MR was also conducted for comparison – particularly as it could also be performed with two-sample, summary-level data.

Figure 5.3. Overview of MR analyses employed; informed by Tian et al., (2022) and A. Zhou & Hyppönen (2022)



5.2.5.1 Linear MR (two-sample and one-sample)

Linear MR was performed for T2D (prevalent baseline cases only, given the absence of time-to-event information in summary-level GWAS data) and HbA1c. Two-sample linear MR was performed using the following meta-analytic tests: 1) Inverse-variance weighted (IVW) meta-analyses (with random effects to allow for separate slopes of each SNP's association with exposure and outcome); 2) Weighted median MR (takes the median as the causal effect, allowing up to 50% of IVs to be invalid); 3) Weighted mode MR (takes the mode as the causal effect, allowing IVs from all other effects to be invalid); 4) MR PRESSO (estimates effects after removing outliers); 5) MR RAPS (allows all IVs to be invalid due to pleiotropy – when a SNP affects more than one phenotype – as long as the pleiotropy is balanced); 6) MR Egger (tolerant to invalid IVs by allowing for non-zero intercepts). Each test offers a degree of resistance to invalid IVs, i.e., SNPs that violate one or more of the IV design assumptions, in different ways. Consistent findings across these complementary methods increase confidence in results.

Variants were harmonised across the three sets of summary statistics, and for any of the 77 SNPs associated with exposure missing in either of the outcome GWASs, proxy SNPs were sought, resulting in 56 SNPs for the T2D analysis and 76 SNPs for the HbA1C analysis. The *TwoSampleMR* (Hemani et al., 2017, 2018) and *mr.raps* (Q. Zhao et al., 2020) R packages were used for these analyses.

One-sample linear MR was also performed in UK Biobank for comparison, using the 79-SNP GRS in combination with the two-stage least-squares method for calculating a causal effect. The relationship between the GRS and drinks per week was calculated in controls only for

the T2D analysis, to mitigate the influence of reverse causation. Additionally, a SNP-based approach – borrowed from its traditional use in two-sample methods – was performed. That is, instead of aggregating all SNPs into a GRS, the individual SNP– X and SNP–Y relationships were calculated. This enabled performing the meta-analytic tests described earlier that allow for invalid IVs: IVW, weighted median, weighted mode, MR PRESSO, and MR RAPS (MR Egger was not used as it is not recommended in one-sample MR; Minelli et al., 2021). The two-stage least-squares and SNP-based methods were performed using the *ivtools* (Sjolander & Martinussen, 2019) and *MendelianRandomization* (Yavorska & Burgess, 2017)/*TwoSampleMR* R packages, respectively. These analyses also controlled for age, sex, genotyping 'array' (each individual's blood samples were subjected to one of two specific genotyping protocols), and ancestry. As the one-sample analyses used individual-level data, analyses were also repeated in men and women separately.

5.2.5.2 Non-linear MR

Non-linear MR involves generating localised average causal effects (LACEs) within strata of observed exposure – essentially, performing several separate MR analyses across the spectrum of alcohol consumption. If LACEs are consistent across strata, this indicates a linear exposure–outcome relationship (i.e., the effect of increasing genetically-predicted alcohol consumption on the outcome does *not* change at different points along the exposure distribution). Conversely, if LACEs differ between strata, this indicates non-linearity.

Non-linear MR was performed for both T2D and HbA1C levels in UK Biobank. The primary MR analysis for T2D was survival analysis with a Cox model using incident cases only, where T2D was a time-to-event outcome. A secondary logistic regression analysis was

performed with baseline prevalent T2D cases as the outcome. Non-linear MR analyses controlled for age, sex, genotyping array, and ancestry, and were performed in the overall sample, as well as males and females separately.

Several steps were involved in this analysis: sorting the sample into strata, calculating LACEs within strata, and then applying plotting methods and statistical tests *across* the strata-specific LACEs to ascertain the shape of the exposure–outcome relationship.

5.2.5.2.1 Stratification

Stratification cannot be based directly on observed drinking level because exposure is a common effect of both the IV and confounders of the exposure–outcome relationship, and would thereby introduce collider bias (refer to Figure 1.5 in **Chapter 1**). Until recently, the 'residual method' was the dominant stratification method that avoided collider bias, involving stratification on the residuals from a regression of exposure on IV. However, this method assumes heterogeneity in the strength of the gene–exposure relationship across strata, which, when violated, leads to biased MR estimates (Tian et al., 2022). Tian et al. (2022) have demonstrated that this assumption is indeed *not* met for genetic IVs for alcohol consumption in the UK Biobank, and as an alternative, propose the 'doubly-ranked' method for stratification.

Specifically, the method first divides individuals into ten-person strata based on their GRS (genetically-predicted drinking), then ranks each person within these strata based on their observed drinking, and lastly, groups individuals based on these ranks into ten final strata (such that all those ranked lowest within each GRS-based stratum – i.e., all those with the lowest amount of actual drinking given how much they are genetically-predicted to drink –

end up in stratum number one, and so on). The doubly-ranked method is illustrated in Figure 5.3.

5.2.5.2.2 Calculating LACEs

The ratio method was used to calculate LACEs within strata. First, within each stratum, the outcome was regressed on the GRS, and then observed drinks-per-week was regressed on the GRS. Again, for non-linear MR examining baseline prevalent T2D, the relationship between the GRS and drinks per week in each stratum was calculated in controls only to mitigate the influence of reverse causation. The LACE was then calculated as the ratio of the two beta estimates from these regressions, with the standard errors of the LACEs calculated as the ratio of the two standard errors from these regressions.

5.2.5.2.3 Ascertaining non-linearity and functional form

These strata-specific LACEs and their standard errors were then used to plot genetically-predicted alcohol consumption against hazards (time-to-event T2D)/odds (baseline prevalent T2D) and betas (HbA1c), allowing for visualisation of relationship shape.

Plots were constructed using both fractional polynomials (involving meta-regression of the LACEs on mean exposure value in each stratum, followed by plotting of the best fitting polynomial of degree one or two) and piecewise linear methods (a continuous function comprised of the individual LACE slopes from each stratum; Mason & Burgess, 2022). The reference was set at the minimum value of drinks per week (for fractional polynomials that was mean drinks per week for the lowest stratum; for piecewise plots that was zero drinks per week).

Formal tests of non-linearity were also performed: 1) fractional polynomial tests to determine the best-fitting model; 2) quadratic tests (the equivalent of trend tests between exposure and LACE values); 3) Cochran Q tests of heterogeneity between the LACE values.

Non-linear MR was performed with the *SUMnlmr* and *NLMR* R packages (Mason & Burgess, 2022; Staley & Burgess, 2017).

5.2.5.2.4 Sensitivity analyses

Sensitivity analyses included removing former drinkers, removing former and never drinkers, and removing those with a T2D diagnosis ascertained via primary care records. Additionally, for incident T2D and HbA1c outcomes, a SNP-based approach to LACE calculation was performed using IVW, weighted median, weighted mode, MR PRESSO, and MR RAPS tests, to again account for the possibility of invalid IVs. As there are no existing R packages for implementing SNP-based methods in a non-linear MR framework, performing these analyses required that the doubly-ranked strata be manually created (adapting code from Tian et al., 2022), the SNP-based tests run separately within each stratum, and then code from *SUMnlmr* and *NLMR* packages adapted to meta-analyse results across strata.

5.2.5.3 Observational analyses

Finally, traditional observational analyses were conducted for comparison with MR results. These modelled non-linear relationships between alcohol consumption and incident T2D risk using a generalised additive survival model, with alcohol entered as a spline term, and controlling for covariates ascertained at baseline via self-report/interview: age, sex, smoking history (never/once or twice/occasionally/most or all days), education (obtainment/not of tertiary degree), SES (Townsend deprivation index), and physical activity (total weekly

metabolic equivalent (MET) minutes). Cox models were constructed with the *rms* R package (Harrell Jr et al., 2017) using restricted cubic splines (three knots). Analyses were repeated in men and women separately.

All analyses were performed with R version 4.1.1. This work was conducted under UK Biobank application project number 86004. The study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) reporting guidelines (see Appendix 4.11 for checklist; Skrivankova et al., 2021).

5.3 Results

5.3.1 Descriptive statistics

305,614 individuals were included in Cox non-linear MR, 310,448 in baseline prevalent T2D non-linear MR, 295,712 individuals in baseline HbA1C non-linear MR, and 233,568 individuals in observational Cox analyses.

There were 2,923 cases of T2D at baseline. For the sample used in the Cox analyses, there were 10,211 incident cases up to 31 May 2016 (the end date of primary care data coverage for most participants with such data). Descriptive statistics for this sample are presented in Table 5.1.

Table 5.1. Characteristics of sample used in survival-based non-linear MR analyses (prevalent T2D cases excluded)

	Current Drinker		**Former Drinker**		**Never Drinker**		**Overall**	
	Male (N=137981)	**Female (N=145115)**	**Male (N=4835)**	**Female (N=6808)**	**Male (N=2644)**	**Female (N=8231)**	**Male (N=145460)**	**Female (N=160154)**
Drinks per week								
Mean (SD)	25.4 (20.9)	15.6 (14.1)	0 (0)	0 (0)	0 (0)	0 (0)	24.1 (21.1)	14.1 (14.1)
Median [Min, Max]	20.2 [0, 291]	12.6 [0, 231]	0 [0, 0]	0 [0, 0]	0 [0, 0]	0 [0, 0]	19.0 [0, 291]	10.5 [0, 231]
Incident diabetes status								
Did not develop incident T2D	132268 (95.9%)	141982 (97.8%)	4485 (92.8%)	6394 (93.9%)	2454 (92.8%)	7820 (95.0%)	139207 (95.7%)	156196 (97.5%)
Developed incident T2D	5713 (4.1%)	3133 (2.2%)	350 (7.2%)	414 (6.1%)	190 (7.2%)	411 (5.0%)	6253 (4.3%)	3958 (2.5%)
HbA1C (winsorised)								
Mean (SD)	34.9 (3.70)	34.8 (3.60)	35.5 (3.98)	35.7 (3.79)	35.4 (4.02)	35.9 (3.70)	34.9 (3.72)	34.9 (3.62)
Median [Min, Max]	34.8 [20.3, 47.9]	34.8 [20.3, 47.9]	35.4 [20.3, 47.9]	35.7 [20.3, 47.9]	35.3 [20.3, 47.9]	35.9 [20.3, 47.9]	34.8 [20.3, 47.9]	34.9 [20.3, 47.9]
Missing	6526 (4.7%)	7020 (4.8%)	208 (4.3%)	326 (4.8%)	126 (4.8%)	386 (4.7%)	6860 (4.7%)	7732 (4.8%)
Age								
Mean (SD)	57.0 (8.03)	56.4 (7.82)	57.2 (7.97)	57.5 (7.66)	58.5 (8.25)	59.4 (7.57)	57.0 (8.04)	56.6 (7.83)

Table 5.1. Characteristics of sample used in survival-based non-linear MR analyses (prevalent T2D cases excluded)

	Current Drinker		**Former Drinker**		**Never Drinker**		**Overall**	
	Male (N=137981)	**Female (N=145115)**	**Male (N=4835)**	**Female (N=6808)**	**Male (N=2644)**	**Female (N=8231)**	**Male (N=145460)**	**Female (N=160154)**
Median [Min, Max]	58.0 [39.0, 73.0]	58.0 [40.0, 70.0]	59.0 [40.0, 70.0]	59.0 [40.0, 70.0]	61.0 [40.0, 70.0]	61.0 [40.0, 71.0]	59.0 [39.0, 73.0]	58.0 [40.0, 71.0]
Smoking history								
Never	46830 (33.9%)	60481 (41.7%)	1429 (29.6%)	2559 (37.6%)	1899 (71.8%)	5929 (72.0%)	50158 (34.5%)	68969 (43.1%)
Once or twice	21000 (15.2%)	23506 (16.2%)	460 (9.5%)	789 (11.6%)	245 (9.3%)	658 (8.0%)	21705 (14.9%)	24953 (15.6%)
Occasionally	18670 (13.5%)	20930 (14.4%)	505 (10.4%)	743 (10.9%)	104 (3.9%)	409 (5.0%)	19279 (13.3%)	22082 (13.8%)
Most or all days	40246 (29.2%)	31059 (21.4%)	1681 (34.8%)	1835 (27.0%)	251 (9.5%)	745 (9.1%)	42178 (29.0%)	33639 (21.0%)
Missing	11235 (8.1%)	9139 (6.3%)	760 (15.7%)	882 (13.0%)	145 (5.5%)	490 (6.0%)	12140 (8.3%)	10511 (6.6%)
Tertiary degree								
No	89913 (65.2%)	97144 (66.9%)	3547 (73.4%)	5217 (76.6%)	1894 (71.6%)	6677 (81.1%)	95354 (65.6%)	109038 (68.1%)
Yes	48068 (34.8%)	47971 (33.1%)	1288 (26.6%)	1591 (23.4%)	750 (28.4%)	1554 (18.9%)	50106 (34.4%)	51116 (31.9%)

Table 5.1. Characteristics of sample used in survival-based non-linear MR analyses (prevalent T2D cases excluded)

	Current Drinker		**Former Drinker**		**Never Drinker**		**Overall**	
	Male (N=137981)	**Female (N=145115)**	**Male (N=4835)**	**Female (N=6808)**	**Male (N=2644)**	**Female (N=8231)**	**Male (N=145460)**	**Female (N=160154)**
Townsend deprivation index								
Mean (SD)	-1.70 (2.87)	-1.78 (2.75)	-0.181 (3.52)	-0.636 (3.30)	-1.15 (3.08)	-1.24 (3.06)	-1.64 (2.91)	-1.71 (2.80)
Median [Min, Max]	-2.46 [-6.26, 10.9]	-2.48 [-6.26, 10.8]	-0.924 [-6.26, 10.5]	-1.44 [-6.26, 9.79]	-1.94 [-6.26, 9.90]	-2.06 [-6.26, 9.79]	-2.42 [-6.26, 10.9]	-2.43 [-6.26, 10.8]
Missing	159 (0.1%)	174 (0.1%)	11 (0.2%)	8 (0.1%)	5 (0.2%)	9 (0.1%)	175 (0.1%)	191 (0.1%)
MET minutes per week								
Mean (SD)	2850 (2910)	2530 (2420)	3110 (3470)	2640 (2770)	2960 (3300)	2770 (2870)	2860 (2940)	2540 (2450)
Median [Min, Max]	1890 [0, 19300]	1770 [0, 19300]	1870 [0, 19300]	1710 [0, 19300]	1820 [0, 19300]	1810 [0, 19300]	1890 [0, 19300]	1770 [0, 19300]
Missing	19294 (14.0%)	29423 (20.3%)	861 (17.8%)	1739 (25.5%)	504 (19.1%)	2252 (27.4%)	20659 (14.2%)	33414 (20.9%)
Genotyping array								
UK Biobank Axiom array	121548 (88.1%)	130641 (90.0%)	4137 (85.6%)	5883 (86.4%)	2225 (84.2%)	7221 (87.7%)	127910 (87.9%)	143745 (89.8%)

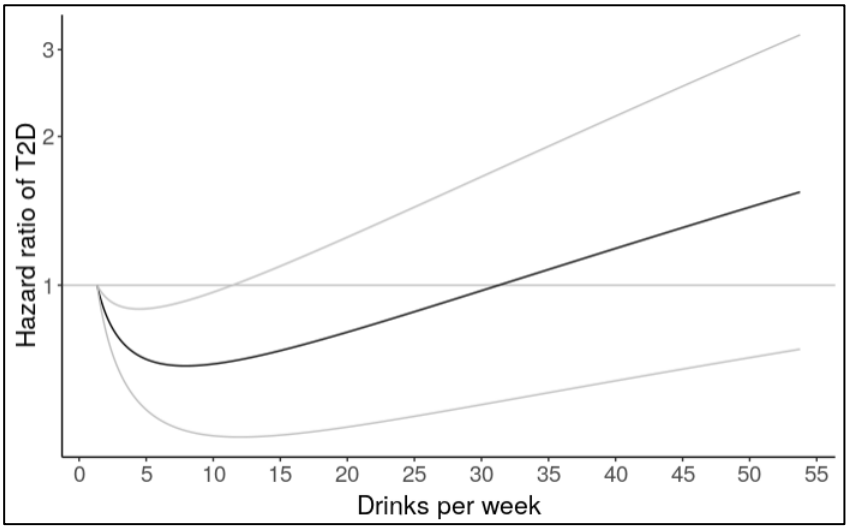
Table 5.1. Characteristics of sample used in survival-based non-linear MR analyses (prevalent T2D cases excluded)

	Current Drinker		**Former Drinker**		**Never Drinker**		**Overall**	
	Male (N=137981)	**Female (N=145115)**	**Male (N=4835)**	**Female (N=6808)**	**Male (N=2644)**	**Female (N=8231)**	**Male (N=145460)**	**Female (N=160154)**
UKBiLEVE Axiom array	16433 (11.9%)	14474 (10.0%)	698 (14.4%)	925 (13.6%)	419 (15.8%)	1010 (12.3%)	17550 (12.1%)	16409 (10.2%)

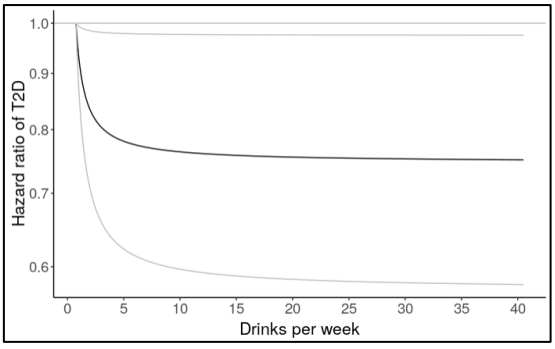
5.3.2 Non-linear MR

Figure 5.4. Best-fitting fractional polynomials depicting genetic associations between alcohol consumption and risk for incident T2D in A) the overall sample; B) females; C) males.

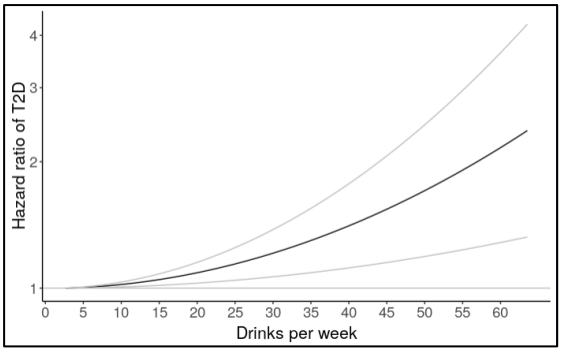
A) Overall



B) Females



C) Males



In the survival model, the association between genetically-predicted alcohol consumption and incident T2D risk was consistent with a J-shape (Figure 5.4A; the second-degree polynomial was a better fit than the first-degree polynomial, p=.01; quadratic p=.40; Cochran Q p=.06). The theoretical minimum risk exposure level (TMREL) for the fractional polynomial lay at 6 drinks per week (hazard ratio=.76), and the non-drinker equivalence (NDE) point occurred

at 21 drinks per week, with drinkers significantly different from non-drinkers to 8 drinks per week. For the secondary analysis using baseline prevalent T2D as the outcome (Appendix 4.3), results were instead consistent with an exponentially increasing function (the first-degree polynomial was a better fit than the linear model, p=.047; quadratic p=.02, Cochran Q p=.29). Piecewise linear functions (Appendix 4.4) were largely concordant with these results (although do indicate a slight protective effect at low-level drinking for the baseline prevalent analysis; Appendix 4.4D), but are not readily interpretable given very large confidence intervals.

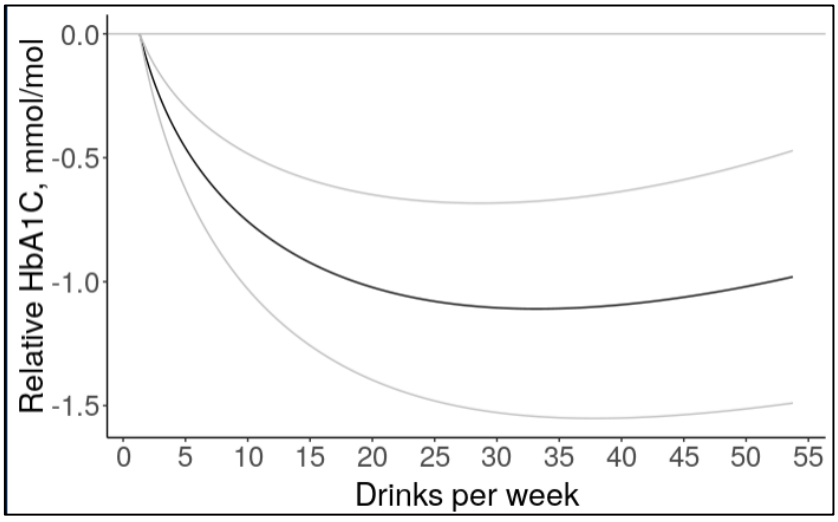
Analyses stratified by sex revealed functional forms that differed greatly between women and men (Figures 5.4B and C). All levels of genetically-predicted alcohol consumption (up to 40 drinks per week) were predicted to decrease T2D risk in women (the first-degree polynomial was a better fit than the linear model, p=.03; quadratic p=.49, Cochran Q p=.59), while for men, risk appeared to increase exponentially (although the first-degree polynomial was not a better fit than the linear model, p=.39; quadratic p=.36, Cochran Q p=.60). Piecewise models indicate a slight protective effect for men at very low levels of drinking that are not apparent in the polynomial (Appendix 4.4C).

Sensitivity analyses excluding former drinkers did not change the functional form for MR results, although it lowered the number of drinks per week associated with reduced risk of incident T2D and magnified the increased risk for heavy drinking (Appendix 4.5A). When excluding both former and never drinkers – i.e., limiting the sample to current drinkers – protective effects were no longer observed for the overall sample or for women, producing a positive linear relationship and no relationship respectively (Appendices 4.6A and B). Most *but not all* of the SNP-based approaches to non-linear MR produced results concordant with

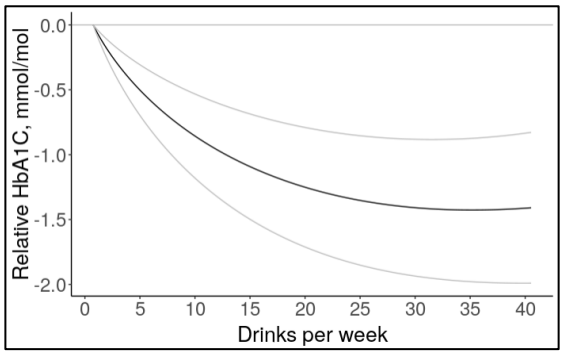
the primary analyses, with the weighted median and weighted mode methods resulting in positive linear relationships in the overall sample and the protective effect in women disappearing/becoming non-significant (Appendices 4.8A and B).

Figure 5.5. Best-fitting fractional polynomial depicting genetic associations between alcohol consumption and levels of HbA1C in A) the overall sample; B) females; C) males.

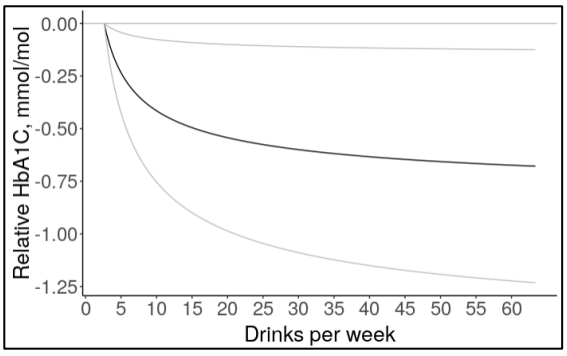
A) Overall



B) Females



C) Males



Non-linear MR analyses indicated HbA1C levels at baseline decreased with increasing genetically-predicted alcohol consumption (Figure 5.5A; the second-degree polynomial was a better fit than the first-degree polynomial, p=.02; quadratic p<.001; Cochran Q p<.001). The functional form was similar in the piecewise model (Appendix 4.4E). The reduction in HbA1C levels associated with moderate drinking was greater for women than men (Figures 5.5B and C). All sensitivity analyses resulted in consistent findings for the overall sample (Appendices 4.5C, 4.6D, and 4.9; although the weighted mode SNP-based approach had wide confidence intervals).

5.3.3 Linear MR

Neither two- nor one-sample linear MR analyses suggested an effect of alcohol on T2D in either direction, including when the latter were stratified by sex (Appendix 4.10). Two-sample methods for HbA1C tended towards a harmful effect of increasing alcohol use, while the reverse was true of one-sample analyses with UK Biobank, but almost none of these were statistically significant. The MR Egger intercept from two-sample analyses did not indicate statistically significant directional pleiotropy in either case (for T2D: intercept=.003; p=.64; for HbA1c: intercept<-.001, p=.765).

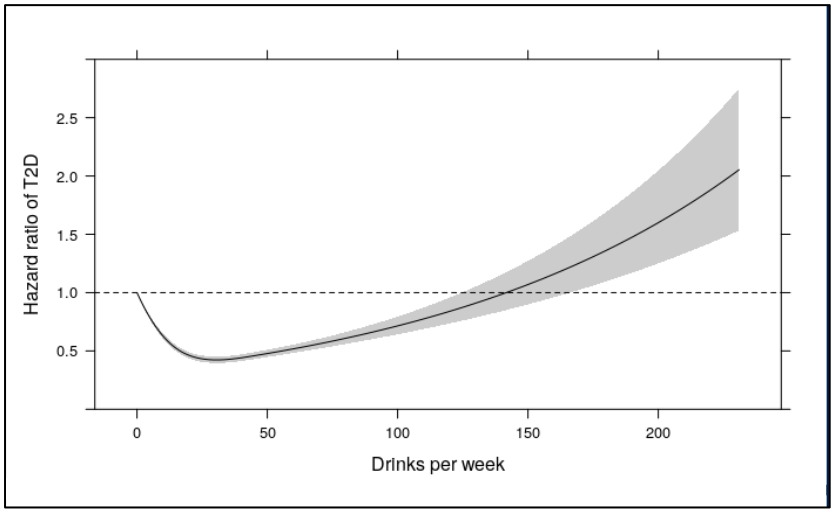
5.3.4 Observational analyses

Survival analyses using observed alcohol consumption and incident T2D cases were consistent with a J-shaped relationship (Figure 5.6A), with the NDE and TMREL corresponding to a very large number of drinks per week. Results were similar in the samples excluding former drinkers, and when excluding both former drinkers and never drinkers (Appendices 4.5D and 4.6E). Analyses stratified by sex revealed a larger protective effect

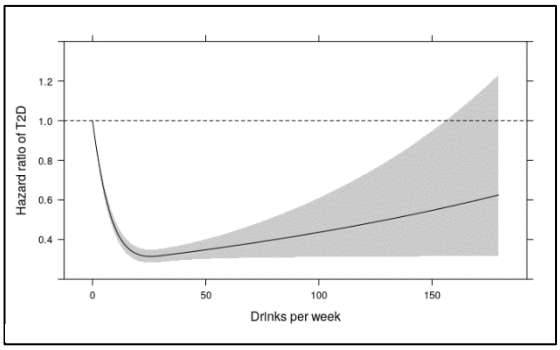
and absence of increased risk for heavy drinking in women, compared with men (Figures 5.6B and C).

Figure 5.6. The observational association between drinks per week and risk for incident T2D over follow-up in A) the overall sample; B) females; C) males.

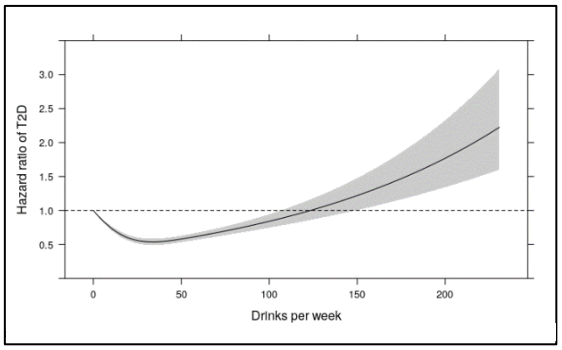
A) Overall



B) Females



C) Males



5.4 Discussion

Initial findings from non-linear MR indicated a J-shaped relationship between alcohol and T2D risk, with a small but significant protective effect up to roughly 8 drinks per week. However, sex-stratified analyses suggested this protective effect may only be present for women. SNP-based sensitivity analyses were not all consistent with the main analyses – in particular, weighted median and mode MR failed to produce significant protective effects in

the main sample and for women specifically. Results for HbA1C, an index of average blood glucose levels, were robust across sensitivity tests, revealing a protective effect of alcohol at all levels of genetically-predicted consumption that was larger for women than for men. Linear MR for both outcomes produced no clear results, but this is expected if the true relationship is non-linear given the averaging-out phenomenon discussed above. This finding is also consistent with other MR studies that have found clear evidence of non-linear epidemiological relationships while linear methods produced null results (Hernáez et al., 2021; A. Zhou et al., 2022).

The presence of a protective effect afforded by low-to-moderate drinking against incident T2D when using genetic-based methods indicates there may be a genuine causal relationship in women, although discordant results from some sensitivity analyses reduce confidence in this finding, suggesting that invalid IVs may underpin the primary results. The positive linear results found when limiting the sample to current drinkers places further bounds on any potential benefit in women (i.e., increasing volume of consumption among those *already* drinking is unlikely to provide added benefit).

Recent studies using causal methods have clarified that, for other health outcomes, there are clear harmful effects of alcohol consumption at all levels. For example, recent non-linear MR in UK Biobank found alcohol exponentially increased cardiovascular risk (Biddinger et al., 2022). That was not the case here. The present findings are instead consistent with the observational literature on T2D, although suggest a much more modest volume limit associated with protection. Indeed, evidence reviews and modelling exercises conducted for recent updates to both Australia and Canada's low-risk drinking guidelines showed T2D to

be one of the few outcomes where the evidence base continues to indicate that some level of drinking confers reduced risk (Angus et al., 2019; Paradis et al., 2023).

As in the present analyses, previous observational studies also determined that protective effects were largely limited to women. It is possible that differential effects of ethanol on insulin sensitivity between the sexes may be responsible for this disparity in both protective effects at lower levels and harm at higher levels of consumption (Schrieks et al., 2015), but more research is required to establish what biological mechanisms mediate this divergence. A competing explanation may be the sex-discrepant beverage preferences in UK Biobank, with women favouring wine, and men favouring beer/cider (T. Zhou et al., 2021). While ethanol – common to all beverages – is generally considered the key health-relevant component of alcoholic drinks (Krenz & Korthuis, 2012), when it comes to metabolic conditions specifically, factors like carbohydrate content (that do differ between beverages) may have additional impacts (Brand-Miller et al., 2007).

The HbA1C analyses in the present study also produced a protective effect more pronounced for women than men. Interestingly, the only previous non-linear MR analysis of alcohol and HbA1C found no evidence of either a linear or non-linear relationship (Peng et al., 2019) – this may be explained by either a genuinely different effect in Asian populations, or a lack of power (n~5,000). As larger, more refined GWASs and additional large-scale, individual-level datasets become available, future non-linear MR could elucidate possible ethnic differences, as well as consolidating our understanding of alcohol–metabolic relationships more generally.

5.4.1 Strengths

This study represents the first formal non-linear MR analysis of alcohol consumption and risk for T2D and has several strengths. It is one of the first non-linear MR studies to implement the doubly-ranked stratification method, that unlike the dominant residual method, does not assume heterogeneity in the strength of the GRS–exposure relationship across strata – an assumption that was indeed violated in the present data. Another strength was the use of SNP-based sensitivity methods to complement the GRS-based ratio method. Additionally, the genetic instrument was carefully selected (chosen as the better-performing of two candidates; see Appendix 4.1) and indicated no evidence of directional pleiotropy. Unlike other recent MR studies that have relied on instruments associated with alcohol use *disorder* (Biddinger et al., 2022; Lankester et al., 2021), the instrument used in the present study is a true proxy for the exposure of interest – volume consumed. Finally, the present study's focus on sex-specific results is a critical addition to the literature.

5.4.2 Limitations

Less-than-weekly drinkers in UK Biobank were only asked about usual volume consumed from 2008 onwards, so those enrolled prior to this date were excluded from analyses. However, given that time of enrolment was directly linked to assessment centre (centres opened in a staggered fashion around the country), it is reasonable to assume equivalence between earlier and later recruits. The other data limitation was that primary care linkage has only been achieved for just under 50% of the cohort, meaning many incident T2D cases are probably unrecorded. Again, it is unlikely that linkage, or lack thereof, was differentially

associated with T2D risk, and a sensitivity analysis excluding cases diagnosed via primary care records (<10% of cases) produced consistent results (Appendix 4.7).

There are also several limitations stemming from the nature of MR itself. Firstly, the instruments used are associated with lifetime average drinks per week, so the findings offer no insight into the differential impacts of exposure at various ages, nor on the impact of drinking *patterns*. Xue et al. (2021) have also demonstrated that misreport of and longitudinal change in alcohol consumption may bias GWAS and subsequent MR estimates. Additionally, the HbA1C analyses should be interpreted with some care given the potential for reverse causation introduced by stratification when performing non-linear MR with a baseline-measured outcome. This may also explain why non-linear MR analyses using baseline prevalent cases produced a different result to those using incident cases, although a competing explanation is healthy survivor bias (i.e., that drinkers who lived to baseline without developing T2D are healthier than drinkers more generally).

Finally, despite the precautions taken to remove SNPs with evidence of association with confounders of the alcohol–T2D relationship (see Appendix 4.1), it is not possible to guarantee the assumptions of exchangeability and exclusion restriction. That is, potential for confounding and/or horizontal pleiotropy remain. Of particular concern are SES factors – strongly associated with both drinking and health outcomes and thus responsible for many spurious findings in the observational alcohol–long-term health literature. It is possible that additional SES-associated SNPs were missed by the filtering process, meaning the present findings could be partially confounded or otherwise explained by the impact of SES on T2D independent of alcohol.

5.4.3 Conclusions

Genetic-based methods for enhancing causal inference do not rule out a possible protective effect of low-level alcohol consumption against T2D risk in women and support a benefit of moderate drinking for average glucose levels in both sexes. Further MR studies in diverse cohorts, and triangulation amongst their findings, are needed for elucidation of these relationships. A small beneficial effect of alcohol for T2D in women (if substantiated), as well as the demonstrated reductions in blood glucose levels for both sexes, is important information for policymakers and clinicians, but must be balanced against the known risks for a host of other conditions that accrue with even low-level drinking, including several cancers (Rehm et al., 2019). Observational evidence substantially overestimates the benefits, and underestimates the harms, of alcohol for diabetes risk.

Chapter 6

General Discussion

Preface

Despite the substantial harms caused by heavy drinking, a considerable body of observational research has long suggested that low-level alcohol consumption may have health benefits for select conditions. Scepticism surrounds these findings and whether they represent true causal effects – a question on which considerable public health and policy outcomes are contingent.

The overarching aim of this thesis was to investigate whether these alcohol–long-term health relationships persist when statistical and design-based methods to improve causal inference are applied. This thesis represents a significant and multifaceted contribution to the literature via its novel application of state-of-the-science methodology to three epidemiological relationships requiring urgent clarification.

In the remaining sections of this final chapter, an overview of this thesis' findings is presented (6.1), followed by an exploration of the implications of these findings for policy, practice, and research (6.2). The strengths and limitations of the thesis are then discussed (6.3), as are future research directions building on the present work (6.4). The chapter ends with some final conclusions (6.5).

6.1 Overview of findings

Chapter 2 presented a detailed overview of enhanced causal inference approaches, and systematically reviewed alcohol–long-term health studies employing these methods (Visontay et al., 2022). Sixteen longitudinal studies, spanning cancer, diabetes, dementia, mental health, cardiovascular health, mortality, HIV seroconversion, and musculoskeletal health outcomes, were critically examined and synthesised. The key finding was the dearth of extant research: most outcomes had been assessed by only one study and one method (with many studies excluded for assuming linearity), prohibiting strong conclusions. Where an outcome was assessed in more than one study, findings tended to be discrepant. Limited evidence of protective effects of low-level drinking were found for a handful of outcomes, including inflammatory marker C-reactive protein and mental health conditions. Triangulation across the causal and non-causal literature produced concordance for some outcomes (e.g., depression) but not others. Overall, this review provides a critical overview of the alcohol–long-term health causal inference landscape, serving as a much-needed foundation for future epidemiological studies and triangulation work.

Chapter 2 identified **inflammation** as one outcome with which alcohol may have a non-linear dose-response relationship, but found only one causally-focussed study. When it comes to the conventional observational evidence base, there is considerable inconsistency in findings. The possibility that this was due to variability in data processing and analysis choices between studies had not been explored. **Chapter 3** tested whether the association between alcohol and inflammation met a preliminary condition for causation – stability in the face of research parameter variation (Visontay, Mewton, Sunderland, et al., 2023). It is

the first study to have employed enhanced single-study sensitivity frameworks (i.e., multiverse analysis) in assessing the robustness of the alcohol–inflammation relationship, and any alcohol–long-term health relationship more generally. The first chief finding was that the association between low-to-moderate drinking and lower inflammation was largely robust to changes in definitions of drinking groups, measurement year, outcome transformation, and breadth of covariate adjustment. However, the association between heavier drinking and inflammation was not as clear, with effect direction and strength being particularly dependent on definitions of drinking groups and interval between exposure and outcome measurement. As inflammation is increasingly linked to a broad range of illnesses, this study has considerable implications for uncovering pathways between drinking and downstream health benefits/harms.

Chapter 2 also identified limited evidence for a protective effect of moderate drinking against **depression**. Again though, existing causal studies were sparse. **Chapter 4** scrutinised the relationship between drinking and depression by applying a marginal structural model – a statistically-based method to improve causal inference – to rich, longitudinal data (Visontay, Mewton, Slade, et al., 2023). The most striking finding was that, contrary to expectations, results resembled a J-shaped relationship despite the use of statistical methods to mitigate complex longitudinal confounding and promote causal inference. The strongest benefit of low-level drinking aligned with the largest interval between exposure and outcome, suggesting these benefits accrue over time, while the greatest harm of above-guidelines consumption occurred for drinking in the few years before depression measurement. This work represented several improvements over the only previous marginal structural model study, in that it incorporated multiple waves of exposure and

covariate data as well as pre-baseline drinking histories, took an explicit target trial approach, and limited covariates in exposure weights to those known to temporally precede drinking measurement. The methodological strengths of this study makes it a valuable contribution to the literature.

Chapter 5 evaluated the relationship between alcohol consumption and risk for **type 2 diabetes** using Mendelian Randomisation, a design-based method to improve causal inference. This approach uses genes associated with an exposure of interest to proxy for that exposure, sidestepping potential confounding and reverse causation. The chapter represents the first non-linear MR analysis of this epidemiological relationship, finding support for a J-shaped relationship – although the finding was not robust to all sensitivity analysis. Crucially, the overall protective effect of low-level drinking was almost entirely driven by women, aligning with the sex-discrepant relationships found in previous observational work. Overall, the study suggests that while low-level drinking may have benefits for average glucose levels, when it comes to type 2 diabetes, at most there is a small benefit limited to women. Observational evidence was found to substantially overestimate the benefits, and underestimate the harms, of alcohol for diabetes risk.

6.1.1 Key themes

Three overarching conclusions can be drawn from the chapters of this thesis:

1. Even with improved methods for promoting causal inference, small benefits of low-level alcohol consumption persist for select health outcomes

Recent years have seen the tide of evidence begin to turn *away* from protective effects of alcohol for most conditions, translating into decreasing national low-risk guidelines over

time. Indeed, several recent studies employing enhanced methods for causal inference appear to conflict with traditionally observed J-shapes, such as a non-linear MR analysis of cardiovascular events and traits (Biddinger et al., 2022). Despite this trend, the three empirical studies in this thesis all produced some evidence of protective effects for low-to-moderate drinking (although the multiverse study is not in itself evidence of a causal relationship). For each condition, results were concordant with the observational literature, and plausible mechanisms can be invoked to explain the relationship. This strengthens the possibility that methodological bias may not be responsible for *all* findings of alcohol's protective effects, and that in the cases of inflammation, depression, and glucose levels/type 2 diabetes in women, small but real causal benefits of low-level drinking may exist**.**

2. Relationships between alcohol and long-term health outcomes are nuanced and multifaceted, requiring equally sophisticated analyses and specialised data

A clear theme emerging from all three empirical studies was that these epidemiological relationships are complex phenomena that require sophisticated analyses and interpretation using specialised data. To test alternative research parameter options, **Chapter 3** required a cohort with rich covariate data, quantity and frequency of alcohol consumption measured at multiple timepoints, and questions that differentiated between former and never drinkers – features that many datasets lack. The study demonstrated the considerable impact of seemingly innocuous research choices on the size, significance, and even direction of results when it comes to estimating alcohol's effect on inflammation. Applying the standard approach in epidemiological research, i.e., testing just one combination of research options or

at best a handful of sensitivity analyses, would not capture that complexity, and making general conclusions based on those results could be misleading.

Given that knowledge, **Chapter 4** embodied the principles of the target trial approach. Rather than asking the vague question: "Is low-level drinking protective against depression?", which may have very different answers depending on its operationalisation, the study explicitly defined all elements of the question that could be addressed by the data, as a comparable RCT would have. That is, it asked: "What is the impact of *stable* levels of consumption – defined according to official low-risk drinking guidelines – throughout *early-to-mid adulthood* on risk for probable depression and depressive symptoms at *age 50*?". This study too required specialised data: not only detailed alcohol consumption recorded at multiple timepoints, but recurring measurements of a validated measure of depressive symptoms and time-varying covariates.

The non-linear MR analysis in **Chapter 5** required individual-level genetic data from a very large sample – currently, only one publicly-available cohort meets these criteria. The study revealed substantially different alcohol–type 2 diabetes relationships for males and females. In this case, valid interpretation and effective health recommendations/policy must account for sex.

These kinds of nuances and considerations are relevant to other statistical and design-based methodologies, as well as other alcohol–long-term health relationships. Indeed, as the epidemiological field more broadly recognises a need to move from purely associational to causally focussed research, it's these kind of nuanced considerations for data collection, selection, analysis, and interpretation that must be front of mind.

3. There is no single 'perfect' method for causal inference

The empirical chapters of this thesis crystallised the knowledge that each methodological approach has unique *advantages* for promoting causal inference from observational data, as well as unique *weaknesses* (as detailed in Table 2.1). Multiverse analysis is crucial to establishing a robust association, but is still vulnerable to confounding, reverse causation, measurement error, and selection bias. The results from **Chapter 3** can therefore tell us about the stability of the alcohol–inflammation association, establishing a preliminary condition for causality, but cannot confirm a causal relationship itself. The G-methods, including marginal structural models, can elucidate causal relationships, but are reliant on appropriate choice of covariates and their accurate measurement. As such, despite accounting for time-varying alcohol consumption, covariates, and depression, the results from **Chapter 4** may still be invalid, particularly in the face of substantial unmeasured confounding (especially social support and interaction variables).

Mendelian Randomisation analyses can also elucidate causal relationships, and largely avoid confounding and reverse causation without the need to even control for covariates. However, they are reliant on the validity of genetic instruments, which can never be established with complete certainty. As a consequence, the analyses in **Chapter 5** may not hold if there is underlying pleiotropy (i.e., the genetic variants are acting on diabetes through paths other than alcohol consumption) or if there are unidentified associations between the genetic instruments and confounders.

The same is true if the alcohol–diabetes relationship is confounded by dynastic effects (when parental genotype influences offspring outcome via means other than offspring genotype),

assortative mating (when mate choice based on particular traits leads to biased exposure–outcome correlation in offspring), or – less likely given restriction to European subsample and control for principle components of ancestry – population stratification (when the gene–exposure relationship is biased by differences in frequency of gene and exposure distribution between ethnic subpopulations; Brumpton et al., 2020). Finally, unless genetic variants have been found that are exclusively associated with exposure at different ages, MR results can only be interpreted with respect to lifetime exposure, meaning the results from **Chapter 5** provide no insights on the effect of drinking *at a particular age* on risk for type 2 diabetes. This is a noteworthy limitation in the context of age-discrepant conclusions in the latest GBD alcohol report (see section 6.4.3).

Consequently, each method offers a unique contribution to developing a nuanced understanding of, and establishing causality for, alcohol–long-term health relationships, but is not sufficient in and of itself.

6.2 Implications

6.2.1 What would genuine protective effects of low-level drinking mean for policy?

This thesis provides *preliminary* evidence that some observed benefits of low-level alcohol consumption may have a causal basis. If these effects were substantiated, what would the implications be for public health policy? As introduced in **Chapter 1**, evidence of protective effects for select conditions is incorporated into the modelling that underpins comprehensive initiatives like the GBD reports, as well as country-specific low-risk drinking guidelines. In

practice, this means that condition-specific risk function estimates are taken from individual studies/meta-analyses and, incorporating each condition's contribution to disease burden in the population (e.g., through weighting), these functions are consolidated into a single risk curve. This curve is then used to identify the theoretical minimum risk exposure level, non-drinker equivalence level, and amount of alcohol consumption associated with any other risk thresholds of interest (e.g., a 1/100 risk of mortality prioritised in the Australian low-risk drinking guidelines). If an individual study or meta-analysis offers evidence of a J-shaped relationship, that too feeds into these calculations, in effect offsetting evidence of monotonically-increasing relationships for other conditions. Substantiating J-shaped findings as causal would validate this 'offsetting' practice, which as illustrated in Figure 1.7, can have considerable impacts on the estimated risks associated with various levels of drinking. However, even if confirmed to be causal, this thesis indicates that potential protective effects of alcohol are likely of small size. That is, while modelling exercises would account for the relatively high disease burden of conditions like depression and type 2 diabetes, they would also reflect the fact that low-level drinking does not reduce much of that burden. Meanwhile, for a host of other conditions, evidence that *any* amount of alcohol increases risk is strong. Conditions in this latter camp include several high-prevalence, high-burden cancers, such as breast, colorectal, stomach, and liver (Rehm et al., 2019), prompting countries like Ireland and South Korea to mandate cancer warnings on alcohol packaging. Indeed, even modelling that assumes protective effects for conditions like diabetes and cardiovascular outcomes still finds an overall net harm of alcohol at the population level (Sohi et al., 2021). Not only is the epidemiological evidence for these harms plain, but the mechanisms that underpin them

(e.g., the carcinogenic effects of alcohol's metabolite acetaldehyde) are better understood than those currently invoked to explain any potential benefits of low-level drinking.

6.2.2 What would genuine protective effects of low-level drinking mean for practice?

The existence of protective effects for some conditions, but not others, as well as their dependence on factors like age and sex, has important implications for clinical settings. Specifically, guidance on drinking may need to be tailored to the individual based on which of those conditions they have underlying risk factors for, as well as demographic characteristics (van der Heide et al., 2023). Importantly, despite pronouncements from some that "...physicians should counsel lifelong non-drinkers at about 40 to 50 years of age to relax and take a drink a day..." (E. Rubin, 2014, p.2890), abstainers should not be encouraged to initiate alcohol use given the potential for unintentionally developing risky or above-guidelines drinking, if not alcohol use disorder, which remain strongly associated with poor health outcomes (Greenfield & Kerr, 2014; Mukamal, 2010).

Finally, the suggested mechanisms via which the protective effects of low-level drinking operate (see Table 1.2) are known to be modified by a range of other exposures. That is, alcohol may be just one set of keys to these mechanisms, but other keys likely exist that do not also carry adverse effects (e.g., increased cancer risk or the potential for developing alcohol use disorder) and would therefore be more appropriate for clinicians to recommend (Mezue et al., 2023). For example, if social interaction does mediate the observed benefits of low-to-moderate alcohol consumption against depression, clinicians could promote other interventions to improve social support as a substitute for alcohol consumption (this would also be facilitated by a greater emphasis on serving and consuming alcohol-free beverages at pubs and clubs). Similarly, if reduced inflammation is substantiated as a key mechanism for

alcohol's protection against depression and other chronic conditions, clinicians may recommend dietary changes and exercise regimes (Paulson et al., 2018). In this sense, findings of protective effects of alcohol consumption can be viewed as encouraging proof of concept that risk for high burden diseases can be successfully modified by lifestyle interventions (van der Heide et al., 2023).

6.2.3 Implications for research: Using triangulation to consolidate complementary strengths and weaknesses

The concept of triangulation – "the strategic use of multiple approaches to address one question" (Munafò & Davey Smith, 2018, p.400) – is a direct response to the fact that there is no single study, nor single method, that is sufficient for causal inference. Triangulation was employed in **Chapter 2** when comparing the findings made via causal inference methods to those of the broader observational literature. Triangulation generally refers to the comparison of findings across several separate studies, but can also be applied within a single study, e.g., the comparison of non-linear MR, linear MR, and observational models conducted in **Chapter 5**. Crucially, the approaches triangulated must have different underlying assumptions/sources of bias, and therefore complementary strengths and weaknesses, which is ideally achieved by employing both statistical and design-based methods to promote causal inference (Hammerton & Munafò, 2021). Consistent findings increase confidence in a causal conclusion, while disparity indicates the opposite.

Relatedly, comparison of findings across settings, cohorts, and researcher-defined parameters can help to establish consensus (Glymour & Whitmer, 2019; although genuine differences in epidemiological relationships based on sex, age, and ethnicity must be considered). The task

of triangulation and diverse evidence accumulation more broadly is therefore crucial to making any genuine headway in elucidating relationships between alcohol and long-term health.

While the chapters of this thesis focussed on diverse health outcomes, its structure can be used as a template for working towards triangulation when it comes to tackling a single epidemiological question. That is, first, a systematic assessment of previous research is required to evaluate the existing evidence base and build a foundation for triangulation. Then, the stability of the longitudinal association between exposure and outcome should be assessed. If stable, a possible causal relationship should be probed using both statistical (e.g., MSMs) and design-based (e.g., MR) methods. Finally, acknowledging the strengths and weaknesses of all approaches used, these findings should be compared and integrated with the existing evidence base. Crucially, this is an *ongoing* process, with triangulation continually reperformed as new datasets, analysis methods, and findings becomes available.

6.3 Strengths and limitations of the research

While it was not possible to be exhaustive in the approaches and epidemiological relationships explored in the empirical studies, this thesis did sample one each of preliminary, statistically-based, and design-based methods for causal inference, each targeting alcohol–long-term health relationships of critical importance. Efforts were made to apply the most up-to-date statistical methodologies, with the empirical chapters incorporating a range of tests and features developed in the last decade: vibration of effects, specification curve, and multiverse analysis methods in **Chapter 3**, E-values and covariate-

balanced propensity scores in **Chapter 4**, and two-sample MR sensitivity tests, non-linear MR analysis, and the doubly-ranked stratification method in **Chapter 5**.

The three large, population-based datasets used in this thesis were especially well-suited to answering the relevant questions in each empirical chapter, and were rich in their measurement of alcohol consumption, covariates, and outcomes. However, despite being among the best observational cohorts to answer these research questions, each was still limited: the 1970 British Birth Cohort Study provided a single inflammatory marker, limiting generalisability of results to inflammation more broadly; the National Longitudinal Survey of Youth lacked social support and interaction data, and changed during the study from year-based measurement occasions to age-based measurement occasions; and in the UK Biobank, earlier-recruited occasional drinkers were not asked about usual volume consumed, and primary care linkage was only achieved for roughly 50% of the cohort. The ideal observational cohort would offer prospective, consistent alcohol consumption measurements from an early age, as well as comprehensive primary care and hospital record linkage for ascertainment of health outcomes – features that are often lacking in publicly available data due to their administrative and financial burden.

Moving from the quality of available data to the analyses themselves, a limitation common to all three empirical studies is that the alcohol exposure variables were – for pragmatic reasons – coarser than what may be needed to fully capture the nuances of the relationships under investigation. Firstly, the studies in **Chapters 3 and 4** relied on a categorised exposure variable, where individuals were classed as either lifetime abstainers, former drinkers, less-than-weekly drinkers, low-to-moderate drinkers, or above-guidelines drinkers.

These categories are broad, and there are likely important differences lost between individuals sitting at different levels of drinking *within* these categories. However, this categorisation was necessary for the multiverse analysis, where the specification curve plots and associated metrics rely on single effect estimates and therefore cannot accommodate non-linear models based on continuous data. Categorisation was also the most reasonable way to examine non-linearity in the alcohol–depression study; generating inverse probability weights is complex for a continuous exposure as they are sensitive to model misspecification and prone to assumption violations (Huling et al., 2021; A. I. Naimi et al., 2014; Thoemmes & Ong, 2016).

Similarly, *pattern of consumption* was only crudely incorporated: in **Chapters 3 and 4**, heavy episodic drinking (HED) was part of the categorisation criteria (the presence of which automatically placed individuals in the highest drinking category). Particularly problematic was that the Mendelian Randomisation paper in **Chapter 5** could not consider pattern at all, given the genetic instruments strictly proxied for drinks per week. Evidence suggests that drinking pattern – how alcohol consumption is spread over a period of time, over and above the volume consumed in that period of time – can substantially modify the effects of average weekly/monthly volume on health outcomes. For conditions like cardiovascular disease, research shows that HED mitigates any protective effects (Bagnardi et al., 2008; Manthey et al., 2022; Roerecke & Rehm, 2014). Interestingly though, for other outcomes such as liver cirrhosis, recent evidence indicates that spreading drinking across the week (i.e., light daily drinking) can lead to *greater* risk than the same total weekly intake split over several heavier drinking days (Simpson et al., 2019), suggesting benefits of 'giving the body a break' with alcohol-free days. Another aspect of pattern that appears to moderate the impact of drinking

level is whether alcohol is consumed with meals (Ma et al., 2022). Findings like these strengthen the case for pattern of consumption being a focus of future research, including consideration of potential divergence in the effects of pattern between organ systems/disease processes.

6.4 Future directions

6.4.1 Substantiating protective effects

This thesis' preliminary findings on the relationships between alcohol and inflammation, alcohol and mental health, and alcohol and glucose levels/type 2 diabetes require urgent corroboration to establish causality. High-quality estimates for mental health are critically needed for modelling and health policy purposes. For example, both the latest GBD alcohol modelling (Bryazka et al., 2022) and Australian low-risk drinking guidelines (National Health and Medical Research Council, 2020) do not incorporate depression or anxiety due to a lack of evidence. While inflammation is not considered a 'condition' itself for modelling and health policy purposes, it is a process increasingly implicated in myriad downstream health outcomes, and is therefore of keen epidemiological interest, justifying future research. In the triangulation endeavour, applying complementary methods to those employed in this thesis are a clear next step. For alcohol's relationship with inflammation and mental health outcomes, that would involve further studies leveraging the potential of non-linear MR and very large biobank data. For type 2 diabetes, statistically-based approaches for causal inference are required for triangulation with the MR findings.

6.4.2 Identifying mediating pathways

Next, for those conditions for which there is emerging causal evidence of protective effects of low-level drinking, suggested mechanisms of those benefits remain largely untested in a long-term context. Causal mediation analysis, introduced briefly in **Chapter 2** but otherwise unexplored in this thesis, is robust to exposure–mediator interaction, can incorporate non-linearity, explicitly tests assumptions, and has the potential to elucidate mechanisms involved in alcohol–long-term health relationships. Two relationships are of particular interest and import: first, benefits of low-level drinking are frequently observed for cardiovascular disease (CVD) and sub-types – conditions with considerable associated burden. Ethanol's anti-inflammatory action is a suggested mediator, but this is unsubstantiated. Second, for the apparent benefits conferred by low-to-moderate drinking against several conditions like dementia and depression, competing explanations promote either alcohol's benefits for social support/and interaction as a mediator, or else invoke direct biological effects (Visontay, Mewton, Slade, et al., 2023). For these conditions, the relative contributions of the two pathways remains largely unclear.

The elucidation of mediating mechanisms is important for three reasons. Firstly, to establish the causal nature of a relationship, evidence of a plausible mechanism is a prerequisite (Hill, 1965). Next, mediation analyses could identify possible downstream risk/protective factors which could be modified by other interventions – as explained in 6.2.2, alcohol may be one set of keys to these mechanisms, but other, safer keys likely exist. Finally, it is important to observe whether the mechanisms at play vary over the spectrum of alcohol consumption. Notably, alcohol at heavier doses begins to exert opposite effects on many of the same

mechanisms posited to underly low-to-moderate alcohol consumption's benefits (Szabo, 2007; Tizabi et al., 2018). Directly testing whether this extends to the mediation of downstream outcomes would establish whether observed J-shaped relationships are best understood as single dose-response relationships, or instead, as the amalgamation of several underlying discrete relationships (Wood et al., 2018).

6.4.3 Towards personalised guidelines: Disaggregated research

Just as this thesis began to explore sex-based differences in alcohol–long-term health relationships, which is crucial information for modelling and health policy, establishing whether these relationships vary by age remains of keen interest, especially as health authorities around the world begin to issue limited age-specific guidelines (e.g., age <65 vs >65). Contentiously, the 2022 GBD alcohol consumption modelling found differences by age: even a fraction of a standard drink per day put those under 40 at increased health risk, while for those aged over 40, low-level alcohol consumption (roughly half a standard drink per day) actually aligned with better health than non-drinking (Bryazka et al., 2022). However, while we know that alcohol consumption's effects on mid-to-older adult health conditions *differ based on when that drinking occurred* (Manthey et al., 2022), the GBD report uses risk estimates that do not incorporate age at alcohol consumption. Instead, the GBD age-specific findings result solely from consideration of differential underlying disease burdens; age is factored into absolute risk, but not relative risk (Manthey et al., 2022). To improve the value of modelling exercises for policymakers and individuals, and in line with the UN's 2030 Agenda for Sustainable Development, robust, age-disaggregated risk inputs are required. This work can be achieved using harmonised life course data, such that alcohol coverage is

spread across adulthood while age at outcome/survival period are aligned. These efforts should also be combined with statistically-based methods to improve causal inference.

More generally, future research should aim to produce country-, sex-, and age-specific estimates of alcohol–long-term health relationships for incorporation into public health policy, but also to inform decision-making for clinicians and individuals.

6.5 Conclusions

This thesis offers a multifaceted exploration of how to evaluate alcohol–long-term health relationships through a causal lens, and provides preliminary evidence on what causally-focussed research can tell us about the functional form of these relationships. Three key conclusions emerged from this thesis. First, that even with improved methods for promoting causal inference, small benefits of low-level alcohol consumption persist for select health outcomes – most clearly, for depression, but possibly also for chronic inflammation and for type 2 diabetes in women. At the same time, it is also clear that standard epidemiological methods often substantially overestimate the benefits, and underestimate the harms, of drinking. Second, relationships between alcohol and long-term health outcomes are nuanced and multifaceted, requiring equally sophisticated data, analyses, and interpretation. Careful consideration of these complexities is required to make progress in elucidating alcohol–long-term health relationships, as well as for effective public health policy and clinical practice. Finally, it is clear that there is no single 'perfect' method for causal inference. Each has its

own strengths and weaknesses, which is why the process of triangulation is critical to establishing robust causal conclusions.

Sceptical researchers, clinicians, and policymakers have increasingly called for the apparent benefits of moderate drinking to be subjected to rigorous, causally-focussed scrutiny. This thesis answers that call, and in doing so lays the foundations for future research on alcohol's J-shaped curve.

References

Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abegaz, K. H., Abolhassani, H., Aboyans, V., Abreu, L. G., Abrigo, M. R. M., Abualhasan, A., Abu-Raddad, L. J., Abushouk, A. I., Adabi, M., Adekanmbi, V., Adeoye, A. M., Adetokunboh, O. O., ... Wiangkham, T. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *396*(10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

Adams, C., Conigrave, J. H., Lewohl, J., Haber, P., & Morley, K. C. (2020). Alcohol use disorder and circulating cytokines: A systematic review and meta-analysis. *Brain, Behavior, and Immunity*, *89*, 501–512. https://doi.org/10.1016/j.bbi.2020.08.002

Ahlskog, R., & Oskarsson, S. (2023). Quantifying bias from measurable and unmeasurable confounders across three domains of individual determinants of political preferences. *Political Analysis*, *31*(2), 181–194.

Alati, R., Lawlor, D. A., Najman, J. M., Williams, G. M., Bor, W., & O'callaghan, M. (2005). Is there really a 'J-shaped' curve in the association between alcohol consumption and symptoms of depression and anxiety? Findings from the Mater-University Study of Pregnancy and its outcomes. *Addiction*, *100*(5), 643–651.

Albert, M. A., Glynn, R. J., & Ridker, P. M. (2003). Alcohol consumption and plasma concentration of C-reactive protein. *Circulation*, *107*(3), 443–447. https://doi.org/10.1161/01.CIR.0000045669.16499.EC

Ali, M. S., Groenwold, R. H. H., & Klungel, O. H. (2014). Propensity score methods and unobserved covariate imbalance: Comments on "squeezing the balloon." *Health Services Research*, *49*(3), 1074–1082. https://doi.org/10.1111/1475-6773.12152

Allison, P. D. (2005). *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. SAS Institute Inc.

Andreasson, S., Chikritzhs, T., Dangardt, F. H. H., Naimi, T., & Stockwell, T. (2014). Evidence about health effects of "moderate" alcohol consumption: Reasons for scepticism and public health implications. *Alcohol and Society*, *6*.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Source: Journal of the American Statistical Association*, *91*(434), 444–455.

Angus, C., Henney, M., Meier, P., Brennan, A., & Holmes, J. (2019). Mortality and morbidity risks from alcohol consumption in Australia: Analyses using an Australian

adaptation of the Sheffield Alcohol Policy Model (v2.7) to inform the development of new alcohol guidelines. *Sheffield: ScHARR, University of Sheffield*.

Ansar, W., & Ghosh, S. (2013). C-reactive protein and the biology of disease. *Immunologic Research*, *56*(1), 131–142. https://doi.org/10.1007/s12026-013-8384-0

Aris, I. M., Sordillo, J. E., Rifas-Shiman, S. L., Young, J. G., Gold, D. R., Camargo, C. A., Hivert, M. F., & Oken, E. (2021). Childhood patterns of overweight and wheeze and subsequent risk of current asthma and obesity in adolescence. *Paediatric and Perinatal Epidemiology*, *35*(5), 569–577. https://doi.org/10.1111/ppe.12760

Australian Institute of Health and Welfare. (2020). *National Drug Strategy Household Survey 2019*. AIHW.

Avan, A., Tavakoly Sany, S. B., Ghayour-Mobarhan, M., Rahimi, H. R., Tajfard, M., & Ferns, G. (2018). Serum C-reactive protein in the prediction of cardiovascular diseases: Overview of the latest clinical studies and public health practice. *Journal of Cellular Physiology*, *233*(11), 8508–8525.

Bagnardi, V., Zatonski, W., Scotti, L., La Vecchia, C., & Corrao, G. (2008). Does drinking pattern modify the effect of alcohol on the risk of coronary heart disease? Evidence from a meta-analysis. *Journal of Epidemiology & Community Health*, *62*(7), 615–619.

Barber, V. (2016). *Young people's beliefs about the health effects of different alcoholic beverages: an exploratory comparison of the UK and France*. [Doctoral dissertation, Kingston University]. Kingston University Research Repository. https://eprints.kingston.ac.uk/id/eprint/37411/.

Bell, S., Mehta, G., Moore, K., & Britton, A. (2017). Ten-year alcohol consumption typologies and trajectories of C-reactive protein, interleukin-6 and interleukin-1 receptor antagonist over the following 12 years: a prospective cohort study. *Journal of Internal Medicine*, *281*(1), 75–85. https://doi.org/10.1111/joim.12544

Bellos, S., Skapinakis, P., Rai, D., Zitko, P., Araya, R., Lewis, G., Lionis, C., & Mavreas, V. (2016). Longitudinal association between different levels of alcohol consumption and a new onset of depression and generalized anxiety disorder: Results from an international study in primary care. *Psychiatry Research*, *243*, 30–34. https://doi.org/10.1016/j.psychres.2016.05.049

Benedetto, U., Head, S. J., Angelini, G. D., & Blackstone, E. H. (2018). Statistical primer: Propensity score matching and its alternatives. *European Journal of Cardio-Thoracic Surgery*, *53*(6), 1112–1117. https://doi.org/10.1093/ejcts/ezy167

Bergeron, C. D. , Tremblay-Antoine, C. , Morin, R., April, N., Leclair, V., Cloutier, A., Rancourt, M.-A., Fréchet, N., Poirier, W., & Ouellet, C. (2021). *La consommation*

d'alcool: qu'en pense la population du Québec? INSPQ: Institut national de santé publique du Québec.

Beulens, J. W. J., Rimm, E. B., Hu, F. B., Hendriks, H. F. J., & Mukamal, K. J. (2008). Alcohol consumption, mediating biomarkers, and risk of type 2 diabetes among middle-aged women. *Diabetes Care*, *31*(10), 2050–2055.

Biddinger, K. J., Emdin, C. A., Haas, M. E., Wang, M., Hindy, G., Ellinor, P. T., Kathiresan, S., Khera, A. v., & Aragam, K. G. (2022). Association of Habitual Alcohol Intake with Risk of Cardiovascular Disease. *JAMA Network Open*, *5*(3), E223849. https://doi.org/10.1001/jamanetworkopen.2022.3849

Boden, J. M., & Fergusson, D. M. (2011). Alcohol and depression. *Addiction*, *106*(5), 906–914. https://doi.org/10.1111/j.1360-0443.2010.03351.x

Brand-Miller, J. C., Fatima, K., Middlemiss, C., Bare, M., Liu, V., Atkinson, F., & Petocz, P. (2007). Effect of alcoholic beverages on postprandial glycemia and insulinemia in lean, young, healthy adults. *The American Journal of Clinical Nutrition*, *85*(6), 1545–1551.

Brien, S. E., Ronksley, P. E., Turner, B. J., Mukamal, K. J., & Ghali, W. A. (2011). Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *BMJ*, *342*, d636.

Brière, F. N., Rohde, P., Seeley, J. R., Klein, D., & Lewinsohn, P. M. (2014). Comorbidity between major depression and alcohol use disorder from adolescence to adulthood. *Comprehensive Psychiatry*, *55*(3), 526–533. https://doi.org/10.1016/j.comppsych.2013.10.007

Brookhart, M. A., Wyss, R., Layton, J. B., & Stürmer, T. (2013). Propensity score methods for confounding control in nonexperimental research. *Circulation: Cardiovascular Quality and Outcomes*, *6*(5), 604–611. https://doi.org/10.1161/CIRCOUTCOMES.113.000359

Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., & Boomsma, D. I. (2020). Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature Communications*, *11*(1), 3519.

Bryazka, D., Reitsma, M. B., Griswold, M. G., Abate, K. H., Abbafati, C., Abbasi-Kangevari, M., Abbasi-Kangevari, Z., Abdoli, A., Abdollahi, M., Abdullah, A. Y. M., Abhilash, E. S., Abu-Gharbieh, E., Acuna, J. M., Addolorato, G., Adebayo, O. M., Adekanmbi, V., Adhikari, K., Adhikari, S., Adnani, Q. E. S., ... Gakidou, E. (2022). Population-level risks of alcohol consumption by amount, geography, age, sex, and year:

a systematic analysis for the Global Burden of Disease Study 2020. *The Lancet*, *400*(10347), 185–235. https://doi.org/10.1016/S0140-6736(22)00847-9

Burgess, S., Davies, N. M., & Thompson, S. G. (2014). Instrumental variable analysis with a nonlinear exposure-outcome relationship. *Epidemiology*, *25*(6), 877–885. https://doi.org/10.1097/EDE.0000000000000161

Burgess, S., Malik, R., Liu, B., Mason, A. M., Georgakis, M. K., Dichgans, M., & Gill, D. (2021). Dose–response relationship between genetically proxied average blood glucose levels and incident coronary heart disease in individuals without diabetes mellitus. *Diabetologia*, *64*, 845–849.

Burgess, S., & Thompson, S. G. (2021). *Mendelian randomization: methods for causal inference using genetic variants*. CRC Press.

Calder, P. C., Ahluwalia, N., Albers, R., Bosco, N., Bourdet-Sicard, R., Haller, D., Holgate, S. T., Jönsson, L. S., Latulippe, M. E., Marcos, A., Moreines, J., Mrini, C., Müller, M., Pawelec, G., van Neerven, R. J. J., Watzl, B., & Zhao, J. (2013). A Consideration of Biomarkers to be Used for Evaluation of Inflammation in Human Nutritional Studies. *British Journal of Nutrition*, *109*(SUPPL. S1). https://doi.org/10.1017/S0007114512005119

Callinan, S., Chikritzhs, T., & Livingston, M. (2019). Consistency of drinker status over time: Drinking patterns of ex-drinkers who describe themselves as lifetime abstainers. *Journal of Studies on Alcohol and Drugs*, *80*(5), 552–556. https://doi.org/10.15288/jsad.2019.80.552

Calvo, E., Allel, K., Staudinger, U. M., Castillo-carniglia, A., Keyes, K. M. (2021). Cross-country differences in age trends in alcohol consumption among older adults: a cross-sectional study of individuals aged 50 years and older in 22 countries. *Addiction*, *116*(6), 1399–1412.

Carlsson, S., Hammar, N., Grill, V., & Kaprio, J. (2003). Alcohol Consumption and the Incidence of Type 2 Diabetes A 20-year follow-up of the Finnish Twin Cohort Study. *Diabetes Care*, *26*(10), 2–7.

Chen, J., Spracklen, C. N., Marenne, G., Varshney, A., Corbin, L. J., Luan, J., Willems, S. M., Wu, Y., Zhang, X., & Horikoshi, M. (2021). The trans-ancestral genomic architecture of glycemic traits. *Nature Genetics*, *53*(6), 840–860.

Chikritzhs, T., Fillmore, K., & Stockwell, T. (2009). A healthy dose of scepticism: Four good reasons to think again about protective effects of alcohol on coronary heart disease. *Drug and Alcohol Review*, *28*(4), 441–444. https://doi.org/10.1111/j.1465-3362.2009.00052.x

Cho, Y., Shin, S.-Y., Won, S., Relton, C. L., Davey Smith, G., & Shin, M.-J. (2015). Alcohol intake and cardiovascular risk factors: a Mendelian randomisation study. *Scientific Reports*, *5*(1), 18422.

Choi, K. W., Stein, M. B., Nishimi, K. M., Ge, T., Coleman, J. R. I., Chen, C. Y., Ratanatharathorn, A., Zheutlin, A. B., Dunn, E. C., Breen, G., Koenen, K. C., Smoller, J. W., Agee, M., Alipanahi, B., Auton, A., Bell, R. K., Bryc, K., Elson, S. L., Fontanillas, P., ... Wilson, C. H. (2020). An exposure-wide and mendelian randomization approach to identifying modifiable factors for the prevention of depression. *American Journal of Psychiatry*, *177*(10), 944–954. https://doi.org/10.1176/appi.ajp.2020.19111158

Chu, L., Ioannidis, J. P. A., Egilman, A. C., Vasiliou, V., Ross, J. S., & Wallach, J. D. (2020). Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk. *International Journal of Epidemiology*, *49*(2), 608–618. https://doi.org/10.1093/ije/dyz271

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, *168*(6), 656–664. https://doi.org/10.1093/aje/kwn164

Coventry, B. J., Ashdown, M. L., Quinn, M. A., Markovic, S. N., Yatomi-Clarke, S. L., & Robinson, A. P. (2009). CRP identifies homeostatic immune oscillations in cancer patients: a potential treatment targeting tool? *Journal of Translational Medicine*, *7*(1), 1–8.

Cruwys, T., Dingle, G. A., Haslam, C., Haslam, S. A., Jetten, J., & Morton, T. A. (2013). Social group memberships protect against future depression, alleviate depression symptoms and prevent depression relapse. *Social Science & Medicine*, *98*, 179–186.

Davey Smith, G., Holmes, M. V, Davies, N. M., & Ebrahim, S. (2020). Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *European Journal of Epidemiology*, *35*(2), 99–111.

Davies, N. M., Holmes, M. V, & Davey Smith, G. (2018). Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ (Online)*, *362*. https://doi.org/10.1136/bmj.k601

de Timary, P., Stärkel, P., Delzenne, N. M., & Leclercq, S. (2017). A role for the peripheral immune system in the development of alcohol use disorders? *Neuropharmacology*, *122*, 148–160.

Delgado-Rodríguez, M., & Llorca, J. (2004). Bias. *Journal of Epidemiology and Community Health*, *58*(8), 635–641. https://doi.org/10.1136/jech.2003.008466

Department of Health. (2016). *UK chief medical officers' low risk drinking guidelines*. Department of Health London.

Dickerman, B. A., Coseo, S., Markku, M., Pukkala, E., Mucci, L. A., & Kaprio, J. (2016). Alcohol intake, drinking patterns, and prostate cancer risk and mortality : a 30-year prospective cohort study of Finnish twins. *Cancer Causes & Control*, *27*(9), 1049–1058. https://doi.org/10.1007/s10552-016-0778-6

Diop, A., Sirois, C., Guertin, J. R., & Talbot, D. (2021). *Marginal structural models with Latent Class Growth Modeling of Treatment Trajectories*. ArXiv. https://doi.org/https://doi.org/10.48550/arXiv.2105.12720

Dodgeon, B. , Elliott, J. , Hancock, M. , Johnson, J. , Parsons, S. , Peters, A. , & Shepherd, P. (2020). *1970 British Cohort Study Age 34 Sweep User Guide*. UCL Centre for Longitudinal Studies.

D'Onofrio, B. M., Sjölander, A., Lahey, B. B., Lichtenstein, P., & Öberg, A. S. (2020). Accounting for Confounding in Observational Studies. *Annual Review of Clinical Psychology*, *16*(1), 25–48. https://doi.org/10.1146/annurev-clinpsy-032816-045030

Dugoff, E. H., Schuler, M., & Stuart, E. A. (2014). Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Services Research*, *49*(1), 284–303. https://doi.org/10.1111/1475-6773.12090

Dunning, T. (2012). *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press.

Eastwood, S. V, Mathur, R., Atkinson, M., Brophy, S., Sudlow, C., Flaig, R., de Lusignan, S., Allen, N., & Chaturvedi, N. (2016). Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. *PloS One*, *11*(9), e0162388.

Ebrahim, S., & Davey Smith, G. (2008). Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Human Genetics*, *123*, 15–33.

Edenberg, H. J., & McClintick, J. N. (2018). Alcohol Dehydrogenases, Aldehyde Dehydrogenases, and Alcohol Use Disorders: A Critical Review. *Alcoholism: Clinical and Experimental Research*, *42*(12), 2281–2297. https://doi.org/10.1111/acer.13904

Elliott, J., & Shepherd, P. (2006). Cohort profile: 1970 British Birth Cohort (BCS70). *International Journal of Epidemiology*, *35*(4), 836–843. https://doi.org/10.1093/ije/dyl174

Ellison, R. C., Grønbæk, M., & Skovenborg, E. (2021). Using Mendelian randomization to evaluate the effects of alcohol consumption on the risk of coronary heart disease. *Drugs and Alcohol Today*, *21*(1), 84–95.

Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2007). Unemployment and suicidal behavior in a New Zealand birth cohort: a fixed effects regression analysis. *Crisis*, *28*(2), 95–101.

Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2009). Tests of Causal Links Between Alcohol Abuse or Dependence and Major Depression. *Archives of General Psychiatry*, *66*(3), 260–266.

Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2015). Psychosocial sequelae of cannabis use and implications for policy : findings from the Christchurch Health and Development Study. *Social Psychiatry and Psychiatric Epidemiology*, *50*(9), 1317–1326. https://doi.org/10.1007/s00127-015-1070-x

Furman, D., Campisi, J., Verdin, E., Carrera-Bastos, P., Targ, S., Franceschi, C., Ferrucci, L., Gilroy, D. W., Fasano, A., Miller, G. W., Miller, A. H., Mantovani, A., Weyand, C. M., Barzilai, N., Goronzy, J. J., Rando, T. A., Effros, R. B., Lucia, A., Kleinstreuer, N., & Slavich, G. M. (2019). Chronic inflammation in the etiology of disease across the life span. *Nature Medicine*, *25*(12), 1822–1832. https://doi.org/10.1038/s41591-019-0675-0

Gage, S. H., Munafo, M. R., & Davey Smith, G. (2016). Causal Inference in Developmental Origins of Health and Disease (DOHaD) Research. *Annual Review of Psychology*, *67*, 567–585. https://doi.org/10.1146/annurev-psych-122414-033352

Garcia-Huidobro, D., & Michael Oakes, J. (2017). Squeezing observational data for better causal inference: Methods and examples for prevention research. *International Journal of Psychology*, *52*(2), 96–105. https://doi.org/10.1002/ijop.12275

Gémes, K., Forsell, Y., Janszky, I., László, K. D., Lundin, A., Ponce De Leon, A., Mukamal, K. J., & Möller, J. (2019). Moderate alcohol consumption and depression-a longitudinal population-based study in Sweden. *Acta Psychiatrica Scandinavica*, *139*(6), 526–535.

Gepner, Y., Golan, R., Harman-Boehm, I., Henkin, Y., Schwarzfuchs, D., Shelef, I., Durst, R., Kovsan, J., Bolotin, A., & Leitersdorf, E. (2015). Effects of initiating moderate alcohol intake on cardiometabolic risk in adults with type 2 diabetes: a 2-year randomized, controlled trial. *Annals of Internal Medicine*, *163*(8), 569–579.

Giollabhui, N. M., Ellman, L. M., Coe, C. L., Byrne, M. L., Abramson, L. Y., & Alloy, L. B. (2020). To exclude or not to exclude: Considerations and recommendations for C-reactive protein values higher than 10 mg/L. *Brain, Behavior, and Immunity*, *87*, 898.

Glymour, M. M., & Whitmer, R. A. (2019). Using cross-cultural studies to improve evidence on dementia prevention: Lessons from the special issue sponsored by the International Research Network on Dementia Prevention (IRNDP). *Journal of Alzheimer's Disease*, *70*(Suppl 1), S5.

González-Reimers, E., Santolaria-Fernández, F., Martín-González, M. C., Fernández-Rodríguez, C. M., & Quintero-Platt, G. (2014). Alcoholism: A systemic proinflammatory condition. *World Journal of Gastroenterology*, *20*(40), 14660–14671. https://doi.org/10.3748/wjg.v20.i40.14660

Greenfield, T. K., & Kerr, W. C. (2014). Physicians' prescription for lifetime abstainers aged 40 to 50 to take a drink a day are not yet justified. *Alcoholism, Clinical and Experimental Research*, *38*(12), 2893.

Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: A problem hiding in plain sight. *BMJ* , *353*, 1–6. https://doi.org/10.1136/bmj.i1981

Griswold, M. G., Fullman, N., Hawley, C., Arian, N., Zimsen, S. R. M., Tymeson, H. D., Venkateswaran, V., Tapp, A. D., Forouzanfar, M. H., & Salama, J. S. (2018). Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *392*(10152), 1015–1035.

Gunasekara, F. I., Richardson, K., Carter, K., & Blakely, T. (2014). Fixed effects analysis of repeated measures data. *International Journal of Epidemiology*, *43*(1), 264–269. https://doi.org/10.1093/ije/dyt221

Hamer, M., O'Donovan, G., Batty, G. D., & Stamatakis, E. (2020). Estimated cardiorespiratory fitness in childhood and cardiometabolic health in adulthood: 1970 British Cohort Study. *Scandinavian Journal of Medicine and Science in Sports*, *30*(5), 932–938. https://doi.org/10.1111/sms.13637

Hammerton, G., & Munafò, M. R. (2021). Causal inference with observational data: the need for triangulation of evidence. *Psychological Medicine*, *51*(4), 563–578.

Handing, E. P., Andel, R., Kadlecova, P., Gatz, M., & Pedersen, N. L. (2015). Midlife Alcohol Consumption and Risk of Dementia Over 43 Years of Follow-Up : A Population-Based Study From the Swedish Twin Registry. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *70*(10), 1248–1254. https://doi.org/10.1093/gerona/glv038

Hanscombe, K. B., Coleman, J. R. I., Traylor, M., & Lewis, C. M. (2019). ukbtools: An R package to manage and query UK Biobank data. *PLoS One*, *14*(5), e0214311.

Harrell Jr, F. E., Harrell Jr, M. F. E., & Hmisc, D. (2017). Package 'rms.' *Vanderbilt University*, *229*, Q8.

He, J., & Crews, F. T. (2008). Increased MCP-1 and microglia in various regions of the human alcoholic brain. *Experimental Neurology*, *210*(2), 349–358. https://doi.org/10.1016/j.expneurol.2007.11.017

Health Council of the Netherlands. (2015). *Dutch dietary guidelines 2015*. Health Council of the Netherlands.

Hemani, G., Tilling, K., & Davey Smith, G. (2017). Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genetics*, *13*(11), e1007081.

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., & Langdon, R. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife*, *7*, e34408.

Hendriks, H. F. J. (2007). Moderate Alcohol Consumption and Insulin Sensitivity: Observations and Possible Mechanisms. *Annals of Epidemiology*, *17*(5 SUPPL.). https://doi.org/10.1016/j.annepidem.2007.01.009

Hernáez, Á., Rogne, T., Skåra, K. H., Håberg, S. E., Page, C. M., Fraser, A., Burgess, S., Lawlor, D. A., & Magnus, M. C. (2021). Body mass index and subfertility: multivariable regression and Mendelian randomization analyses in the Norwegian Mother, Father and Child Cohort Study. *Human Reproduction*, *36*(12), 3141–3151.

Hernán, M. A. (2018). The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *American Journal of Public Health*, *108*(5), 616–619.

Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*, *183*(8), 758–764.

Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If.* Chapman & Hill/CRC.

Hill, A. B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine.*, *58*(2), 295–300.

Holmes, M. V, Dale, C. E., Zuccolo, L., Silverwood, R. J., Guo, Y., Ye, Z., Prieto-Merino, D., Dehghan, A., Trompet, S., & Wong, A. (2014). Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ*, *349*, g4164.

Huang, B.-H., Inan-Eroglu, E., Shaban, R. Z., Hamer, M., Britton, A., & Stamatakis, E. (2022). Alcohol intake and mortality risk of COVID-19, pneumonia, and other infectious diseases: An analysis of 437191 UK biobank participants. *Preventive Medicine Reports*, *26*, 101751.

Huling, J. D., Greifer, N., & Chen, G. (2021). *Independence weights for causal inference with continuous exposures*. ArXiv. https://doi.org/https://doi.org/10.48550/arXiv.2107.07086

Ilomäki, J., Hajat, A., Kauhanen, J., Kurl, S., Kaufman, J. S., Tuomainen, T. P., & Korhonen, M. J. (2012). Relationship between alcohol consumption and myocardial infarction among ageing men using a marginal structural model. *European Journal of Public Health*, *22*(6), 825–830. https://doi.org/10.1093/eurpub/ckr013

Imhof, A., Froehlich, M., Brenner, H., Boeing, H., Pepys, M. B., & Koenig, W. (2001). Effect of alcohol consumption on systemic markers of inflammation. *Lancet*, *357*(9258), 763–767. https://doi.org/10.1016/S0140-6736(00)04170-2

Jiang, T., Gill, D., Butterworth, A. S., & Burgess, S. (2022). An empirical investigation into the impact of winner's curse on estimates from Mendelian randomization. *MedRxiv*, https://doi.org/10.1101/2022.08.05.22278470

Julien, J., Ayer, T., Bethea, E. D., Tapper, E. B., & Chhatwal, J. (2020). Projected prevalence and mortality associated with alcohol-related liver disease in the USA, 2019–40: a modelling study. *The Lancet Public Health*, *5*(6), e316–e323. https://doi.org/10.1016/S2468-2667(20)30062-1

Kadlecová, P., Andel, R., Mikulík, R., Handing, E. P., & Pedersen, N. L. (2015). Alcohol Consumption at Midlife and Risk of Stroke During 43 Years of Follow-Up: Cohort and Twin Analyses. *Stroke*, *46*(3), 627–633. https://doi.org/10.1161/STROKEAHA.114.006724

Kerr, W. C., Lui, C. K., Williams, E., Ye, Y., Greenfield, T. K., & Lown, E. A. (2017). Health risk factors associated with lifetime abstinence from alcohol in the 1979 National Longitudinal Survey of Youth Cohort. *Alcoholism: Clinical and Experimental Research*, *41*(2), 388–398.

Kerr, W. C., Ye, Y., Williams, E., Lui, C. K., Greenfield, T. K., & Lown, E. A. (2019). Lifetime Alcohol Use Patterns and Risk of Diabetes Onset in the National Alcohol Survey. *Alcoholism: Clinical and Experimental Research*, *43*(2), 262–269.

Keyes, K. M., Allel, K., & Staudinger, U. M. (2019). Alcohol consumption predicts incidence of depressive episodes across 10 years among older adults in 19 countries. *International Review of Neurobiology*, *148*, 1–38. https://doi.org/10.1016/bs.irn.2019.09.001

Kezer, C. A., Simonetto, D. A., & Shah, V. H. (2021). Sex differences in alcohol consumption and alcohol-associated liver disease. *Mayo Clinic Proceedings*, *96*(4), 1006–1016.

Khantzian, E. (1985). The self-medication hypothesis of addictive disorders: focus on heroin and cocaine dependence. *Am J Psychiatry*, *142*(11), 1259–1264.

Kilian, C., Manthey, J., Probst, C., Brunborg, G. S., Bye, E. K., Ekholm, O., Kraus, L., Moskalewicz, J., Sieroslawski, J., & Rehm, J. (2020). Why is per capita consumption underestimated in alcohol surveys? Results from 39 surveys in 23 European countries. *Alcohol and Alcoholism*, *55*(5), 554–563.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52. https://doi.org/10.5395/rde.2013.38.1.52

Klau, S., Hoffmann, S., Patel, C. J., Ioannidis, J. P. A., & Boulesteix, A. L. (2021). Examining the robustness of observational associations to model, measurement and sampling uncertainty with the vibration of effects framework. *International Journal of Epidemiology*, *50*(1), 266–278. https://doi.org/10.1093/ije/dyaa164

Kleinbaum, D. G., Morgenstern, H. A. L., & Kupper, L. L. (1981). Selection bias in epidemiologic studies. *American Journal of Epidemiology*, *113*(4), 452–463.

Knott, C., Bell, S., & Britton, A. (2015). Alcohol consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of more than 1.9 million individuals from 38 observational studies. *Diabetes Care*, *38*(9), 1804–1812.

Krenz, M., & Korthuis, R. J. (2012). Moderate ethanol ingestion and cardiovascular protection: from epidemiologic associations to cellular mechanisms. *Journal of Molecular and Cellular Cardiology*, *52*(1), 93–104.

Lankester, J., Zanetti, D., Ingelsson, E., & Assimes, T. L. (2021). Alcohol use and cardiometabolic risk in the UK Biobank: A Mendelian randomization study. *PLoS One*, *16*(8), e0255801.

Larsson, S. C., Wallin, A., Wolk, A., & Markus, H. S. (2016). Differing association of alcohol consumption with different stroke types : a systematic review and meta-analysis. *BMC Medicine*, *14*(1). https://doi.org/10.1186/s12916-016-0721-4

Levine, S. Z. (2013). Evaluating the seven-item Center for Epidemiologic Studies Depression Scale short-form : a longitudinal US community study. *Social Psychiatry and Psychiatric Epidemiology*, *48*(9), 1519–1526. https://doi.org/10.1007/s00127-012-0650-2

Li, J., Wang, H., Li, M., Shen, Q., Li, X., Zhang, Y., Peng, J., Rong, X., & Peng, Y. (2020). Effect of alcohol use disorders and alcohol intake on the risk of subsequent depressive symptoms: a systematic review and meta-analysis of cohort studies. *Addiction*, *115*(7), 1224–1243. https://doi.org/10.1111/add.14935

Liang, L., Hua, R., Tang, S., Li, C., & Xie, W. (2021). Low-to-Moderate Alcohol Intake Associated with Lower Risk of Incidental Depressive Symptoms: A Pooled Analysis of Three Intercontinental Cohort Studies. *Journal of Affective Disorders*, *286*, 49–57. https://doi.org/10.1016/j.jad.2021.02.050

Lipsitch, M., Tchetgen, E. T., & Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.)*, *21*(3), 383.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., & Tian, C. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237–244.

Liu, Q., He, H., Yang, J., Feng, X., Zhao, F., & Lyu, J. (2020). Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study. *Journal of Psychiatric Research*, *126*, 134–140. https://doi.org/10.1016/j.jpsychires.2019.08.002

Livingston, M., Callinan, S., Raninen, J., Pennay, A., & Dietze, P. M. (2018). Alcohol consumption trends in Australia: comparing surveys and sales-based measures. *Drug and Alcohol Review*, *37*, S9–S14.

Ma, H., Wang, X., Li, X., Heianza, Y., & Qi, L. (2022). Moderate alcohol drinking with meals is related to lower incidence of type 2 diabetes. *The American Journal of Clinical Nutrition*, *116*(6), 1507–1514.

Mamluk, L., Edwards, H. B., Savovi, J., Leach, V., Jones, T., Moore, T. H. M., Ijaz, S., Lewis, S. J., Donovan, J. L., Lawlor, D., Davey Smith, G., Fraser, A., & Zuccolo, L. (2017). Low alcohol consumption and pregnancy and childhood outcomes: time to change guidelines indicating apparently ‘ safe ’ levels of alcohol during pregnancy ? A systematic review and meta-analyses. *BMJ Open*, *7*, e015410. https://doi.org/10.1136/bmjopen-2016-015410

Mamluk, L., Jones, T., Ijaz, S., Edwards, H. B., Savovi, J., Leach, V., Moore, T. H. M., Hinke, S. Von, Lewis, S. J., Donovan, J. L., Lawlor, D. A., Davey Smith, G., Fraser, A., & Zuccolo, L. (2020). Evidence of detrimental effects of prenatal alcohol exposure on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies. *International Journal of Epidemiology*, 1–24. https://doi.org/10.1093/ije/dyz272

Mansournia, M. A., Etminan, M., Danaei, G., Kaufman, J. S., & Collins, G. (2017). Handling time varying confounding in observational research. *BMJ* , *j4587*(October), 359. https://doi.org/10.1136/bmj.j4587

Manthey, J., Shield, K., & Rehm, J. (2022). Alcohol and health. *The Lancet*, *400*(10365), 1764–1765.

Mason, A. M. (2021). *UKBB outcomes*. https://github.com/amymariemason/UKBB-outcomes

Mason, A. M., & Burgess, S. (2022). *Software Application Profile: SUMnlmr, an R package that facilitates flexible and reproducible non-linear Mendelian randomization analyses*. Oxford University Press.

Masur, P. K., & Scharkow, M. (2020). *specr: Conducting and Visualizing Specification Curve Analyses*.

Mcgue, M., Osler, M., & Christensen, K. (2010). Causal Inference and Observational Research: The Utility of Twins. In *Source: Perspectives on Psychological Science* (Vol. 5, Issue 5).

McIntosh, C., & Chick, J. (2004). Alcohol and the nervous system. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(suppl 3), iii16–iii21.

McQuire, C., & de Vocht, F. (2019). Methodological advances to mitigate some of the challenges of research on alcohol and all-cause mortality: Commentary on Rehm. *Drug and Alcohol Review*, *38*(1), 7–8. https://doi.org/10.1111/dar.12898

Melberg, H. O. (2006). Does moderate alcohol intake reduce mortality? In O. G. Jon Elster & A. H. and K. Moene (Eds.), *Understanding Choice, Explaining Behaviour: Essays in honour of Ole-Jorgen Kkog* (pp. 191–210). Oslo Academic Press.

Menezes, A. M. B., Oliveira, P. D., Wehrmeister, F. C., Assunção, M. C. F., Oliveira, I. O., Tovo-Rodrigues, L., Ferreira, G. D., & Gonçalves, H. (2019). Association of modifiable risk factors and IL-6, CRP, and adiponectin: Findings from the 1993 birth cohort, Southern Brazil. *PLoS ONE*, *14*(5). https://doi.org/10.1371/journal.pone.0216202

Mezue, K., Osborne, M. T., Abohashem, S., Zureigat, H., Gharios, C., Grewal, S. S., Radfar, A., Cardeiro, A., Abbasi, T., Choi, K. W., Fayad, Z. A., Smoller, J. W., Rosovsky, R., Shin, L., Pitman, R., & Tawakol, A. (2023). Reduced Stress-Related Neural Network Activity Mediates the Effect of Alcohol on Cardiovascular Risk. *Journal of the American College of Cardiology*, *81*(24), 2315–2325. https://doi.org/10.1016/j.jacc.2023.04.015

Milaneschi, Y., Kappelmann, N., Ye, Z., Lamers, F., Moser, S., Jones, P. B., Burgess, S., Penninx, B. W. J. H., & Khandaker, G. M. (2021). Association of inflammation with depression and anxiety: evidence for symptom-specificity and potential causality from UK Biobank and NESDA cohorts. *Molecular Psychiatry*, *26*(12), 7393–7402. https://doi.org/10.1038/s41380-021-01188-w

Millwood, I. Y., Walters, R. G., Mei, X. W., Guo, Y., Yang, L., Bian, Z., Bennett, D. A., Chen, Y., Dong, C., & Hu, R. (2019). Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. *The Lancet*, *393*(10183), 1831–1842.

Minelli, C., Del Greco M, F., van der Plaat, D. A., Bowden, J., Sheehan, N. A., & Thompson, J. (2021). The use of two-sample methods for Mendelian randomization analyses on single large datasets. *International Journal of Epidemiology*, *50*(5), 1651–1659.

Minzer, S., Losno, R. A., & Casas, R. (2020). The Effect of Alcohol on Cardiovascular Risk Factors: Is There New Information ? *Nutrients*, *12*(4), 1–22.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., ... Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moodie, E. E., & Stephens, D. A. (2022, April 5). *Causal inference: critical developments, past and future*. ArXiv. https://doi.org/https://doi.org/10.48550/arXiv.2204.02231

Mukamal, K. J. (2010). A 42-year-old man considering whether to drink alcohol for his health. *JAMA*, *303*(20), 2065–2073.

Mukamal, K. J., Stampfer, M. J., & Rimm, E. B. (2020). Genetic instrumental variable analysis : time to call mendelian randomization what it is . The example of alcohol and cardiovascular disease. *European Journal of Epidemiology*, *35*(2), 93–97.

Munafò, M. R., & Davey Smith, G. (2018). Repeating experiments is not enough. *Nature*, *553*(7689), 399–401.

Muralidharan, S., Ambade, A., Fulham, M. A., Deshpande, J., Catalano, D., & Mandrekar, P. (2014). Moderate Alcohol Induces Stress Proteins HSF1 and hsp70 and Inhibits Proinflammatory Cytokines Resulting in Endotoxin Tolerance. *The Journal of Immunology*, *193*(4), 1975–1987. https://doi.org/10.4049/jimmunol.1303468

Naimi, A. I., Cole, S. R., & Kennedy, E. H. (2017). An introduction to g methods. *International Journal of Epidemiology*, *46*(2), 756–762. https://doi.org/10.1093/ije/dyw323

Naimi, A. I., Moodie, E. E. M., Auger, N., & Kaufman, J. S. (2014). Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology*, 292–299.

Naimi, T., Chikritzhs, T., & Stockwell, T. (2022). Commentary on Di Castelnuovo et al: Implications of using low volume drinkers instead of never drinkers as the reference group. *Addiction*, *117*(2), 327–329.

Naimi, T. S., Brown, D. W., Brewer, R. D., Giles, W. H., Mensah, G., Serdula, M. K., Mokdad, A. H., Hungerford, D. W., Lando, J., Naimi, S., & Stroup, D. F. (2005). Cardiovascular risk factors and confounders among nondrinking and moderate-drinking U.S. adults. *American Journal of Preventive Medicine*, *28*(4), 369–373. https://doi.org/10.1016/j.amepre.2005.01.011

Naimi, T. S., Stockwell, T., Zhao, J., Xuan, Z., Dangardt, F., Saitz, R., Liang, W., & Chikritzhs, T. (2017). Selection biases in observational studies affect associations

between 'moderate' alcohol consumption and mortality. *Addiction*, *112*(2), 207–214. https://doi.org/10.1111/add.13451

National Health and Medical Research Council. (2020). Australian Guidelines to Reduce Health Risks from Drinking Alcohol. In *Australian guidelines to reduce health risks from drinking alcohol*. http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/ds10-alcohol.pdf

Ng Fat, L. (2014). *Investigating the health of non-drinkers: the sick-quitter and sick non-starter hypotheses*. UCL (University College London).

Ng Fat, L., Cable, N., & Shelton, N. (2015). Worsening of health and a cessation or reduction in alcohol consumption to special occasion drinking across three decades of the life course. *Alcoholism: Clinical and Experimental Research*, *39*(1), 166–174.

NHMRC Clinical Trials Centre at the University of Sydney. (2019). *Evaluating the evidence on the health effects of alcohol consumption: Evidence evaluation report*.

Nicholls, J. (2012). Time for reform? Alcohol policy and cultural change in England since 2000. *British Politics*, *7*(3), 250–271.

Nutt, D. (2020). *Drink?: the new science of alcohol and your health*. Hachette UK.

Oh, B. K., Lee, S. J., Kim, H., Choi, H. I., Lee, J. Y., Lee, S. H., Kim, B. J., Kim, B. S., Kang, J. H., Lee, M. Y., & Sung, K. C. (2021). Relationship between alcohol consumption and insulin resistance measured using the homeostatic model assessment for insulin resistance: A retrospective cohort study of 280,194 people. *Nutrition, Metabolism and Cardiovascular Diseases*, *31*(10), 2842–2850. https://doi.org/10.1016/j.numecd.2021.06.023

Oppenheimer, G. M., & Bayer, R. (2020). Is moderate drinking protective against heart disease? The science, politics and history of a public health conundrum. *The Milbank Quarterly*, *98*(1), 39.

Owens, M. M., Potter, A., Hyatt, C. S., Albaugh, M., Thompson, W. K., Jernigan, T., Yuan, D., Hahn, S., Allgaier, N., & Garavan, H. (2021). Recalibrating expectations about effect size: A multi-method survey of effect sizes in the ABCD study. *PLoS ONE*, *16*(9), e0257535. https://doi.org/10.1371/journal.pone.0257535

Pacek, L. R., Martins, S. S., & Crum, R. M. (2013). The bidirectional relationships between alcohol, cannabis, co-occurring alcohol and cannabis use disorders with major depressive disorder: Results from a national sample. *Journal of Affective Disorders*, *148*(2–3), 188–195. https://doi.org/10.1016/j.jad.2012.11.059

Pai, J. K., Hankinson, S. E., Thadhani, R., Rifai, N., Pischon, T., & Rimm, E. B. (2006). Moderate alcohol consumption and lower levels of inflammatory markers in US men and

women. *Atherosclerosis*, *186*(1), 113–120. https://doi.org/10.1016/j.atherosclerosis.2005.06.037

Paradis, C., Butt, P. , Shield, K. , Poole, N. , Wells, S. , Naimi, T. , & Sherk, A. ,. (2023). *Canada's Guidance on Alcohol and Health: Final Report*.

Patel, C. J., Burford, B., & Ioannidis, J. P. A. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology*, *68*(9), 1046–1058. https://doi.org/10.1016/j.jclinepi.2015.05.029

Patel, V., Chisholm, D., Parikh, R., Charlson, F. J., Degenhardt, L., Dua, T., Ferrari, A. J., Hyman, S., Laxminarayan, R., Levin, C., Lund, C., Medina Mora, M. E., Petersen, I., Scott, J., Shidhaye, R., Vijayakumar, L., Thornicroft, G., & Whiteford, H. (2016). Addressing the burden of mental, neurological, and substance use disorders: Key messages from Disease Control Priorities, 3rd edition. *The Lancet*, *387*(10028), 1672–1685. https://doi.org/10.1016/S0140-6736(15)00390-6

Paulson, D., Shah, M., Herring, D., Scott, R., Herrera, M., Brush, D., & Bassett, R. (2018). The relationship between moderate alcohol consumption, depressive symptomatology, and C-reactive protein: the Health and Retirement Study. *International Journal of Geriatric Psychiatry*, *33*(2), 316–324. https://doi.org/10.1002/gps.4746

Pearce, N., & Lawlor, D. A. (2017). Causal inference — so much more than statistics. *International Journal of Epidemiology*, *45*(6), 1895–1903. https://doi.org/10.1093/ije/dyw328

Pearl, R. (1926). *Alcohol and longevity*. AA Knopf.

Peele, S., & Brodsky, A. (2000). Exploring psychological benefits associated with moderate alcohol use: A necessary corrective to assessments of drinking outcomes? *Drug and Alcohol Dependence*, *60*(3), 221–247. https://doi.org/10.1016/S0376-8716(00)00112-5

Peng, M., Zhang, J., Zeng, T., Hu, X., Min, J., Tian, S., Wang, Y., Liu, G., Wan, L., & Huang, Q. (2019). Alcohol consumption and diabetes risk in a Chinese population: a Mendelian randomization analysis. *Addiction*, *114*(3), 436–449.

Pietikäinen, S., Silventoinen, K., Svedberg, P., Alexanderson, K., Huunan-Seppala, A., Koskenuvo, K., Koskenuvo, M., Kaprio, J., & Ropponen, A. (2011). Health-Related and Sociodemographic Risk Factors for Disability Pension due to Low Back Disorders: a 30-year prospective Finnish Twin Cohort Study. *Journal of Occupational and Environmental Medicine*, *53*(5), 488–496. https://doi.org/10.1097/JOM.0b013e31821576dd

Radloff, L. S. (1977). The CES-D Scale : A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401.

Rehm, J. (2019). Why the relationship between level of alcohol-use and all-cause mortality cannot be addressed with meta-analyses of cohort studies. *Drug and Alcohol Review*, *38*(1), 3–4. https://doi.org/10.1111/dar.12866

Rehm, J., Gmel Sr, G. E., Gmel, G., Hasan, O. S. M., Imtiaz, S., Popova, S., Probst, C., Roerecke, M., Room, R., & Samokhvalov, A. V. (2017). The relationship between different dimensions of alcohol use and the burden of disease—an update. *Addiction*, *112*(6), 968–1001.

Rehm, J., Irving, H., Ye, Y., Kerr, W. C., Bond, J., & Greenfield, T. K. (2008). Are lifetime abstainers the best control group in alcohol epidemiology? On the stability and validity of reported lifetime abstention. *American Journal of Epidemiology*, *168*(8), 866–871.

Rehm, J., Probst, C., Shield, K. D., & Shuper, P. A. (2017). Does alcohol use have a causal effect on HIV incidence and disease progression ? A review of the literature and a modeling strategy for quantifying the effect. *Population Health Metrics*, *15*(1), 1–7. https://doi.org/10.1186/s12963-017-0121-9

Rehm, J., Soerjomataram, I., Ferreira-Borges, C., & Shield, K. D. (2019). Does alcohol use affect cancer risk? *Current Nutrition Reports*, *8*(3), 222–229.

Réus, G. Z., Fries, G. R., Stertz, L., Badawy, M., Passos, I. C., Barichello, T., Kapczinski, F., & Quevedo, J. (2015). The role of inflammation and microglial activation in the pathophysiology of psychiatric disorders. In *Neuroscience* (Vol. 300, pp. 141–154). Elsevier Ltd. https://doi.org/10.1016/j.neuroscience.2015.05.018

Richiardi, L., Bellocco, R., & Zugna, D. (2013). Mediation analysis in epidemiology: Methods, interpretation and bias. *International Journal of Epidemiology*, *42*(5), 1511–1519. https://doi.org/10.1093/ije/dyt127

Richmond, R. C., Al-amin, A., Davey Smith, G., & Relton, C. L. (2014). Approaches for drawing causal inferences from epidemiological birth cohorts: A review. *Early Human Development*, *90*, 769–780. https://doi.org/10.1016/j.earlhumdev.2014.08.023

Ridgeway, G., Kovalchik, S. A., Griffin, B. A., & Kabeto, M. U. (2015). Propensity score analysis with survey weighted data. *Journal of Causal Inference*, *3*(2), 237–249.

Ridker, P. M. (2004). Inflammation in atherothrombosis: how to use high-sensitivity C-reactive protein (hsCRP) in clinical practice. *The American Heart Hospital Journal*, *2*(4 Suppl 1), 4–9. http://europepmc.org/abstract/MED/15539969

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Roerecke, M., & Rehm, J. (2014). Alcohol consumption, drinking patterns, and ischemic heart disease: a narrative review of meta-analyses and a systematic review and meta-

analysis of the impact of heavy drinking occasions on risk for moderate drinkers. *BMC Medicine*, *12*(1), 182.

Roerecke, M., Tobe, S. W., Kaczorowski, J., Bacon, S. L., Vafaei, A., Hasan, O. S. M., Krishnan, R. J., Raifu, A. O., & Rehm, J. (2018). Sex-specific associations between alcohol consumption and incidence of hypertension: a systematic review and meta-analysis of cohort studies. *Journal of the American Heart Association*, *7*(13), e008202.

Rohleder, N. (2019). Stress and inflammation – The need to address the gap in the transition between acute and chronic stress effects. *Psychoneuroendocrinology, 105*, 164–171. https://doi.org/10.1016/j.psyneuen.2019.02.021

Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation : Graphical Causal Models for Observational Data. *Advances in methods and practices in psychological science*, *1*(1), 27-42. https://doi.org/10.1177/2515245917745629

Ropponen, A., Silventoinen, K., Svedberg, P., Alexanderson, K., Koskenvuo, K., Huunan-Seppälä, A., Koskenvuo, M., & Kaprio, J. (2011). Health-related risk factors for disability pensions due to musculoskeletal diagnoses : A 30-year Finnish twin cohort study. *Scandinavian Journal of Public Health*, *39*(8), 839–848. https://doi.org/10.1177/1403494811418283

Ropponen, A., & Svedberg, P. (2013). Single and additive effects of health behaviours on the risk for disability pensions among Swedish twins. *European Journal of Public Health*, *24*(4), 643–648. https://doi.org/10.1093/eurpub/ckt168

Rothstein, D. S., Carr, D., & Cooksey, E. (2019). Cohort Profile: The National Longitudinal Survey of Youth 1979 (NLSY79). *International Journal of Epidemiology*, *48*(1), 22-22E. https://doi.org/10.1093/ije/dyy133

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688.

Rubin, E. (2014). To drink or not to drink: that is the question. *Alcoholism, Clinical and Experimental Research*, *38*(12), 2889.

Samuelsson, Å., Ropponen, A., Alexanderson, K., & Svedberg, P. (2013). A prospective cohort study of disability pension due to mental diagnoses : the importance of health factors and behaviors. *BMC Public Health*, *13*(1), 1. https://doi.org/10.1186/1471-2458-13-621

Sander, P. M., Cole, S. R., Stall, R. D., Jacobson, L. P., Eron, J. J., Napravnik, S., Gaynes, B. N., Johnson-hill, L. M., Bolan, R. K., & Ostrow, D. G. (2013). Joint effects of alcohol consumption and high-risk sexual behavior on HIV seroconversion among men who have sex with men. *AIDS*, *27*(5), 815–823. https://doi.org/10.1097/QAD.0b013e32835cff4b

Sarma, M. A. (2021). *Package 'multiverse.'*

Sayon-Orea, C., Martinez-gonzalez, M. A., & Bes-rastrollo, M. (2011). Alcohol consumption and body weight : a systematic review. *Nutrition Reviews*, *69*(8), 419–431. https://doi.org/10.1111/j.1753-4887.2011.00403.x

Schrieks, I. C., Heil, A. L. J., Hendriks, H. F. J., Mukamal, K. J., & Beulens, J. W. J. (2015). The Effect of alcohol consumption on insulin sensitivity and glycemic status: A systematic review and meta-analysis of intervention studies. *Diabetes Care*, *38*(4), 723–732. https://doi.org/10.2337/dc14-1556

Schuler, M. S., & Rose, S. (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *American Journal of Epidemiology*, *185*(1), 65–73. https://doi.org/10.1093/aje/kww165

Schwarzinger, M., Pollock, B. G., Hasan, O. S. M., Dufouil, C., Rehm, J., Baillot, S., Guibert, Q., Planchet, F., & Luchini, S. (2018). Contribution of alcohol use disorders to the burden of dementia in France 2008–13: a nationwide retrospective cohort study. *The Lancet Public Health*, *3*(3), e124–e132. https://doi.org/10.1016/S2468-2667(18)30022-7

Sciensano. (2022, February 17). *Alcohol use*.

Seligman, F., & Nemeroff, C. B. (2015). The interface of depression and cardiovascular disease: therapeutic implications. *Annals of the New York Academy of Sciences*, *1345*(1), 25–35.

Shaper, A. G., Wannamethee, G., & Walker, M. (1988). Alcohol and mortality in British men: explaining the U-shaped curve. *The Lancet*, *332*(8623), 1267–1273. https://doi.org/https://doi.org/10.1016/S0140-6736(88)92890-5

Sherk, A., Gilmore, W., Churchill, S., Lensvelt, E., Stockwell, T., & Chikritzhs, T. (2019). Implications of cardioprotective assumptions for national drinking guidelines and alcohol harm monitoring systems. *International Journal of Environmental Research and Public Health*, *16*(24), 1–17. https://doi.org/10.3390/ijerph16244956

Shuper, P. A., Neuman, M., Kanteres, F., Baliunas, D., Joharchi, N., & Rehm, J. (2010). Causal Considerations on Alcohol and HIV / AIDS — A Systematic Review. *Alcohol and Alcoholism*, *45*(2), 159–166. https://doi.org/10.1093/alcalc/agp091

Silverwood, R. J., Holmes, M. V., Dale, C. E., Lawlor, D. A., Whittaker, J. C., Davey Smith, G., Leon, D. A., Palmer, T., Keating, B. J., Zuccolo, L., Casas, J. P., & Dudbridge, F. (2014). Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: Application to alcohol and cardiovascular traits. *International Journal of Epidemiology*, *43*(6), 1781–1790. https://doi.org/10.1093/ije/dyu187

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Simpson, R. F., Hermon, C., Liu, B., Green, J., Reeves, G. K., Beral, V., & Floud, S. (2019). Alcohol drinking patterns and liver cirrhosis risk: analysis of the prospective UK Million Women Study. *The Lancet Public Health*, *4*(1), e41–e48.

Sipilä, P., Rose, R. J., & Kaprio, J. (2016). Drinking and mortality : long-term follow-up of drinking- discordant twin pairs. *Addiction*, *111*(2), 245–254. https://doi.org/10.1111/add.13152

Sjolander, A., & Martinussen, T. (2019). Instrumental variable estimation with the R package ivtools. *Epidemiologic Methods*, *8*(1).

Skrivankova, V. W., Richmond, R. C., Woolf, B. A. R., Yarmolinsky, J., Davies, N. M., Swanson, S. A., VanderWeele, T. J., Higgins, J. P. T., Timpson, N. J., & Dimou, N. (2021). Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *Jama*, *326*(16), 1614–1621.

Sohi, I., Franklin, A., Chrystoja, B., Wettlaufer, A., Rehm, J., & Shield, K. (2021). The global impact of alcohol consumption on premature mortality and health in 2016. *Nutrients*, *13*(9), 3145.

Spanish Ministry of Health. (2020). *Low risk alcohol consumption thresholds: Update on the risks related to alcohol consumption levels, consumption patterns and type of alcoholic beverages.*

Spiegelman, D., Lovato, L. C., Khudyakov, P., Wilkens, T. L., Adebamowo, C. A., Adebamowo, S. N., Appel, L. J., Beulens, J. W. J., Coughlin, J. W., & Dragsted, L. O. (2020). The Moderate Alcohol and Cardiovascular Health Trial (MACH15): Design and methods for a randomized trial of moderate alcohol consumption and cardiometabolic risk. *European Journal of Preventive Cardiology*, *27*(18), 1967–1982.

Staley, J. R., & Burgess, S. (2017). Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genetic Epidemiology*, *41*(4), 341–352. https://doi.org/10.1002/gepi.22041

State Agency for the Prevention of Alcohol-Related Problems (PARPA). (2019). *Alcohol - is it safe to drink?*

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Steffensen, S. C., Walton, C. H., Hansen, D. M., Yorgason, J. T., Gallegos, R. A., & Criado, J. R. (2009). Contingent and non-contingent effects of low-dose ethanol on GABA neuron activity in the ventral tegmental area. *Pharmacology Biochemistry and Behavior*, *92*(1), 68–75. https://doi.org/10.1016/j.pbb.2008.10.012

Stockwell, T., & Room, R. (2012). Constructing and responding to low-risk drinking guidelines: Conceptualisation, evidence and reception. *Drug and Alcohol Review*, *31*(2), 121–125. https://doi.org/10.1111/j.1465-3362.2011.00416.x

Stockwell, T., Zhao, J., Panwar, S., Roemer, A., Naimi, T., & Chikritzhs, T. (2016). Do "moderate" drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. *Journal of Studies on Alcohol and Drugs*, *77*(2), 185–198.

Stockwell, T., Zhao, J., Sherk, A., Rehm, J., Shield, K., & Naimi, T. (2018). Underestimation of alcohol consumption in cohort studies and implications for alcohol's contribution to the global burden of disease. *Addiction*, *113*(12), 2245–2249.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., & Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, *12*(3), e1001779.

Sun, Y.-Q., Burgess, S., Staley, J. R., Wood, A. M., Bell, S., Kaptoge, S. K., Guo, Q., Bolton, T. R., Mason, A. M., & Butterworth, A. S. (2019). Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. *BMJ*, *l1042*, 364.

Szabo, G. (2007). Moderate Drinking, Inflammation, and Liver Disease. *Annals of Epidemiology*, *17*(5 SUPPL.). https://doi.org/10.1016/j.annepidem.2007.01.012

Taghdiri, S. (2017). *Association of Alcohol Consumption with Glucose Homeostasis: A Systematic Review and Meta-Analysis.* [Master's thesis, University of Toronto.] TSpace. https://tspace.library.utoronto.ca/bitstream/1807/79786/3/Taghdiri_Soudeh_201706_MSc_thesis.pdf.

Textor, J., Zander, B. Van Der, & Gilthorpe, M. S. (2017). Robust causal inference using directed acyclic graphs : the R package ' dagitty .' *International Journal of Epidemiology*, *45*(6), 1887–1894. https://doi.org/10.1093/ije/dyw341

Thoemmes, F., & Ong, A. D. (2016). A Primer on Inverse Probability of Treatment Weighting and Marginal Structural Models. *Emerging Adulthood*, *4*(1), 40–59. https://doi.org/10.1177/2167696815621645

Tian, H., Mason, A. M., Liu, C., & Burgess, S. (2022). *Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method*. BioRxiv. https://doi.org/https://doi.org/10.1101/2022.06.28.497930

Tinajero, M. G., & Malik, V. S. (2021). An Update on the Epidemiology of Type 2 Diabetes: A Global Perspective. In *Endocrinology and Metabolism Clinics of North America* (Vol. 50, Issue 3, pp. 337–355). W.B. Saunders. https://doi.org/10.1016/j.ecl.2021.05.013

Tizabi, Y., Getachew, B., Ferguson, C. L., Csoka, A. B., Thompson, K. M., Gomez-Paz, A., Ruda-Kucerova, J., & Taylor, R. E. (2018). Low Vs. High Alcohol: Central Benefits Vs. Detriments. *Neurotoxicity Research*, *34*(4), 860–869. https://doi.org/10.1007/s12640-017-9859-x

Topiwala, A., Taschler, B., Ebmeier, K. P., Smith, S., Zhou, H., Levey, D. F., Codd, V., Samani, N. J., Gelernter, J., Nichols, T. E., & Burgess, S. (2022). Alcohol consumption and telomere length: Mendelian randomization clarifies alcohol's effects. *Molecular Psychiatry*, *27*(10), 4001–4008. https://doi.org/10.1038/s41380-022-01690-9

Turner, S., Mota, N., Bolton, J., & Sareen, J. (2018). Self-medication with alcohol or drugs for mood and anxiety disorders: A narrative review of the epidemiological literature. *Depression and Anxiety*, *35*(9), 851–860. https://doi.org/10.1002/da.22771

US Department of Agriculture and US Department of Health and Human Services. (2020). *Dietary Guidelines for Americans, 2020-2025. 9th Edition.* https://doi.org/10.1093/ajcn/34.1.121

van de Luitgaarden, I. A. T., van Oort, S., Bouman, E. J., Schoonmade, L. J., Schrieks, I. C., Grobbee, D. E., van der Schouw, Y. T., Larsson, S. C., Burgess, S., & van Ballegooijen, A. J. (2021). Alcohol consumption in relation to cardiovascular diseases and mortality: a systematic review of Mendelian randomization studies. *European Journal of Epidemiology*, *37*, 655–669.

van der Heide, F. C. T., Eussen, S. J. P. M., Houben, A. J. H. M., Henry, R. M. A., Kroon, A. A., van der Kallen, C. J. H., Dagnelie, P. C., van Dongen, M. C. J. M., Berendschot, T. T. J. M., & Schouten, J. S. A. G. (2023). Alcohol consumption and microvascular dysfunction: a J-shaped association: The Maastricht Study. *Cardiovascular Diabetology*, *22*(1), 67.

Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2016). Time-Varying Treatments in Observational Studies: Marginal Structural Models of the

Effects of Early Grade Retention on Math Achievement. *Multivariate Behavioral Research*, *51*(6), 843–864. https://doi.org/10.1080/00273171.2016.1155146

VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner ' s Guide. *Annual Review of Public Health*, *37*, 17–32. https://doi.org/10.1146/annurev-publhealth-032315-021402

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine*, *167*(4), 268–274.

van't Klooster, C. C., van der Graaf, Y., Ridker, P. M., Westerink, J., Hjortnaes, J., Sluijs, I., Asselbergs, F. W., Bots, M. L., Kappelle, L. J., & Visseren, F. L. J. (2020). The relation between healthy lifestyle changes and decrease in systemic inflammation in patients with stable cardiovascular disease. *Atherosclerosis*, *301*, 37–43. https://doi.org/10.1016/j.atherosclerosis.2020.03.022

Visontay, R., Mewton, L., Slade, T., Aris, I. M., & Sunderland, M. (2023). Moderate Alcohol Consumption and Depression: A Marginal Structural Model Approach Promoting Causal Inference. *American Journal of Psychiatry*, *180*(3), 209–217.

Visontay, R., Mewton, L., Sunderland, M., Bell, S., Britton, A., Osman, B., North, H., Mathew, N., & Slade, T. (2023). A comprehensive evaluation of the longitudinal association between alcohol consumption and a measure of inflammation: Multiverse and vibration of effects analyses. *Drug and Alcohol Dependence*, 109886.

Visontay, R., Rao, R. T., & Mewton, L. (2021). Alcohol use and dementia: new research directions. *Current Opinion in Psychiatry*, *34*(2), 165–170. https://doi.org/10.1097/yco.0000000000000679

Visontay, R., Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2021). Are there non-linear relationships between alcohol consumption and long-term health? Protocol for a systematic review of observational studies employing approaches to improve causal inference. *BMJ Open*, *11*(3), e043985.

Visontay, R., Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2022). Are there non-linear relationships between alcohol consumption and long-term health?: a systematic review of observational studies employing approaches to improve causal inference. *BMC Medical Research Methodology*, *22*(1), 1–28.

Volpato, S., Pahor, M., Ferrucci, L., Simonsick, E. M., Guralnik, J. M., Kritchevsky, S. B., Fellin, R., & Harris, T. B. (2004). Relationship of Alcohol Intake with Inflammatory Markers and Plasminogen Activator Inhibitior-1 in Well-Functioning Older Adults: The Health, Aging, and Body Composition Study. *Circulation*, *109*(5), 607–612. https://doi.org/10.1161/01.CIR.0000109503.13955.00

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & Initiative, S. (2007). The Strengthening the Reporting of Observational Studies in

Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Annals of Internal Medicine*, *147*(8), 573–577.

Vu, K. N., Ballantyne, C. M., Hoogeveen, R. C., & Nambi, V. (2016). Causal Role of Alcohol Consumption in an Improved Lipid Profile : The Atherosclerosis Risk in Communities (ARIC) Study. *PLoS ONE*, *11*(2), e0148765. https://doi.org/10.1371/journal.pone.0148765

Wallach, J. D., Serghiou, S., Chu, L., Egilman, A. C., Vasiliou, V., Ross, J. S., & Ioannidis, J. P. A. (2020). Evaluation of confounding in epidemiologic studies assessing alcohol consumption on the risk of ischemic heart disease. *BMC Medical Research Methodology*, *20*(1), 1–10. https://doi.org/10.1186/s12874-020-0914-6

Watkins, T. R. (2019). *Understanding uncertainty and bias to improve causal inference in health intervention research*. [Doctoral dissertation, The University of Sydney]. Sydney Digital Theses (Open Access). https://ses.library.usyd.edu.au/handle/2123/20772.

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (n.d.). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Williamson, T., & Ravani, P. (2017). Marginal structural models in clinical research: When and how to use them? *Nephrology Dialysis Transplantation*, *32*(suppl_2), ii84–ii90. https://doi.org/10.1093/ndt/gfw341

Wilson, D. F., & Matschinsky, F. M. (2020). Ethanol metabolism: The good, the bad, and the ugly. *Medical Hypotheses*, *140*. https://doi.org/10.1016/j.mehy.2020.109638

Wood, A. M., Kaptoge, S., Butterworth, A. S., Willeit, P., Warnakula, S., Bolton, T., Paige, E., Paul, D. S., Sweeting, M., & Burgess, S. (2018). Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *The Lancet*, *391*(10129), 1513–1523.

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

Xue, A., Jiang, L., Zhu, Z., Wray, N. R., Visscher, P. M., Zeng, J., & Yang, J. (2021). Genome-wide analyses of behavioural traits are subject to bias by misreports and longitudinal changes. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20237-6

Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K. E., Zheng, Z., Yengo, L., Lloyd-Jones, L. R., Sidorenko, J., & Wu, Y. (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications*, *9*(1), 1–14.

Yang, Y., Liu, D.-C., Wang, Q.-M., Long, Q.-Q., Zhao, S., Zhang, Z., Ma, Y., Wang, Z.-M., Chen, L.-L., & Wang, L.-S. (2016). Alcohol consumption and risk of coronary artery disease : A dose-response meta-analysis of prospective studies. *Nutrition*, *32*(6), 637–644. https://doi.org/10.1016/j.nut.2015.11.013

Yavorska, O. O., & Burgess, S. (2017). MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *International Journal of Epidemiology*, *46*(6), 1734–1739.

Yuan, S., Merino, J., & Larsson, S. C. (2023). Causal factors underlying diabetes risk informed by Mendelian randomisation analysis: evidence, opportunities and challenges. *Diabetologia*, *66*(5), 800–812.

Zhao, J., Stockwell, T., Roemer, A., & Chikritzhs, T. (2016). Is alcohol consumption a risk factor for prostate cancer ? A systematic review and meta – analysis. *BMC Cancer*, *16*(1), 1–13. https://doi.org/10.1186/s12885-016-2891-z

Zhao, Q., Wang, J., Hemani, G., Bowden, J., & Small, D. S. (2020). *Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score*. ArXiv. https://doi.org/https://doi.org/10.48550/arXiv.1801.09652

Zhong, Q. Y., Gelaye, B., Weele, T. J. V., Sanchez, S. E., & Williams, M. A. (2018). Causal model of the association of social support with antepartum depression: Amarginal structural modeling approach. *American Journal of Epidemiology*, *187*(9), 1871–1879. https://doi.org/10.1093/aje/kwy067

Zhou, A., & Hyppönen, E. (2022). Vitamin D deficiency and C-reactive protein: a bidirectional Mendelian randomization study. *International Journal of Epidemiology*, *52*(1), 260–271.

Zhou, A., Selvanayagam, J. B., & Hyppönen, E. (2022). Non-linear Mendelian randomization analyses support a role for vitamin D deficiency in cardiovascular disease risk. *European Heart Journal*, *43*(18), 1731–1739.

Zhou, T., Sun, D., Li, X., Ma, H., Heianza, Y., & Qi, L. (2021). Educational attainment and drinking behaviors: Mendelian randomization study in UK Biobank. *Molecular Psychiatry*, *26*(8), 4355–4366.

Appendix A

Causal evidence for non-linear alcohol–long-term health relationships

Preface

This systematic review was published as **Visontay, R.,** Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2022). Are there non-linear relationships between alcohol consumption and long-term health?: a systematic review of observational studies employing approaches to improve causal inference. *BMC Medical Research Methodology, 22*(1), 1-28.

RV conceptualised the study with assistance from MS, TS, and LM. RV and JW completed the screening, data extraction, and quality assessment process. RV conducted the narrative synthesis and drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Visontay *et al.*
BMC Medical Research Methodology (2022) 22:16
https://doi.org/10.1186/s12874-021-01486-5

BMC Medical Research Methodology

RESEARCH

Open Access

Are there non-linear relationships between alcohol consumption and long-term health?: a systematic review of observational studies employing approaches to improve causal inference

Rachel Visontay[1,2*], Matthew Sunderland[1], Tim Slade[1], Jack Wilson[1] and Louise Mewton[2,1]

Abstract

Background: Research has long found 'J-shaped' relationships between alcohol consumption and certain health outcomes, indicating a protective effect of moderate consumption. However, methodological limitations in most studies hinder causal inference. This review aimed to identify all observational studies employing improved approaches to mitigate confounding in characterizing alcohol–long-term health relationships, and to qualitatively synthesize their findings.

Methods: Eligible studies met the above description, were longitudinal (with pre-defined exceptions), discretized alcohol consumption, and were conducted with human populations. MEDLINE, PsycINFO, Embase and SCOPUS were searched in May 2020, yielding 16 published manuscripts reporting on cancer, diabetes, dementia, mental health, cardiovascular health, mortality, HIV seroconversion, and musculoskeletal health. Risk of bias of cohort studies was evaluated using the Newcastle-Ottawa Scale, and a recently developed tool was used for Mendelian Randomization studies.

Results: A variety of functional forms were found, including reverse J/J-shaped relationships for prostate cancer and related mortality, dementia risk, mental health, and certain lipids. However, most outcomes were only evaluated by a single study, and few studies provided information on the role of alcohol consumption pattern.

Conclusions: More research employing enhanced causal inference methods is urgently required to accurately characterize alcohol–long-term health relationships. Those studies that have been conducted find a variety of linear and non-linear functional forms, with results tending to be discrepant even within specific health outcomes.

Trial registration: PROSPERO registration number CRD42020185861.

Keywords: Systematic review, Alcohol drinking, Alcohol abstinence, Causality, Risk factors, Protective factors

Background

While the contribution of heavy alcohol consumption to the burden of disease is well-known [1], findings that moderate alcohol consumption is associated with health benefits for a wide range of outcomes persist. This relationship is often characterized as 'J-shaped', where

*Correspondence: rachel.visontay@sydney.edu.au
[1] The Matilda Centre for Research in Mental Health and Substance Use, The University of Sydney, Level 6, Jane Foss Russell Building, G02, Sydney, NSW 2006, Australia
Full list of author information is available at the end of the article

low-to-moderate consumption coincides with the lowest risk, compared to a slightly higher risk for alcohol abstainers and much greater risk for heavy consumers. However, findings are inconsistent, and many functional forms have been reported for alcohol–long-term health relationships (see Fig. 1 for exemplar forms). Much recent research supports positive linear/monotonically increasing relationships for many outcomes, including most cancers [1, 2]. However, methodologically rigorous individual studies continue to find exceptions, e.g., for mortality [3], and indeed for dementia, diabetes, and particular cardiovascular conditions, evidence remains largely consistent with a J/U-shape [1, 2]. Which of the myriad reported functional forms reflect true causal relationships, and which are merely methodological artefacts, remains unclear.

In addition to these inconsistent findings, J-shaped relationships are sometimes found for certain outcomes (e.g., cirrhosis of the liver) which lack plausible biological mechanisms [4]. More generally, that studies across a broad array of health outcomes (with different underlying biological pathways) arrive at similar functional forms has prompted scrutiny of biases in observational studies – specifically confounding, reverse causality, selection bias and measurement error [5–7]. Confounding may constitute the biggest threat to causal inference, i.e., confidence that a relationship's observed functional form/strength reflects the actual causal effect of exposure on outcome. Indeed, several confounders (e.g., socioeconomic disadvantage [8, 9] and limited health care access [9]) may be driving the relationship between alcohol abstention and poor health outcomes.

Standard approaches to counter confounding include control for covariates via multivariable regression adjustment, stratification or (exact) matching [10]. These methods are limited; regression with adjustment requires correct specification of the functional form (i.e., algebraic form in regression equations) of covariate–outcome relationships [11, 12], and relies on extreme extrapolation when there are insufficient observations for all combinations of exposure, covariates and outcomes [12–14]. Similarly, the 'curse of dimensionality' – that the number of groups to compare increases exponentially with the inclusion of additional covariates, resulting in too few observations per group – is a limitation for matching and stratification [14]. It is also difficult to know if all (and only) relevant covariates have been identified, and whether they may be imperfectly measured, leading to

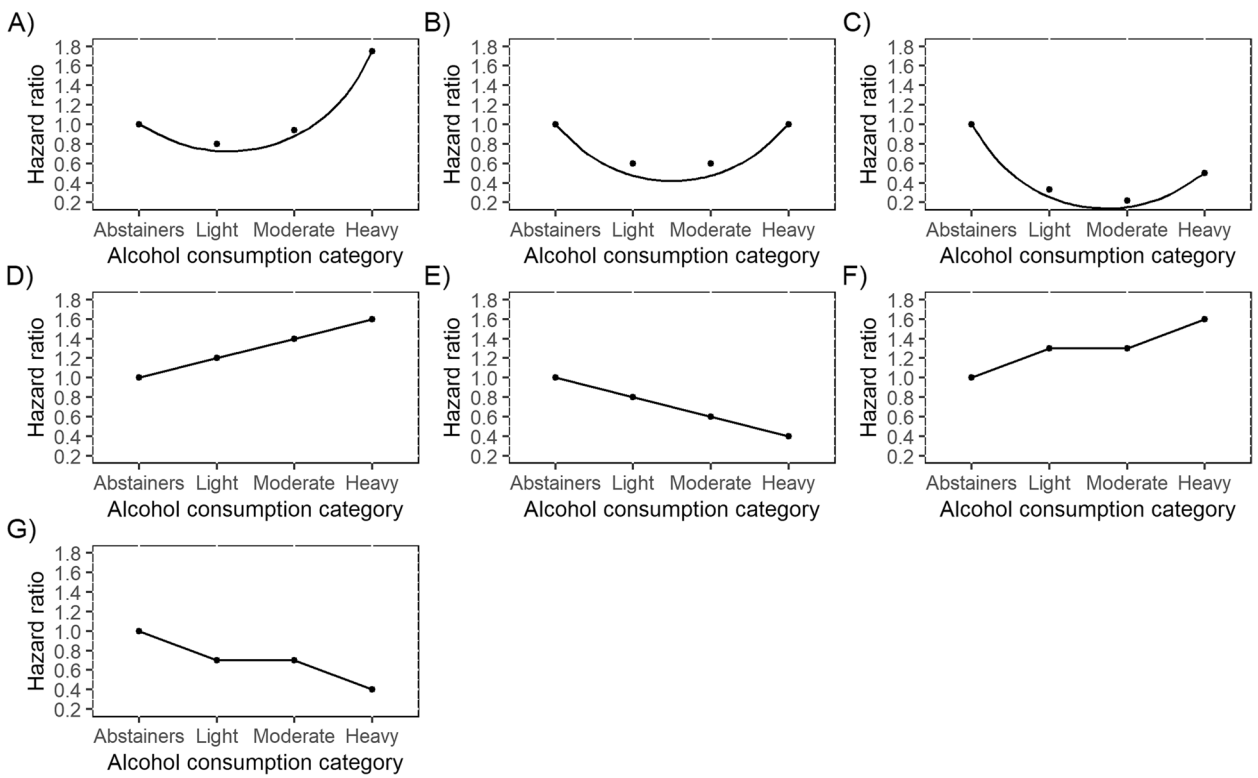


Fig. 1 Exemplar functional forms between alcohol consumption and health outcomes. Legend: A = J-shaped; B=U-shaped; C = reverse J-shaped; D = positive linear; E = negative linear; F = monotonically increasing; G = monotonically decreasing

residual confounding. Indeed, some propose that certain relevant confounders are not even measurable [10]. While confounding is generally obviated by the randomization mechanism in randomized controlled trials (RCTs), no long-term RCTs have been performed evaluating the relationships between alcohol use and long-term health outcomes because of ethical and compliance concerns.

To investigate alcohol–long-term health relationships, the field is therefore limited to observational studies. As such, efforts to improve causal inference have centered on mitigating bias. There is increased acknowledgement that data collection and analysis decisions can substantially affect conclusions about relationship strength and form [15, 16], so should be made in a considered, literature-informed manner. A particularly hazardous decision here is treating lifetime abstainers and former drinkers (whose abstinence is often precipitated by illness) as a homogenous group, thereby inducing a 'sick quitter bias' which effectively shifts poor health outcomes that have accrued to former drinkers to the abstaining group [7]. Certain tools and strategies can assist with limiting bias, such as creating directed acyclic graphs (DAGs) at study outset, and, following primary analysis, assessing robustness to methodological decisions ('sensitivity analysis'), bias ('bias analysis'), sample-specific confounding ('cross-cohort comparison') or research type ('triangulation') [17]. Indeed, the impact of analytical decisions such as how exposures are categorized and compared has been the focus of recent reviews/meta-analyses of alcohol–health research [15, 16].

Particularly promising, however, for addressing the identified limitations of existing research, are modern methods for data analysis and alternative observational designs. Conventional designs (e.g., prospective cohort studies) can be enhanced with modern analysis methods, such as propensity scores used for matching or weighting, and 'G-methods' such as marginal structural models (MSMs; which can account for time-varying variables that act as both confounders and mediators). Regarding alternative designs, twin studies and other family-based designs control automatically for shared confounders, as do negative controls [18]. Natural experiments are another alternative, mimicking the random allocation of RCTs and thus guarding against confounding and reverse causation. These include instrumental variables (IV) designs, where as-if/randomly allocated proxies for exposures are used in place of exposures themselves. Mendelian Randomization (MR), a kind of IV design, offers particular promise given the potential of genetic proxies for alcohol consumption.

While these methods are still limited in their approach to inferring causal relationships from observational data, they represent significant improvements over conventional analyses (see Table 1 for a full list of methods of interest, their advantages, and their limitations). Some of these approaches are gaining popularity [19], but they are not routinely applied to alcohol–long-term health research [20]. Importantly, reviews in this area rarely focus on improved analytical methods to counter confounding and tend to exclude novel study designs. This review therefore aims to identify all observational studies employing such approaches, and to synthesize their findings on the functional form and strength of alcohol–long-term health relationships.

Methods

Search strategy and study selection

This review's methods are reported in detail in the study protocol, which was registered with PROSPERO (CRD42020185861) and published [38]. Briefly, searches for peer-reviewed, English-language journal articles and grey literature on MEDLINE, PsycINFO, Embase and Scopus were performed in May 2020 with no limits on publication date. Choice of causal inference methods of interest incorporated expert feedback. Search terms were generated by adapting those from recent reviews and searching keywords/index terms of key eligible papers known to the authors, with iterative refinement. These included controlled vocabulary terms and free text words, and related to: 1) alcohol; 2) levels/patterns of drinking 3) observational, longitudinal studies; 4) analytical approaches to improve causal inference that are used in conjunction with conventional study designs; and 5) design-based approaches to improve causal inference. Groups of terms were combined as follows: 1 and 2 and ((3 and 4) or 5). MEDLINE search terms are provided in Table S1. Additionally, reference lists of eligible, retrieved studies were manually searched.

Only human research was eligible. The exposure of interest was level of alcohol consumption (volume over a given period), or level and pattern of consumption (incorporating frequency/heavy episodic drinking). While studies were eligible regardless of their findings on functional form, their methods must have been capable of detecting non-linearity – were it to be present. For this reason, studies were only eligible if they categorized alcohol consumption, and subsequently performed comparisons between a chosen reference category and the other levels of consumption. This approach does not require assuming a functional form (unlike a single regression using a continuous predictor). Specifically, a non-drinking/light drinking reference was required in addition to at least two other levels of consumption (alternative methods of comparison allowing for the detection of non-linearity were permitted for IV/MR designs). Any long-term

Table 1 Methods to enhance causal inference in observational research

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
			Confounding	Reverse causality	Selection bias	Measurement error		
Analytical methods applied to traditional longitudinal study designs								
Propensity scores (PS) [11, 19]	Covariate balancing propensity scores (CBPS)	-The PS is a single value reflecting the probability of exposure for an individual given their values on all relevant covariates -PS generation occurs as a data 'pre-processing' step prior to main analysis -Usually generated via logistic regression -Once generated, the PS can be used for matching, stratification, weighting, or as a covariate for adjustment in regression	✓				-Unlike standard methods, can handle large numbers of covariates -Not reliant on correctly modelling covariate- outcome relationships -Covariate balance after matching/weighting can be assessed	-Still relies on appropriate choice of covariates and accurate measurement -PS matching may not find matches for some treatment cases, leading to reduced sample size and limiting generalizability -Effect estimation with PS doesn't always perform better than regular adjustment -Most PS methods rely on manual covariate balance checking and refitting
G-methods [21–24]		-A family of methods intended for use with time-dependent variables -Developed as a solution to the problem of time-varying covariates affected by past exposure, including those that act as both confounders and mediators over time -The three G-methods are the G-formula, marginal structural models, and G-estimation, each relying on its own modelling assumptions					-Unlike standard methods to control for confounding, G-methods do not fix values of covariates, thus do not block mediation via the covariate, and avoid introducing collider bias -Accounting for changes in variables over time mitigates misclassification	-Still relies on appropriate choice of covariates and accurate measurement
	G-formula (aka G-computation or G-standardization) [21, 23, 25, 26]	-First models relationship given observed data (using actual exposure for each individual), and then predicts outcomes under counterfactual exposures, with the difference taken as the causal effect -Is a generalization of standardization (conditioning on covariates and then marginalizing) that accounts for dynamic variables by considering covariate distribution over follow-up time	✓		✓ (specifically, bias due to censoring)	✓	-Can be used to calculate risk ratios and risk differences -Well-suited to assess time-varying exposures	-Vulnerable to the 'g-null paradox'- null hypotheses tend to be rejected in large studies even when true -Usually requires specifying a statistical model, hence also being known as the 'parametric' G-formula

Table 1 (continued)

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
			Confounding	Reverse causality	Selection bias	Measurement error		
	Marginal structural models (MSMs) [21, 23, 25]	-Use weights based on inverse probability of exposure at each time point to create a pseudo-population where each combination of covariates is equally present in each exposure condition -Using these weights, MSMs then estimate the causal effect -The most popular of the G-methods	✓		✓ (specifically, bias due to censoring)	✓	-Simplest of the G-methods to understand and implement -Can also integrate censoring weights to account for differential attrition	-Not ideal for assessing exposure-confounder interactions, with standard MSMs unable to estimate interactions involving dynamic variables -Requires checking weight distribution, may require refitting (as with PS methods) -Cannot be used if all participants are exposed/unexposed on a particular level of a confounder
	G-estimation of structural nested models [23–25]	-At each wave assesses the relationship between exposure and likelihood of outcome given covariates, adjusting for exposure and covariate values from past waves, thus accounting for dynamic confounders affected by past exposure -Considered semi-parametric in that mean counterfactual outcomes under no exposure are unspecified	✓			✓	-Can be used even if all participants are exposed/unexposed on a particular level of a confounder -Can be used to assess exposure-confounder interactions	-Unlike the other G-methods, cannot account for selection bias arising from censoring, so data requires preliminary weighting to account for bias from censoring
Doubly robust methods [26]	Targeted maximum likelihood estimation; Augmented inverse probability weighting	-Incorporates both an estimation of the outcome mechanism (as in regression adjustment) and the exposure mechanism (as in propensity scores)	✓				-Only the outcome mechanism or the exposure mechanism need be consistently estimated to generate an unbiased estimate of exposure effect	-Still relies on appropriate choice of covariates and accurate measurement

Table 1 (continued)

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
			Confounding	Reverse causality	Selection bias	Measurement error		
Fixed effects regression [27–31]		-A technique developed in the econometrics literature for use with longitudinal data with repeat outcome measurements, only using information on within-subject variation, thus controlling for all time-invariant sources of confounding -Treats time-invariant characteristics that differ between individuals as fixed parameters (unlike in mixed models), allowing estimation of parameters of interest net of stable confounders -Each participant serves as own control	✓				-Removes the threat of all observed and unobserved time-invariant confounding -Models can be extended to include time-varying covariates	-Cannot overcome time-varying confounding without extending the model, and these variables must be observed/measured -Individuals with stable exposure values do not contribute to estimates; also leads to imprecision when exposures change little over time -Reducing confounding comes at the cost of more sampling variability -Cannot generate parameter estimates for stable characteristics like race
Causal mediation analysis [32–34]		-Integrates traditional mediation analysis (which separately estimates total effect of exposure on outcome, indirect effect via mediators, and direct effect unexplained by mediators) with the potential outcomes framework to allow for exposure-mediator interaction and non-linear relationships (i.e., is a non-parametric method) -Uses the concepts of 'controlled direct effect', 'natural direct effect', and 'natural indirect effect' -Makes explicit underlying assumptions related to unmeasured confounding, and encourages sensitivity analyses to test robustness to assumption violations	✓ (see advantages and limitations*)				-Effect decomposition is still possible given exposure-mediator interaction, nonlinearity, and categorical variables -Makes underlying assumptions explicit -Helps identify the mechanism/s of an exposure's effect; especially useful when heterogenous causal mechanisms at play -Can be extended to situations with multiple mediators -*If a variable completely mediates the exposure-outcome relationship and is shielded from confounders, confounder measurement is not needed	-*In practice, likely there will always be exposure-mediator or mediator-outcome confounding, so still need to observe and accurately measure covariates -Analyses make strong assumptions, necessitating sensitivity analyses

Table 1 (continued)

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
			Confounding	Reverse causality	Selection bias	Measurement error		
Alternative observational study designs								
Natural experiments [17, 35–37]		-Mimic RCTs by exploiting exogenous events that are truly randomized/approximate random assignment -Differ from true experiments in that exposure is not assigned by the researcher -Assignment may be as a result of naturally occurring phenomena (e.g., a weather event), or of human intervention implemented for reasons other than the research question (e.g., army draft lottery)					-In approximating randomization, obviates the need for accurate measurement of confounders -Potential to overcome measurement error, reverse causation, and selection bias	-Rare to find truly random or as if-random exposure assignment
	Standard natural experiments	-Natural experiments where individuals are as-if/randomly assigned to exposure and control groups	✓	✓	✓	✓		-May be difficult to find a standard natural experiment that maps on well to the actual research question of interest
	Instrumental variable analysis	-Assesses the relationship between an as-if/ randomly assigned *proxy* for the exposure of interest and the outcome -A valid instrumental variable must be associated with the exposure of interest, be independent of confounders of the exposure-outcome relationship, and should affect the outcome only via the exposure	✓	✓	✓	✓	-Useful when the exposure itself is difficult to manipulate or measure	-Difficult to find valid instrumental variables -Potential for weak instrument bias i.e., when the instrument explains a small amount of variance in exposure -Relies on assumption that the instrumental variable is not associated with exposure-outcome confounding

Table 1 (continued)

Method	Relevant sub-methods	Description	Can address:				Advantages	Limitations
			Confounding	Reverse causality	Selection bias	Measurement error		
	Genetic instrumental variables	-Kind of instrumental variable analysis using genetic variants as proxies for exposure -The most prominent technique is Mendelian Randomisation	✓	✓	✓	✓	-Genes cannot be confounded by environment, cannot be subject to reverse causality, and are stable over time -Multiple variants can be combined to explain more variance in exposure, mitigating weak instrument bias	-Genetic instrumental variables are proxies for lifelong exposure - this period may be longer than what the research question is interested in -Potential for weak instrument bias, pleiotropy (gene affecting more than one phenotype) and linkage disequilibrium (genes more likely to be inherited together) -Possible population stratification (there may be population subgroups with different distributions of genes)
Quasi-experiments [35]		-Like natural experiments, exploit exogenous events to assess relationships between exposures and outcomes, but lack random or as-if random assignment	✓	✓		✓	-Same as for natural experiments	-Without random assignment, confounding is still possible i.e., cause of exposure may also contribute to outcome
Family-based designs [17, 36]	Twin studies; Sibling comparison	-By comparing genetically related participants discordant for the exposure of interest, accounts for confounding from genetic or shared environmental sources	✓				-Controls for unmeasured/unobservable confounding (for shared covariates) -Comparing monozygotic and dizygotic twins enhances understanding of genetic vs. environmental confounding	-Still need to observe and accurately measure non-shared environmental covariates to control for this kind of confounding -May be difficult to find family members discordant for exposure of interest
Negative controls [18, 36]	Negative control exposures; Negative control outcomes	-Have the same confounding structures as the exposure-outcome relationship of interest, but lack a plausible causal mechanism -If association is greater for the relationship of interest than for the negative control, a causal relationship is likely; if not, suggests confounding/other shared biases responsible -May take the form of a negative control exposure or a negative control outcome	✓		✓ (specifically, immortal time bias)		-Can identify when confounding (or assumed shared bias) is responsible for apparent causal effects -Does not require observation/measurement of covariates	-Relies on assumption of no plausible causal mechanism in the negative control relationship -Relies on assumption that same confounding structure shared by relationship of interest and negative control relationship

health outcome was eligible; studies only reporting on acute/short-term conditions (e.g., injury) were excluded. Eligible studies needed to employ one of the pre-specified approaches to improving causal inference (see Table 1). Studies needed to be longitudinal cohort or case-control designs (excepting IV/MR designs). IV/MR studies must have performed formal IV analysis or otherwise provided estimates in terms of predicted alcohol consumption. Reviews and interventional studies were excluded.

Retrieved titles and abstracts were screened by one reviewer (RV), with a second reviewer (JW) additionally screening a random 25%. Full-text articles were independently assessed by two reviewers (RV and JW). A third reviewer (LM) was consulted regarding unresolved discrepancies.

Data extraction

Extraction was performed independently by two reviewers (RV and JW) using pre-piloted forms. Extracted data included publication details (author/s, year), participant characteristics (sample size, setting, mean age, eligibility criteria, cohort name), exposure details (number and spread of measurement occasions, nature of discretized categories), study design and analysis methods, health outcome/s (how assessed, whether binary/continuous, interval to measurement), and results (relationship strength and form). Study authors were contacted if further information was required (see Table S2).

Quality assessment

Given the range of designs targeted by this review, two risk of bias assessment tools were used. Cohort studies were assessed using the relevant Newcastle-Ottawa Scale (NOS) [39], and a recently developed tool specific to MR [40] was employed. One reviewer (RV) applied the tools to all studies, with a second reviewer (JW) additionally assessing a random 25%. In line with other similar reviews, formal assessment of evidence quality was limited to risk of bias [40, 41].

Synthesis and reporting

Data synthesis was limited to narrative description given the heterogeneity in health outcomes and methods employed by included studies. Reporting of this review complies with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) [42], the checklist for which can be found in Table S3.

Results

Characteristics of included studies

Sixteen articles met inclusion criteria (see Fig. 2), comprising four MR studies, nine twin designs, and three prospective cohort studies employing MSMs, reporting on health outcomes broadly related to cancer, diabetes, dementia, mental health, cardiovascular health, mortality, HIV, and musculoskeletal health. Two cohorts provided all twin study data, with all but one non-MR study conducted with Swedish or Finnish populations. Study characteristics are summarized in Table 2, and exclusion reasons for key ineligible papers are provided in Table S4. The two reviewers were in agreement on title and abstract screening for 93.13% of cases, and all discrepancies at full-text screening were resolved by discussion between reviewers.

Cancer

One twin study reported on two prostate cancer outcomes, prostate cancer and prostate cancer mortality, and results were consistent with reverse J- and J-shaped relationships, respectively, with light drinking at the nadir [43]. Hazard ratios (HRs) for abstainers ranged from 2.85–2.98 (monozygotic (MZ) and combined twin analyses) compared to light drinkers, and for heavy drinkers compared to light drinkers ranged from 1.63–2.00. With the maximum sample (all twin analyses), the confidence interval (CI) for the abstainer comparison was large. Results were similar when restricting analyses to twins discordant for prostate cancer outcome.

Diabetes

Two studies reported on diabetes, including one twin study, and one MR study. Carlsson et al.'s twin-based findings on the risk for type 2 diabetes (T2D) resemble a J-shape [44], but were critically underpowered, preventing interpretation. In Peng et al.'s MR study, local average treatment effects (LATEs) were employed to detect non-linearity – indicated by a LATE slope (effect of genetically-predicted alcohol consumption plotted against discretized observed alcohol consumption) significantly different to zero [45]. Results did not support non-linearity for diabetes-related biomarkers. Substantive effects were only interpreted for men, with women used as a negative control due to their lack of alcohol consumption (effect of genetic instrument on diabetic markers should be observed in men but not women if alcohol consumption is the only causal pathway). In linear IV analyses, small positive linear relationships with narrow CIs were found for fasting blood glucose (FBG), 2-h post-load plasma glucose (P2hBG) and insulin resistance (HOMA-IR), while there were no relationships for haemoglobin A1c (HbA1c) or beta-cell function (HOMA-beta).

Dementia

One twin study reported on dementia [46]. Analyses were consistent with a J-shape, with HRs for abstainers compared to light drinkers between 1.37–1.39, and

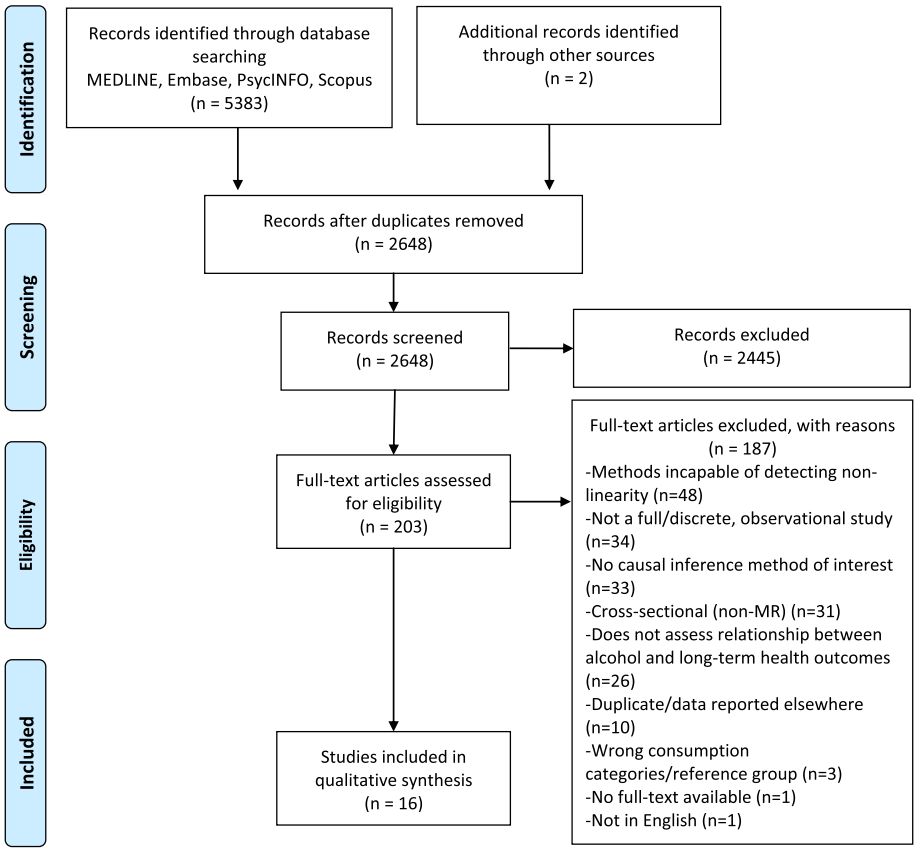


Fig. 2 PRISMA flow chart

for moderate-to-very-heavy drinkers compared to light drinkers between 1.57–3.07. Analyses of dementia-concordant twins demonstrated that only very heavy alcohol consumers had a much earlier age of onset than their light-drinking co-twins (10.67-year discrepancy in diagnosis compared to 6.79 years when both twins were light drinkers).

Mental health

Two articles reported on mental health outcomes: a prospective cohort study of depression employing MSMs, and a twin study assessing disability pension due to mental health diagnoses (MHD). Gemes et al. employed DAG-informed MSMs incorporating inverse probability of exposure and attrition weights [47]. Results support a U-shaped relationship (although there were few excessive consumers), with relative risks (RR) of 1.60 and 1.77 for abstainers and excessive drinkers respectively compared to light drinkers. Excluding those with baseline depression increased risk for excessive consumers such that the form approximated a J-shape, with RRs of 1.46 and 2.83 for abstainers and excessive drinkers respectively. Stratifying by gender, only abstainers were at increased risk for men, while both abstainers and excessive consumers remained at increased risk for women.

Samuelsson et al. discretized alcohol consumption into categories based on both volume and frequency [48]. In analyses of outcome-discordant twins, abstainers were at increased risk for pension due to MHD compared to light frequent consumers (HRs of 1.93–2.17 for MZ and combined twin analyses), as were heavy infrequent consumers with an HR of 2.10 (disappeared in MZ-only analyses; HR of 1.09), and light infrequent consumers (HRs of 2.80 and 3.67 for MZ and combined twin analyses respectively). Heavy frequent consumers were at *decreased* risk, although there were too few pairs to conduct MZ-specific analyses.

Cardiovascular events/diagnoses

Four studies reported on cardiovascular events/diagnoses, including two twin studies, one MSM and one MR study. Ilomaki et al. reported on myocardial infarction (MI), comparing various crude, adjusted and MSM models [49]. The DAG-informed MSM incorporated time-varying consumption and both time-varying and invariant covariates. Results were consistent with a

Table 2 Characteristics of included papers ($n = 16$)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Cancer									
Dickerman (2016) [43]	Prostate cancer (PC) and prostate cancer mortality -median 30 yrs. of follow-up	Twin	Male twins; mean age 40.1 at baseline	Older Finnish Twin Cohort	11,372; 225 and 43 discordant twin pairs for PC and PC-mortality respectively	-Co-twin (discordant for both alcohol consumption level and either time to diagnosis among outcome-concordant pairs concordant or time to event vs death/end of follow-up among outcome-discordant pairs) and pooled cohort Cox analyses to examine risk for PC and PC-mortality -Alcohol consumption measured twice (6 years apart) and averaged; categories: abstainers, **light** (0.01–3 drinks/wk), moderate (> 3–14 drinks/wk), heavy (> 14 drinks/wk); 1 drink considered = 12 g of alcohol	x	-PC-risk: HR (95% CI) Cohort MZ twins DZ twins All twins Abstainers 1.27 (.94,1.71) 2.85 (.67,12.1) **3.80** (1.36,20.6) **2.98** (1.35,6.60) Moderate 1.2 (.99,1.46) 1.28 (.6,2.74) 1.54 (0.92,2.57) 1.36 (.91,2.04) Heavy **1.46** (1.12,1.91) 2.00(.62,6.45) 1.71 (.87,3.39) 1.63 (.92,2.88) -PC-mortality: HR (95% CI) Cohort MZ twins DZ twins All twins Abstainers **1.90** (1.04,3.47) 2.31 (.19, 27.4) 1.83 (.42,8.03) 1.37 (.44,4.28) Moderate 1.22 (.76,1.97) 9.13 (.70,119) 1.43 (.37,5.62) 2.44 (.79, 7.52) Heavy 1.32 (.66,2.62) - 2.39 (.33,17.3) 7.31 (1.3, 41)	**Reverse J-shaped relationships, with light drinking at the nadir, were evident in all twin analyses of PC risk. J-shaped curves were evident in twin analyses of PC mortality. The evidence is consistent with a causal protective effect of light drinking, but twin analyses in general lacked power.**
Diabetes									
Carlsson (2003) [44]	Type 2 diabetes (T2D) −20 yrs. of follow-up	Twin	Same-sex twins without diabetes at baseline; mean age 34.3 (M) and 35.4 (F) at baseline	Finnish Twin Cohort	22,788; 27 discordant twin pairs analysed	-Co-twin (discordant for alcohol category) and pooled cohort Cox analyses to examine risk for T2D -Categories for pooled cohort (based on 3 questionnaires over 15 years): abstainers, **< 5 g/day**, 5–19.9 g/day (women)/5–29.9 g/day (men), ≥ 20 g/day (women)/ ≥ 30 g/day (men); for twin analyses (based on baseline questionnaire only): **low** (< 5 g/day), moderate (15–29.9 g/day), high (≥ 30 g/day)	✓	RR (95% CI) Cohort Men Cohort Women OR (95% CI) Twin Abstainers 1.1 (.7,1.5) 1.1 (.9,1.5) Moderate .5 (.2,1.3) 5–29.9 g/day (m); 5–19.9 g/day (f) .8 (.6,1.1) .7 (.4,1.1) High 1.2 (.4,3.9) ≥ 30 g/day (m); ≥ 20 g/day (f) .9 (.6,1.4) 1.6 (.8,3.5)	**Twin analyses were critically underpowered, preventing causal inference.**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Peng (2019) [45] *See also cardiovascular outcomes	Diabetes-related biomarkers (FBG, P2hBG, HbA1c, HOMA-IR, HOMA-beta) -cross-sectional	MR	Chinese adults living in the Yi-Ling district of Yichang; mean age 55 (SD 0.1) at baseline	One community from the Risk Evaluation of cAncers in Chinese diabeTic Individuals: a LONgitudinal (REACTION) study	4536	—1-sample MR using a single variant and standard IV analysis (2SLS); local average treatment effects (LATEs) computed for subgroups of observed alcohol consumption (non-zero LATE slopes indicate non-linearity) -Only standard, linear IV analysis was performed for categorical diabetes risk, not allowing for detection of non-linearity -Analysis performed in women as a negative control due to their lack of alcohol consumption -No conventional analyses conducted for comparison	✓	-Non-linear analyses: the LATE slopes for all diabetes-related biomarkers were not sig. Different to zero, indicating no non-linear relationships Unstandardizedβ (95% CI) Using log-transformed alcohol intake FBG (log-transformed) -.01 (—.04,.01) P2hBG (log-transformed) .01 (—.04,.06) HbA1c (log-transformed) -.01 (—.03,.00) HOMA-IR (log-transformed) -.00 (—.09,.08) HOMA-beta (log-transformed) .03 (—.04, .11) -Standard IV analyses: Unstandardizedβ (95% CI)Per 1-unit increase in log-transformed genetically-predicted alcohol consumption FBG (log-transformed)**.04** (.02,.05) P2hBG(log-transformed)**.07** (.04,.11) HbA1c (log-transformed).00 (—.01,.01) HOMA-IR(log-transformed)**.10** (.04, .17) HOMA-beta (log-transformed).00 (—.05,.06)	**No evidence of non-linear relationships between genetically predicted alcohol consumption and diabetes-related biomarkers. Alcohol appears to raise FBG, P2hBG and HOMA-IR in a linear fashion in males, although these are small effects.**
Dementia									
Handing (2015) [46]	Dementia —43 yrs. of follow-up	Twin	Twins without dementia prior to baseline; <=65 at baseline and >=60 at study end; mean age 54.2 (SD 5.9) at baseline; reported drinking <=100 g/day of alcohol	Swedish Twin Registry	12,362; 576 dementia discordant pairs (177 MZ); 396 concordant pairs (160 MZ)	-Co-twin (discordant for dementia) logistic regressions and pooled cohort Cox models used to examine risk for dementia; co-twin (concordant for dementia) mixed-effects analyses used to examine age at onset -Categories for pooled cohort + twin age of onset analyses: abstainers, **light** (> 0 and ≤ 5 g/d), moderate (> 5 and ≤ 12 g/d), heavy (> 12 and ≤ 24 g/d), and very heavy (> 24 g/d). For twin dementia risk analyses: abstainers, **light** (> 0 and ≤ 5 g/d), and moderate-to-very heavy (> 5 g/d)	✓	-Dementia risk: HR (95% CI) Cohort OR (95% CI) MZ twins All twins Abstainers 1.05 (1.00,1.11) Abstainers 1.37 (.6,3.16) 1.39 (.89,2.16) Moderate .98 (.92–1.04) Moderate-to-very heavy **3.07** (1.37,6.86) **1.57** (1.04,2.37) Heavy **1.1** (1.01,1.19) Very heavy **1.18** (1.01,1.36) -Dementia age at onset: Mean difference in years between twin diagnosed first vs twin diagnosed later (*p* value) MZ twins All twins Abstainers — 5.37 (.83) -5.49 (.68) Light —6.28 (NA – reference group) -6.79 (NA – reference group) Moderate — 5.41 (.13) -7.00 (.98) Heavy — 6.33 (.99) -6.73 (.32) Very heavy — 12.67 (.09) **-10.67** (.02)	**Results are consistent with a J-shaped curve. Twin analyses do support a causal role for increased risk of dementia for moderate-to-very heavy drinking, and faster dementia onset for very heavy drinkers.**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Mental health									
Gemes (2019) [47]	Depression —7-9yrs. of follow-up	MSM	General population aged 20–64 at recruitment; mean age 43.3 (SD 12.2) at baseline	Psykisk Hälsa–Arbete–Relationer (PART) cohort (Sweden)	5087	-MSM (weighted logistic regression) + standard logistic regression for comparison -Alcohol consumption measured pre-baseline (for use as time-variant confounder) and baseline; categories: no consumption, **light** (> 0- ≤ 7 drinks/wk), moderate (> 7- ≤ 14 drinks/wk), and excessive (> 14 drinks/wk); 1 drink = 12 g	✓	-Adjusting for baseline MDI score: RR (95% CI) Non-MSM MSM Abstainers 1.10 (.69,1.74) **1.60** (1.27,2.01) Moderate .54 (.28,1.04) 1.05 (.83,1.32) Excessive .61 (.21,3.07) **1.77** (1.13,2.78) -Excluding those with baseline depression (MDI > 26): RR (95% CI) Non-MSM MSM Abstainers 1.24 (.72,2.14) **1.46** (1.10,1.95) Moderate .63 (.29,1.41) .75 (.55,1.03) Excessive 2.72 (.61,12.19) **2.83** (1.80,4.44)	**MSM results support a non-linear (U/J-shaped) relationship between alcohol consumption and depression.**
Samuelsson 2013 [48] [a]	Disability pension (DP) due to mental health diagnoses (MHD) -median 10 yrs. of follow-up from prior study (baseline)	Twin	Twins with data from a prior study, and at time of that study were living in Sweden, < 65, and without DP/old age pension; mean age 52.9 (SD 5.6)	Swedish Twin Study of Disability Pension and Sickness Absence (STODS), drawn from the Swedish Twin Registry	28,613; 229 DP-discordant twin pairs (95 MZ pairs)	-Co-twin (discordant for dementia) Cox regressions + pooled cohort Cox regressions performed -Differentiated between frequent and infrequent drinkers (have/not consumed alcohol in previous two months); categories for pooled cohort + twin analyses: abstainers, **light frequent** (12-84 g/wk), moderate frequent (M: 85-168 g/wk.; F: 85–108 g/wk) and ≤ 12 g/d), heavy frequent (M: > 168 g/wk.; F: > 108 g/wk), light infrequent (12 g/occasion), moderate infrequent (M:13-48 g/occasion; F: 13-36 g/occasion), heavy infrequent (M: > 48 g/occasion; F:> 36 g/occasion)	✓	HR (95% CI) Cohort MZ twins DZ twins All twins Abstainers **1.99** (1.57,2.54) 1.93 (.54,6.96) **2.63** (1.05,6.62) **2.17** (1.06,4.45) Moderate frequent 1.07 (.78,1.49) .46 (.11,1.87) 2.70 (.81–9.02) 1.24 (.53,2.91) Heavy frequent .98 (.61,1.54) - .46 (.09,2.36) .48 (.12,1.90) Light infrequent 1.09 (.69,1.73) 2.80 (.24,32.6) 3.98 (.71,22.4) 3.67 (.91,14.8) Moderate infrequent 1.18 (.91,1.54) .52 (.18,1.49) 1.71 (.75,3.91) 1.03 (.55,1.93) Heavy infrequent 1.20 (.92,1.57) 1.09 (.33,3.53) **3.61** (1.28,10.2) 2.10 (.99,4.46)	**Results support a non-linear relationship between alcohol consumption and DP due to MHD, such that abstainers are at increased risk compared to light frequent drinkers. Light and heavy infrequent drinkers are also at increased risk, but CIs are very wide for the former, and effect disappears for the latter in MZ-only twins.**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Cardiovascular events/diagnoses									
Ilomaki (2011) [49]	Myocardial infarction (MI) −12-14 yrs. of follow-up	MSM	General population males; mean age 52 (SD 6.7) at initial exam (4 yrs. before 'baseline')	Kuopio Ischaemic Heart Disease Risk Factor Study (KIHD); Finland	1030	−5 discrete-time hazard models run and compared: M1 (baseline alcohol consumption with no covariate adjustment or inclusion of t-4 consumption), M2 (baseline alcohol consumption adjusted for covariates and t-4 consumption), M3 (baseline and t + 7 alcohol consumption with no covariate adjustment or inclusion of t-4 consumption), M4 (baseline and t + 7 alcohol consumption adjusted for covariates and t-4 consumption), M5 (MSM with baseline and t + 7 alcohol consumption with stabilized IP weights at t and t + 7) -Baseline alcohol categories: < 12 g/wk., **12-83 g/wk**, 84-167 g/wk., ≥168 g/wk.; additional alcohol measurements at t-4 and t + 7 included in various models	✓	RR (95% CI) M1 M2 M3 M4 M5 (MSM) <12 g/wk. 1.20 (.86,1.67) 1.01 (.70,1.45) **1.55** (1.08,2.21) 1.27 (.88,1.81) 1.27 (.88,1.83) 84-167 g/wk. 1.05 (.71,1.56) 1.13 (.75,1.72) 1.20 (.78,1.84) 1.27 (.81,2.00) 1.18 (.75,1.87) ≥168 g/wk. .98 (.60,1.58) 1.20 (.68,2.12) 1.40 (.91,2.18) **1.71** (1.03,2.85) 1.59 (.93,2.72)	**Results are generally consistent with increased risk for MI for those drinking less than weekly and those drinking heavily. However, effect sizes varied with model specifications.**
Kadlecova (2015) [50]	Stroke and transient ischaemic attack (TIA) events −43 yrs. of follow-up	Twin	Same-sex twins ≤60 with no history of stroke at baseline and with ≥5 yrs. of follow-up; mean age 50.5 (SD 5.29) at baseline	Swedish Twin Registry	11,644; 370 stroke/TIA discordant pairs (all MZ); 167 stroke/TIA concordant pairs (all MZ)	-Co-twin (discordant for stroke) logistic regressions and pooled cohort Cox models used to examine risk for stroke/TIA; co-twin (concordant for stroke) mixed-effects analyses used to examine time to stroke/TIA -Categories for pooled cohort + twin analyses: abstainers, **very light** (> 0-5 g/day), light (> 5 − 12 g/day), moderate (> 12-24 g/day), heavy (> 24 g/day)	✓	-Stroke/TIA risk: HR (95% CI) Cohort OR (*p* value) Twins Abstainers 1.11 (.98,1.23) Abstainers 2.22 (.058) Light .98 (.85,1.15) Light 1.56 (.17) Moderate .99 (.85,1.15) Moderate 1.59 (.26) Heavy **1.34** (1.04,1.70) Heavy 1.23 (.78) -Time to stroke/TIA: Mean difference in yrs. to stroke/TIA (*p* value) Twins Abstainers 2.07 (.11) Light .77 (.54) Moderate 1.15 (.47) Heavy **−5.68** (.029)	**Twin analyses are consistent with an increased risk of stroke/TIA for abstainers compared to very light drinkers, with somewhat increased risks for heavier drinking groups as well (reverse J-shape). Results support a causal role of heavy alcohol consumption in hastening time to event among those who experience stroke/TIA.**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Millwood 2019 [51]	Ischaemic stroke, intracerebral haemorrhage (ICH), total stroke, acute myocardial infarction (AMI), total coronary heart disease (CHD) -roughly 10 yrs. of follow-up	MR	Permanent residents from 10 Chinese regions aged roughly 35–74 and without major disabilities, (those with a history of CVD were excluded from analyses of disease incidence); mean age 52 (SD 11) at baseline	China Kadoorie Biobank	512,715; 161,498 of which had genotype data (male and female combined)	-1-sample MR using Cox models -Instrument composed of 9 combinations of 2 SNPs (ALDH2 rs671 and ADH1B rs1229984) within each of 10 geographic areas, producing 90 combinations overall; then, based on mean 'usual' alcohol consumption within each of those combinations (incorporating repeat measurements to account for measurement error), six final categories were produced, aligned with increasing genetically-predicted alcohol consumption: **Category1: 0-10 g/wk**, Category2: 10-25 g/wk., Category 3: 25-50 g/wk., Category 4: 50-100 g/wk., Category 5100-150 g/wk., Category 6 > 150 g/wk. -Conventional analyses using observed alcohol consumption conducted for comparison using Cox models, with consumption categories (for men): ex-drinker, non-drinker, occasional drinker (less than weekly), **< 140 g/wk**, 140-279 g/wk., 280-419 g/wk., ≥ 420 g/wk	✓	-Ischaemic stroke: Log RR (95% CI) MR Conventional C2 1.00 (.91,1.10) Ex-drinker **1.39** (1.32,1.46) C3 1.03 (.96,1.11) Non-drinker **1.21** (1.16,1.25) C4 **1.11** (1.02,1.20) Occasional 1.00 (.97,1.03) C5 **1.23** (1.15,1.33) 140-279 g/wk. **1.13** (1.07,1.19) C6 **1.23** (1.12,1.35) 280-419 g/wk. **1.23** (1.15, 1.32) ≥420 g/wk. **1.31** (1.21,1.41) Per 280 g/wk. (assuming linearity) **1.27** (1.13,1.43) **1.28** (1.19,1.38) -ICH: Log RR (95% CI) MR Conventional C2 1.01 (.88,1.14) Ex-drinker **1.52** (1.38, 1.68) C3 1.02 (.88,1.14) Non-drinker **1.33** (1.25,1.43) C4 1.08 (.96,1.22) Occasional 1.02 (.96,1.09) C5 **1.29** (1.15,1.44) 140-279 g/wk. **1.35** (1.21,1.52) C6 **1.54** (1.36,1.76) 280-419 g/wk. **1.52** (1.32,1.74) ≥420 g/wk. **1.73** (1.52,1.97) Per 280 g/wk. (assuming linearity) **1.58** (1.36,1.84) **1.59** (1.37,1.85) -Total stroke: Log RR (95% CI) MR Conventional C2 1.02 (.95,1.10) Ex-drinker **1.40** (1.34,1.46) C3 1.05 (.95,1.10) Non-drinker **1.23** (1.19,1.27) C4 **1.13** (1.06,2.10) Occasional 1.00 (.98, 1.03) C5 **1.27** (1.19,1.34) 140-279 g/wk. **1.16** (1.10,1.21) C6 **1.35** (1.26,1.45) 280-419 g/wk. **1.28** (1.20,1.36) ≥420 g/wk. **1.39** (1.31,1.48) Per 280 g/wk. (assuming linearity) **1.38** (1.26,1.51) **1.35** (1.27,1.44) -AMI: Log RR (95% CI) MR Conventional C2 1.02 (.87,1.19) Ex-drinker **1.66** (1.48,1.85) C3 1.05 (.93,1.19) Non-drinker **1.63** (1.51,1.76) C4 .93 (.81,1.07) Occasional **1.23** (1.16,1.31) C5 .94 9.81,1.09) 140-279 g/wk 1.11 (.97,1.26) C6 .97 (.83,1.15) 280-419 g/wk **1.19** (1.01,1.41) ≥420 g/wk 1.14 (.95,1.37) -Total CHD: Log RR (95% CI) MR Conventional C2 1.03 (.93,1.13) Ex-drinker **1.45** (1.38,1.52) C3 **1.08** (1.01,1.15) Non-drinker **1.31** (1.26,1.36) C4 .94 (.87,1.02) Occasional **1.07** (1.04,1.11) C5 1.04 (.97,1.11) 140-279 g/wk 1.03 (.97,1.09) C6 1.06 (.98,1.15) 280-419 g/wk **1.11** (1.03,1.19) ≥420 g/wk **1.13** (1.04,1.22)	**In contrast to conventional analyses, MR results are consistent with a monotonically increasing relationship between alcohol and stroke events. The results do not support a causal relationship between alcohol consumption and AMI or total CHD.**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Ropponen 2014 [52] *See also musculoskeletal health	DP due to circulatory system diagnoses −5 − 10 years of follow-up	Twin	Twins with data from a prior study, and at time of that study were living in Sweden, < 65, working and without DP/old age pension; mean age at baseline 53.7 (SD 5.7) [a]	Swedish Twin Registry	31,206; 216 DP due to circulatory system diagnoses-discordant pairs (of which 95 are MZ)	-Co-twin (discordant for DP due to MSD) Cox models and pooled cohort Cox models used to examine risk (stratified for sex in pooled model) -Categories for pooled cohort + twin analyses: abstainers, **light** (≤3 drinks/wk), moderate (F: >3- ≤ 7 drinks/wk.; M: >3- ≤ 14 drinks/wk), heavy (F: >7drinks/wk.; M: > 14 drinks/wk)	✓	HR (95% CI) Cohort MZ twins DZ twins All twins Abstainers 1.22 (.91,1.64) .97 (.31,3.01) 1.55 (.70,3.44) 1.3 (.69,2.45) Moderate .85 (.63,1.15) - .86 (.39,1.91) 1.54 (.78,3.04) Heavy **.79** (.63,.98) .91 (.49,1.68) .89 (.52,1.51) .86 (.57,1.28)	**Results do not offer clear support for causal relationships between alcohol consumption and later receipt of DP due to circulatory system diagnoses.**
Cardiovascular disease biomarkers									
Peng (2019) 28 *See also diabetic outcomes	Lipids: HDL-C, non-HDL-C, triglycerides (TG), total cholesterol (TC); blood pressure: systolic blood pressure (SBP), diastolic blood pressure (DBP); obesity anthropometric measures: BMI, waist circumference (WC), hip circumference (HC), waist-to-hip ratio (WHR) -cross-sectional	MR	Chinese adults living in the Yi-Ling district of Yichang; mean age 55 (SD 0.1) at baseline	One community selected from the Risk Evaluation of cAncers in Chinese diabeTic Individuals: a LONgitudinal (REACTION) study	4536	−1-sample MR using ALDH2 to instrument for alcohol consumption with standard IV analysis (2SLS); local average treatment effects (LATEs) computed for subgroups of observed alcohol consumption (non-zero LATE slopes indicate non-linearity) -Analysis performed in women as a negative control due to their lack of alcohol consumption -No conventional analyses conducted for comparison	✓	-Non-linear analyses: the LATE slopes for all lipids, blood pressure and obesity parameters in both sexes were not sig. Different to zero, indicating no presence of non-linear relationships Unstandardizedβ (95% CI) Using log-transformed alcohol intake HDL-C −.01 (−.06,.04) Non-HDL-C −.01 (−.13,.10) TG (log-transformed) .03 (−.05,.10) TC −.03 (−.16,.08) SBP −.85 (− 3.15,1.3) DBP −.70 (− 2.32,.71) BMI −.15 (−.51,.21) WC −.02 (− 1.09,1.09) HC .34 (−.37,1.07) WHR −.00 (−.01,.00) -Standard IV analyses: Unstandardizedβ (95% CI) Per 1-unit increase in log-transformed genetically-predicted alcohol consumption HDL-C .04 (−.00,.08) Non-HDL-C **.11** (.02,.20) TG (log-transformed) **.11** (.05,.16) TC **.15** (.06,.25) SBP **2.91** (1.06,4.76) DBP **3.03** (1.87,4.19) BMI **.57** (.28,.87) WC **2.37** (1.47,3.27) HC **1.01** (.39,1.62) WHR .02 (.00,.02)	**LATE results offer no evidence of non-linear relationships between genetically predicted alcohol consumption and lipid markers, blood pressure or obesity parameters. Alcohol appears to raise BMI, WC, HC, non-HDL-C, TG, TC, SBP and DBP in a linear fashion in males, with little effect on HDL-C or WHR.**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Silverwood (2014) [53]	Lipids: non-HDL-C, HDL-C, TG; blood pressure: SBP; obesity anthropometric measures: BMI, WC; inflammatory markers: CRP, interleukin 6 (IL-6) -cross-sectional	MR	Individuals of European descent from Europe and North America; mean age 56.75 (calculated from Holmes et al. [cite])	22 individual studies (18 cohorts, 2 nested case-control, 1 RCT, 1 case-control) from the Alcohol-ADH1B consortium	80,057 individuals total; 78,172 for SBP, 60,140 for non-HDL-C, 60,227 for HDL-C, 79,454 for BMI, 57,172 for WC, 63,367 for CRP, 23,535 for IL-6, 63,667 for TG	-1-sample MR using the rs1229984 polymorphism in ADH1B to instrument for alcohol consumption and standard IV analysis (2SLS); local average treatment effects (LATEs) computed for subgroups of observed alcohol consumption (non-zero LATE slopes indicate non-linearity) -Where non-linearity is present, the difference in outcome between no alcohol and median observed consumption in low (> 0–7 units/wk), moderate (7–21 units/wk), heavy (21–70 units/wk) and very heavy (> 70 units/wk) groups is predicted, as well as curve nadir, difference in outcome between nadir and abstinence, and level of consumption matching outcome for abstinence -No conventional analyses conducted for comparison	✓	-Non-linear analyses: the LATE slopes for SBP, non-HDL-C, BMI, WC and CRP were all sig. Different to zero, indicating non-linearity Unstandardizedβ (95% CI) Using log-transformed alcohol intake HDL-C .00 (−.06,.06) Non-HDL-C **.37** (.19,.55) TG (log-transformed) -.02 (−.10,.06) SBP **3.30** (1.0,5.5) BMI **.90** (.3,1.4) WC **2.00** (.6,3.6) CRP (log-transformed) **.26** (.10,.43) IL-6 (log-transformed) -.13 (−.34,.29) -Predicted differences in outcome between category medians and abstinence (unstandardized): 3.04 units/wk. 12.15 units/wk. 31.90 units/wk. 84.52 units/wk. Non-HDL-C −.39 (−.79,.06) -.15 (−.72,.47) .40 (−.28,1.10) **1.30** (.45,2.16) SBP .1 (− 5.5,6.1) 5.2 (− 2.6,13.9) **12.4** (3.4,22.1) **22.8** (12.2,34.6) BMI −.6 (− 2.2,.8) .2 (− 2.0,2.1) 1.6 (−.8,3.8) **3.9** (1.2,6.3) WC −.6 (− 4.7,3.5) 1.9 (− 3.9,7.8) 5.7(−.6,12.5) **11.5** (4.5,12.5) CRP (log-transformed) -.29 (−.68,.15) -.15 (−.68,.5) .22 (−.37,.95) **.83** (.15,1.69) -Predicted curve features for those outcomes with evidence of non-linearity (unstandardized): Nadir (units/wk.; 95% CI) Difference in outcome at nadir Units/wk. (95% CI) with outcome equivalent to abstinence Non-HDL-C **3.2** (.7,6.0) **−.39** (−.85,-.03) **16.9** (2.1, 48.2) SBP 1.00 (.0,3.6) −.7(− 5.4,.0) 2.8 (.0,19.6) BMI 2.3 (.0,6.0) −.6 (− 2.3,.0) 10.1 (.0,48.4) WC 1.5 (.0,5.4) −.8 (− 4.9,.0) 5.3 (.0,37.4) CRP 3.5 (.0,7.2) −.30 (−.75,.00) 19.4 (.0,66.0) -Standard IV analyses for those outcomes with no evidence of non-linearity: β (95% CI) Per 1-unit increase in log-transformed genetically-predicted alcohol consumption HDL-C −.02 (−.07,.03) TG (log-transformed) −.01 (−.06,.07) IL-6 (log-transformed) **.30** (.16,.45)	**Results support non-linear relationships between alcohol consumption and SBP, non-HDL-C, BMI, WC and CRP, with nadirs for these outcomes falling in the low drinking range. These relationships are best characterised by J-shapes with shallow nadirs, followed by gentle, elongated inclines. Results are consistent with a positive linear relationship between alcohol and IL-6, and do not support any causal relationships between alcohol and HDL-C or TG.**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Vu (2016) [54]	Lipids: TG, total cholesterol, HDL-C, HDL2-C, HDL3-C, LDL-C, sdLDL-C, apoB, Lp(a) -cross-sectional	MR	European Americans; mean age 54.3 (SD 5.7) at baseline	Atherosclerosis Risk in Communities (ARIC); USA	10,893 individuals total; 9911 for TG, 9751 for total cholesterol and LDL-C, 10,132 for HDL-C, 10,120 for HDL2-C and HDL3-C, 8102 for sdLDL-C, 7663 for apoB, 9924 for Lp(a)	−1-sample MR using a genetic risk score composed of 5 SNPs (rs2066702, rs1693457, rs1789891, rs698, and rs1126671) and standard IV analysis (2SLS); model fitted separately for quartiles of genetically-predicted alcohol consumption (**q1 = 1.49–3.63 g/wk,** q2 = 3.63–4.66 g/wk., q3 = 4.66–10.57 g/wk., q4 = 10.57–19.54 g/wk) to determine presence of non-linearity -Conventional analyses regressing outcomes on observed alcohol consumption were also conducted, using consumption categories: **never drinkers**, former/infrequent drinkers (< 1 drink/wk), low-to-moderate current drinkers (M: ≤210 g/wk., F: ≤105 g/wk), heavy current drinkers (M: > 210 g/wk., F: > 105 g/wk)	✓	-TG (log-transformed): Unstandardized β (95% CI) MR Using log-transformed alcohol intake Conventional Q2–**.06** (−.09,-.02) Former/ infrequent **−.08** (−.11,-.05) Q3–**.13** (−.20,-.07) Low-to-moderate **−.16** (−.19,-.13) Q4–**.08** (−.17,.00) Heavy **−.13** (−.17,-.09) -TC: Unstandardized β (95% CI) MR Using log-transformed alcohol intake Conventional Q2–**5.54** (−8.23,-2.85) Former/ infrequent **− 2.75** (−4.99,-.51) Q3–**7.71** (− 13.26,-2.15) Low-to-moderate −.73 (− 3.03,1.57) Q4–4.56 (− 11.36,2.25) Heavy **4.05** (.73,7.37) -HDL-C (log-transformed): Unstandardized β (95% CI) MR Using log-transformed alcohol intake Conventional Q2 .01 (−.01,.03) Former/ infrequent **.03** (.01,.04) Q3 .04 (.00,.07) Low-to-moderate **.14** (.12,.15) Q4 .03 (−.02,.07) Heavy **.26** (.23,.28) -HDL2-C (log-transformed): Unstandardized β (95% CI) MR Using log-transformed alcohol intake Conventional Q2 **.04** (.00,.07) Former/ infrequent **.07** (.04,.10) Q3 **.10** (.03,.17) Low-to-moderate **.17** (.14,.20) Q4 .06 (−.03,.15) Heavy **.30** (.26,.35) -HDL3-C: Unstandardized β (95% CI) MR Conventional Q2–.19(−.87,.49) Former/ infrequent .40 (−.14,.94) Q3 .08 (− 1.23,1.38) Low-to-moderate **4.16** (3.60,4.73) Q4 .11 (− 1.51,1.74) Heavy **8.70** (7.84,9.55) -LDL-C: Unstandardized β (95% CI) MR Using log-transformed alcohol intake Conventional Q2 **−4.60** (−7.18,-2.03) Former/ infrequent **−2.59** (−4.69,-.49) Q3 **−6.87** (− 12.24,-1.50) Low-to-moderate **−4.38** (−6.53,-2.23) Q4 −4.57 (− 11.11,1.96) Heavy **−7.48** (− 10.70,-4.25) -sdLDL-C (log-transformed): Unstandardized β (95% CI) MR Using log-transformed alcohol intake Conventional Q2 **−.04** (−.08,-.01) Former/ infrequent **−.04** (−.07,-.01) Q3 **−.08** (−.15,-.01) Low-to-moderate **−.05** (−.08,-.01) Q4 −.08 (−.17,.01) Heavy .01 (−.04,.06) -apoB (log-transformed): Unstandardized β (95% CI) MR Using log-transformed alcohol intake Conventional Q2 **−.03** (−.04,−.01) Former/ infrequent -.01 (−.02,.01) Q3 −.04 (−.07,.00) Low-to-moderate −.01 (−.03,.01) Q4 −.04 (−.08,.01) Heavy −.02 (−.04,.01) -Lp(a) (log-transformed): Unstandardized β (95% CI) MR Using log-transformed alcohol intake Conventional Q2 −.02 (−.09,.06) Former/ infrequent −.03 (−.10,.03) Q3 −.05 (−.20,.10) Low-to-moderate .01 (−.06,.08) Q4 −.01 (−.20,.18) Heavy −.06 (−.16,.03)	**Results support causal relationships between alcohol and TG, TC, HDL2-C, LDL-C, sdlDL-C and apoB, such that some level of consumption leads to more favorable levels than abstinence/very low consumption does. Analyses suggest non-linearity, with benefits peaking for most outcomes at roughly .5–.1 drinks per week, although effect sizes vary. Results do not support causal relationships between alcohol and HDL-C, HDL3-C and Lp(a).**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Mortality									
Sipila 2016 [55]	All-cause mortality -median 30.2 yrs. of follow-up	Twin	Same-sex twins aged 18–54 at baseline and free of chronic disease 6 years post-baseline; mean age 35.9 at baseline	Older Finnish Twin Cohort	14,787; 3389 drinking-discordant pairs (of which 926 pairs are MZ) [a]	-Co-twin (discordant for alcohol consumption) Cox models and pooled cohort Cox models used to examine risk for mortality -Categories for pooled cohort + twin analyses based on average of two measurements 6 yrs. apart: abstainers, **1-69 g/mnth**, 70-139 g/mnth, 140-209 g/mnth, 210-419 g/mnth, 420-839 g/mnth, 840-1199 g/mnth, ≥1200 g/mnth	✓	HR (95% CI) Cohort MZ twins All twins 0 g/mnth 1.02 (.85,1.22) .43 (.17,1.11) .96 (.63,1.45) 70–139 g/mnth .95 (.81,1.10) .64 (.29,1.40) .96 (.68,1.36) 140–209 g/mnth 1.08 (.91,1.29) 1.12 (.47,2.65) .85 (.57,1.26) 210–419 g/mth **1.29** (1.09,1.53) 1.55 (.65,3.71) 1.40 (.94,2.09) 420–839 g/mnth **1.56** (1.31,1.85) 1.70 (.69,4.22) **1.60** (1.06,2.43) 840–1199 g/mnth **2.17** (1.74,2.70) 1.65 (.54,5.08) **2.65** (1.50,4.69) ≥1200 g/mnth **2.81** (2.26,3.50) 3.18 (.81,12.43) **2.99** (1.60,5.59)	**Results are consistent with a monotonically increasing risk function. Twin analyses support a causal role for increased risk of all-cause mortality for moderate-to-very heavy drinking.**
HIV seroconversion									
Sander (2013) [56]	HIV seroconversion -median 10.5 yrs. of follow-up	MSM	Men who have sex with men and who were sexually active and HIV-seronegative at baseline; median age 33.4 at baseline	Multicenter AIDS Cohort Study (MACS); USA	3752	-MSM (Cox models) + standard Cox models for comparison -Categories for analyses based on average of two measurements 1 yr apart: **abstainers**, moderate (1–14 drinks/wk), heavy (>14 drinks/wk); 1 drink considered = 14 mL	✓	RR (95% CI) Non-MSM MSM Moderate .91 (.65,1.27) 1.10 (.78,1.54) Heavy 1.19 (.83,1.70) **1.61** (1.12,2.29)	**Results are consistent with a monotonically increasing risk function. Abstainers and moderate drinkers appear to have similar risk for HIV seroconversion, while heavy drinkers are at increased risk.**
Musculoskeletal health									
Pietikainen 2011 [57]	DP due to low back disorders (LBD) − 29 years of follow-up	Twin	Same-sex twins aged 18–64 and not receiving pension at baseline; mean age 33.2 (SD 12) at baseline	Finnish Twin Cohort	24,043; 504 (284 M) pairs discordant for DP due to LBD	-Co-twin (discordant for DP due to LBD) Cox models and pooled cohort Cox models used to examine risk for mortality -Categories for pooled cohort + twin analyses: abstainers, **light (≤3 drinks/wk),** moderate (F: >3-≤7 drinks/wk.; M: >3-≤14 drinks/wk), heavy (F: >7drinks/wk.; M: >14 drinks/wk)	✓	HR (95% CI) Cohort All twins Abstainer .85 (.61,1.20) .79 (.49,1.27) Moderate 1.07 (.79,1.43) .94 (.62,1.42) Heavy 1.08 (.80,1.46) 1.07 (.69,1.66)	**Results suggest a roughly monotonically increasing relationship between alcohol and later receipt of DP due to LBP, with reduced risk for abstainers the clearest feature.**

Table 2 (continued)

First author (year)	Health outcome/s and interval to follow-up	Study design	Participant characteristics	Cohort/s name	Sample size	Statistical methodology (reference group bolded)	Explicit testing for non-linearity?	Main results (statistically significant findings bolded)	Conclusion
Ropponen 2014 [52] *see also CVD section	DP due to musculoskeletal diagnoses (MSD) — 5-10 years of follow-up	Twin	Twins with data from a prior study, and at time of that study were living in Sweden, < 65, working and without DP/old age pension; mean age at baseline 53.7 (SD 5.7) [b]	Swedish Twin Registry	31,206; 922 DP due to MSD-discordant pairs (of which 357 are MZ)	-Co-twin (discordant for DP due to MSD) Cox models and pooled cohort Cox models used to examine risk (stratified for sex in pooled model) -Categories for pooled cohort + twin analyses: abstainers, **light** (≤3 drinks/wk), moderate (F: >3- ≤ 7 drinks/wk.; M: >3- ≤ 14 drinks/wk), heavy (F: >7drinks/wk.; M: > 14 drinks/wk)	✓	HR (95% CI) Cohort MZ twins DZ twins All twins Abstainers .93 (.80,1.07) 1.49 (.98,2.26) .8 (.54,1.19) 1.07 (.81,1.42) Moderate **.8** (.69,.93) **2.33 (**1.39,3.91) **.67** (.48,.95) 1.02 (.78,1.34) Heavy **.73** (.67,.81) 1.11 (.82,1.51) **.77** (.60,.98) .88 (.73,1.07)	**Results do not support a clear relationship between alcohol and later receipt of DP due to MSD, particularly given the discrepancy in direction of effects between MZ and DZ twins.**
Ropponen 2011 [58]	DP due to musculoskeletal disorder (MSD) and osteoarthritis specifically (OA) —29 years of follow-up	Twin	MZ and same-sex DZ twins aged ≥18 and working at baseline; mean age 33.2 (SD 12) at baseline	Finnish Twin Cohort	24,043; 1317 pairs discordant for DP due to MSD, 461 pairs discordant for DP due to OA	-Co-twin (discordant for DP due to MSD/OA) Cox models and pooled cohort Cox models used to examine risk for MSD and OA in men and women separately -Categories for pooled cohort + twin analyses: **abstainers**, light (≤3 drinks/wk), moderate (F: >3- ≤ 7 drinks/wk.; M: >3- ≤ 14 drinks/wk), heavy (F: >7drinks/wk.; M: > 14 drinks/wk)	✗	-For DP due to MSD: HR (95% CI) Cohort (M) Cohort (F) All twins (M) All twins (F) Light .91 (.59,1.40) 1.15 (.93,1.42) **2.04** (1.09,3.82) 1.07 (.81,1.41) Moderate .95 (.70,1.30) 1.23 (1.00,1.51) 1.50 (.94,2.38) .97 (.74,1.27) Heavy 1.01 (.73,1.39) 1.08 (.84,1.39) 1.58 (.99,2.53) 1.37 (.98,1.91) -For DP due to OA specifically: HR (95% CI) Cohort (M) Cohort (F) All twins (M) All twins (F) Light .99 (.50,1.96) 1.19 (.86,1.63) **4.07** (1.15,14.36) 1.33 (.82,2.14) Moderate .78 (.48,1.28) .96 (.69, 1.32) 2.01 (.82,4.93) 1.12 (.70,1.82) Heavy 1.04 (.64,1.71) .86 (.60,1.25) 2.39 (.97,5.89) **2.32** (1.21,4.47)	**Results do not offer enough support for a clear functional form, but are consistent with abstinence representing the lowest risk for DP due to MSD and OA specifically. Heavy drinkers were also at increased risk for both outcomes and sexes.**

[a] Ropponen et al. (2014) also used this cohort to examine the relationship between alcohol and disability pension due to mental health diagnoses – not reported on here as Samuelsson et al. used the more informative exposure categorization

[b] From correspondence with authors

Note: Where multiple models are reported in the original paper, the most adjusted analyses are reflected in the above table if (excluding for MR studies)

J-shape, with RRs of 1.27 for the lowest group and 1.59 for the highest group respectively, both with fairly narrow CIs (although they included the null). Models 2, 3 and 4 (non-MSM but with assorted incorporation of time-varying exposures/confounders) were consistent with monotonically increasing, reverse-J/U, and J-shaped relationships respectively.

Kadlecova et al. examined the relationship between midlife alcohol consumption and later stroke/transient ischaemic attack (TIA) using MZ twins [50]. In co-twin analyses, all groups had higher odds ratios (ORs) for stroke/TIA than very light consumers, with the highest estimate for abstainers (OR of 2.22; consistent with a reverse J-shape). In twins concordant for stroke/TIA, heavy consumers had shorter time to event (5.68 years), while all other groups had slightly *longer* time to event than the very light drinking group.

Milwood et al. also reported on stroke (ischaemic, intracerebral haemorrhage, total stroke), in addition to acute myocardial infarction (AMI), and total coronary heart disease (CHD) [51]. As for Peng et al., women were negative controls. Comparing categories of genetically-predicted alcohol consumption, MR results were consistent with monotonically increasing relationships for stroke and subtypes, and with no causal relationships for AMI/CHD. Log RRs for those relationships with evidence of linearity ranged from 1.27–1.58 per 280 g of alcohol per week consumed.

Finally, Ropponen et al. reported on disability pension due to circulatory system diagnoses using a twin design [52]. For same-sex twins discordant for outcome, there appeared to be little clear relationship in MZ-only analyses, while the dizygotic (DZ)-only analyses were consistent with a reverse J-shape (both moderate and heavy consumers had HRs < 1 compared to abstainers).

Continuous cardiovascular measures

Three MR studies reported on lipids, with two of these also reporting on blood pressure and obesity anthropometrics. Peng et al., using the same methods as for diabetes outcomes, found no evidence of non-linear relationships for any lipids, blood pressure measures or obesity anthropometrics, but did find positive linear relationships for BMI, waist circumference, hip circumference, non-HDL-C, triglycerides (TG), total cholesterol (TC), systolic blood pressure (SBP) and diastolic blood pressure (DBP).

Silverwood et al. applied the LATE method to pooled data from 22 studies to examine cardiovascular and inflammatory measures, finding non-linearity (J-shapes) for SBP, non-HDL-C, BMI, WC and C-reactive protein (CRP) [53]. Nadirs for these relationships corresponded to small volumes of alcohol, ranging from 1 to 3.5 units of alcohol/week, and the differences in biomarker outcomes at the nadir compared with abstinence were also small. For those outcomes with no evidence of non-linearity, standard IV analysis revealed a positive linear relationship between alcohol consumption and IL-6 (an inflammatory marker), but a lack of relationship with HDL-C and triglycerides.

Finally, Vu et al. discretized genetically-predicted alcohol consumption into quartiles, comparing lipids in each with the lowest quartile [54]. Results provide evidence of non-linearity for TG, TC, HDL2-C, LDL-C, sdLDL-C and apoB. For these outcomes, all quartiles had more favorable levels than quartile 1. Benefits peaked at quartile 3, equivalent to .5–.1 genetically-predicted units per week. Results do not support causal relationships between alcohol and HDL-C overall, HDL3-C or Lp(a).

Mortality

One twin study assessed all-cause mortality, finding the three heaviest alcohol consuming groups had greater mortality risk compared to their lighter consuming reference, with HRs ranging from 1.60–2.99 [55]. This pattern replicated in the MZ-only sample, but with less precision and with CIs crossing the null. Abstainers were at decreased risk (HR of .43) in the MZ-only sample, but the CI included the null.

HIV seroconversion

One study employing MSMs reported on HIV seroconversion in men, incorporating both inverse probability of exposure and censoring weights [56]. Results were consistent with a monotonically increasing risk function, with an RR of 1.61 for heavy drinkers compared with abstainers.

Musculoskeletal health (MSD)

Three twin studies reported on MSD – all using receipt of disability pension due to MSD conditions as the outcome. Pietikainen et al. found a roughly monotonically increasing relationship for pension due to lower back disorders, with lower risk for abstainers the clearest effect (HR of .79), although all CIs included the null [57]. When stratified by sex, the protective effect of abstinence was more pronounced in men (HR .45; CI .13,1.48), while the functional form in women changed such that there was also reduced risk for moderate consumers (HR .76; CI .45,2.32).

Using the same cohort, Ropponen et al. (2011) examined disability pension due to osteoarthritis and due to MSD more generally [58]. Results were not consistent with a clear functional form but do support abstainers having the lowest risk for both outcomes.

Finally, in a different cohort, Ropponen et al. (2014) evaluated risk for disability pension due to MSD [52]. Again, outcome-discordant twin analyses did not reveal a clear functional form, with discrepant results between MZ and DZ samples.

Risk of bias

Cohort studies ranged in scores from 7 to 9 out of 9 on the NOS tool (see Table S5), with most losing marks for self-reported ascertainment of exposure. MR studies all had a combination of low and moderate risk across the five domains (see Table S6).

Discussion

This review found that improved causal inference methods have been applied minimally to research on alcohol–long-term health relationships. Non-linearity was apparent for several outcomes: prostate cancer and related mortality (reverse J-shaped and J-shaped respectively), dementia risk (J-shaped; although age of onset better characterized by monotonically increasing relationship), mental health (U/J-shaped for depression; increased risk for abstainers for disability pension due to MHD), and certain lipids (LDL-C; reverse J-shaped, sdLDL-C and apoB; monotonically decreasing, and HDL-2C; inverted reverse J-shaped). However, many of the individual comparisons from which these overall forms were of small effect size or were imprecise. While the level of consumption coinciding with lowest risk varied between outcomes, it tended to fall in the light range – as little as .5–.1 units/week [54]. Positive linear/monotonically increasing relationships were found for DBP, hip circumference, IL-6, all-cause mortality, and HIV seroconversion (although it is not possible to partition short-term pathways via risky sexual behavior and longer-term effects on immune function). No relationships were found between alcohol and HDL-3C, Lp(a), or waist-to-hip ratio.

Where multiple studies reported on an outcome, findings were inconsistent. For diabetes-related biomarkers, there was a positive linear relationship, but the one study reporting on T2D itself lacked power to support a clear functional form. For cardiovascular events/diagnoses, one twin study found preliminary evidence for a J-shaped relationship with myocardial infarction, while MR failed to find any relationship. For stroke, one twin study found a reverse J-shape (monotonically increasing for time to stroke), while MR found monotonically increasing relationships. For cardiovascular disease more generally, no clear causal relationship emerged. The results of Millwood et al. imply that broad outcomes (in this case total CHD) mask various discrepant sub-functional forms (as is likely for all-cause mortality) [59]. For cardiovascular biomarkers, all three studies evaluating HDL-C were consistent with a lack of causal relationship. There was little consistency across other cardiovascular biomarkers, lipids and obesity anthropometric measures, with conflicting functional forms found for non-HDL-C, triglycerides, total cholesterol, SBP, BMI and waist circumference. This was the case even when the same MR method was used, which may reflect the impact of using different ethnic populations and genetic instruments. Finally, for musculoskeletal health, results varied between a monotonically increasing form, no clear functional form, and no clear functional form with nadir at abstinence.

Some of these findings are roughly consistent with the conclusions on functional form made by recent reviews of the broader observational literature, but there were also outcomes where the present findings do *not* concord with the broader literature; evidence has been triangulated in Table 3. Triangulation with the broader observational research is key, as evidence for most health outcomes was only available from one or two studies in the present review, and thus not definitive.

Importantly though, where included studies performed conventional analyses for comparison with modern methods to address confounding, results were often discrepant. Most starkly, Millwood et al. found typical J/U/reverse-J-shaped relationships via conventional analyses, but monotonically increasing (or no) relationships when using MR. Even when functional forms roughly replicated across methods, the strength of individual comparisons differed. With pooled cohort analysis, Dickerman et al. found abstainers had slightly increased risk for prostate cancer, compared to much greater risk when utilising discordant twins. The application of improved causal inference methods is therefore essential to accurately characterize alcohol–health relationships.

Also of note, there were several methods of interest for which no eligible studies were identified. It may be that certain methods are ill-suited to address this research question – for example, it may be difficult to find a non-genetic IV to proxy for multiple levels of alcohol consumption, while others, such as G-estimation, may not yet have gained traction in research more generally [21]. Negative controls, while not identified as a primary design approach, were incorporated into two of the MR studies. Of the methods that were represented, there were fewer eligible studies than expected – particularly for MR, where many articles were excluded for only performing linear IV analyses, or for providing estimates in terms of the effects of genetic variants (e.g., ADH1B A-allele carriers vs non-carriers) rather than genetically-predicted alcohol consumption. Consistent with other reviews of the literature [10], covariates controlled for across studies varied considerably (see Table S7).

Table 3 Triangulation of included studies with reviews of the broader observational literature

Outcome	Study/ies first author (year)	Simplified summary of findings	Consistent with broader literature?	Details
Prostate cancer	Dickerman (2018)	-Reverse J-shaped relationship	No	A 2016 meta-analysis found a monotonically increasing relationship [60]
All-cause mortality	Sipila (2016)	-Monotonically increasing	Yes	A 2016 meta-analysis found a monotonically increasing relationship [16]
Diabetes biomarkers:				
–FBG	Peng (2016)	-Positive linear relationship	Yes	A 2017 review found results consistent with a positive linear relationship [61]
–Glucose tolerance, insulin sensitivity and glycated haemoglobin	Peng (2016)	-Positive linear relationships for P2hBG (glucose tolerance) and HOMA-IR (insulin sensitivity) -Lack of relationship for HbA1c (glycated haemoglobin) and HOMA-beta (insulin sensitivity)	N/A	A 2017 review found that, for the other diabetes biomarkers included in this review, there was not enough/consistent evidence from which to draw conclusions [61]
CVD biomarkers:				
–Blood pressure	Silverwood (2014) Peng (2019)	-Non-linear relationship (small protective effect of light drinking) for SBP -Positive linear relationships for both SBP and DBP	No Yes	A 2018 meta-analysis found no protective effect for hypeternsion [62]
–Obesity anthropometrics	Silverwood (2014) Peng (2019)	-Non-linear relationship (small protective effect of light drinking) for BMI and waist circumference -Positive linear relationships for BMI, waist circumference, and hip circumference, with no relationship for waist-to-hip ratio	Mixed Mixed	A 2011 review found mixed evidence on the relationship between alcohol and weight/measures of abdominal adiposity, with heavy drinking generally associated with increased values, but moderate consumption either associated with lower values or not associated at all (suggested that discrepancies may depend on type of alcohol consumed) [63]
–Inflammatory markers	Silverwood (2014)	-Non-linear relationship (small protective effect of light drinking) for CRP but a positive linear relationship for IL-6	N/A	A 2021 review of studies from the previous decade found no relevant longitudinal observational studies that were capable of detecting non-linear relationships [64]
–Lipids: HDL-C	Silverwood (2014) Peng (2019) Vu (2016)	-No relationship for HDL-C -No relationship for HDL-C -No relationship for HDL-C, or HDL-3C, but an inverted reverse J-shaped relationship for HDL-2C (a beneficial effect)	No No Mixed	A 2021 review of studies from the previous decade found just one longitudinal observational study, which reported a reverse J-shaped relationship in decreases in HDL-C from baseline [64]
–Lipids: other	Silverwood (2014) Peng (2019) Vu (2016	-No relationship for TG, but a non-linear relationship (small protective effect of light drinking) for non-HDL-C -Positive linear relationships for TG, non-HDL-C, and TC -Reverse J-shaped relationships for LDL-C, TG and TC, no relationship for Lp(a), and monotonically decreasing relationships for sdLDL-C and apoB	N/A	A 2021 review of studies from the previous decade found no relevant longitudinal observational studies [64]

Table 3 (continued)

Outcome	Study/ies first author (year)	Simplified summary of findings	Consistent with broader literature?	Details
CVD events/diagnoses:				
–Myocardial infarction	Ilomaki (2011) Millwood (2019)	-J-shaped relationship -No relationship	No	A 2017 review found alcohol was detrimental for myocarditis [2], but no reviews on myocardial infarction were found
–Stroke	Kadlecova (2015) Millwood (2019)	-Reverse J-shaped relationship for stroke and TIA -Monotonically increasing relationships for ischaemic stroke, intracerebral haemorrhage and total stroke	Mixed	A 2016 meta-analysis found J-shaped relationships for ischaemic stroke, but monotonically increasing relationships for intracerebral haemorrhage and subarachnoid haemorrhage [65]
–Total coronary heart disease	Millwood (2019)	-No relationship	No	A 2016 meta-analysis found a U-shaped relationship [66]
–Circulatory system diagnoses	Ropponen (2014)	-No clear functional form	N/A	Most reviews evaluate CVD sub-conditions, finding divergent functional forms
Dementia	Handing (2015)	-J-shaped relationship	Yes	A 2021 review found most recent evidence is consistent with a J-shaped relationship [67]
Mental health	Gemes (2019) Samuelsson (2013)	-U—/J-shaped relationship -Non-linear relationship with abstainers at increased risk over light frequent drinkers	Yes	A 2020 meta-analysis found a J-shaped relationship with depressive symptoms [68]
HIV seroconversion	Sander (2013)	-Monotonically increasing relationship	N/A	There is a lack of dose-response information on alcohol's relationship with HIV [69, 70]
Musculoskeletal disorders	Pietikainen (2011) Ropponen (2014) Ropponen (2011)	-Monotonically increasing relationship -No clear functional form -No clear functional form, but abstainers have lowest risk	N/A	No reviews on the relationship between alcohol and musculoskeletal disorders were found

Note: Risk for type 2 diabetes (Carlsson et al.) is excluded from this table as the analyses were critically underpowered, and thus the results are not appropriate for inclusion in the triangulation process. Findings from RCTs were not integrated into triangulation as they are almost exclusively conducted in short-term time frames, which is not of interest to this review

Strengths

This review applied a novel framework to examining alcohol–health relationships, identifying and synthesizing information from those observational studies that best mitigate confounding and thus promote causal inference. The search strategy included terms for a broad range of analytical and design-based approaches informed by the literature and consultation with experts. As many included studies performed both conventional analyses and causal inference approaches, this review was able to highlight the difference that such methods make. A further strength was that all long-term health outcomes were eligible, providing a comprehensive picture of the state of the evidence base, and importantly, on the large gaps in the literature where the aforementioned methods require application.

Limitations

While the included studies mitigate confounding, other methodological limitations (not necessarily captured by the NOS) may be present. For example, the prospective cohort studies likely suffered from sick quitter bias in failing to separate ex-drinkers from lifelong abstainers – exacerbated in those studies where baseline mean age was over 50. Misclassification was likely in many of the twin studies as most based classification on a single measurement, and most of these focused on shared confounding, without additionally controlling for measured covariates. While MSMs can account for consumption and covariates at multiple timepoints, these studies were still vulnerable to residual confounding, with Gemes et al. noting that unmeasured social confounders may partially underpin their findings. As approaches to minimize these biases consist largely of literature-informed, considered researcher decisions and sensitivity analyses, they are not suited to systematic database searching, and were not the focus of this review. While MR is largely immune to both misclassification and confounding, it suffers from its own idiosyncratic limitations, with controversy over its application to alcohol–health research specifically [71–73]. For example, two of the included MR studies discretized genetically-predicted alcohol consumption, resulting in the lowest categories aligning with occasional consumption – not strictly comparable with abstinence.

Additionally, despite evidence of the importance of accounting for pattern of consumption [74], MR studies are limited in their ability to do so [73], and only one cohort study [48] used drinking pattern as the exposure, rather than volume alone (or volume and frequency separately). Finally, several of the included studies evaluated alcohol's relationship with condition-specific disability pension, rather than the condition itself. This is an imperfect proxy, with receipt of pension also reflecting the interference of the disease with one's ability to work, as well as incentive to apply [57]. Given that all included studies evaluating musculoskeletal health used disability pension as a proxy, findings with respect to these outcomes should be interpreted with caution.

Future directions

This review has identified clear gaps in alcohol–long-term health research, demonstrating great potential for further application of enhanced causal inference methods. Analysis methods such as MSMs are particularly promising as they do not require the establishment of twin registries or large genetic datasets, but are able to mitigate confounding, differential censoring and misclassification. And as evidenced by studies included in this review, they are suitable for examining alcohol–long-term health relationships.

Analysis and design approaches that mitigate confounding should be combined with sensitivity analyses such as multiverse analyses (to quantify robustness to data processing/analysis decisions) [75], as well as bias analyses such as e-value generation (to quantify robustness to unmeasured confounding) [76]. Given the unique advantages and limitations of each analytical and design-based approach, triangulation of findings across observational evidence is crucial. Combining data across studies through data harmonization techniques should also be considered to mitigate power issues, which were evident in several included studies with rare exposure–outcome combinations. Finally, while acknowledging the limitations of the included studies, the identification of some evidence consistent with causal protective effects of light-to-moderate alcohol consumption for several health outcomes justifies further exploration of the biological mechanisms that could underpin these (potential) effects. This is particularly true of those outcomes for which findings were concordant with the broader observational literature (see Table 3).

Conclusions

This novel review found that, when enhanced causal inference approaches are applied, a variety of functional forms – including linear, J-shaped, and no relationship – are found between alcohol consumption and various long-term health outcomes. However, few studies have employed these methods, with covariate-adjusted, conventional cohort analyses remaining dominant, preventing a conclusive picture of the nature of these relationships from emerging. Given that associations found between moderate alcohol consumption and good health impact safe drinking guidelines and public health policy [77, 78], further research employing methods to mitigate confounding and other biases is urgently required to establish whether such findings are truly causal.

Abbreviations
RCT: Randomised controlled trial; DAG: Directed acyclic graph; MSM: Marginal structural model; IV: Instrumental variables; MR: Mendelian randomisation; NOS: Newcastle-Ottawa Scale; PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis; T2D: Type 2 diabetes; LATE: Local average treatment effects; FBG: Fasting blood glucose; P2hBG: 2-h post-load plasma glucose; HOMA-IR: insulin resistance; HbA1c: Haemoglobin A1c; HOMA-beta: beta-cell function; MHD: Mental health diagnoses; MZ: Monozygotic; MI: Myocardial infarction; TIA: Transient ischaemic attack; AMI: Acute myocardial infarction; CHD: Coronary heart disease; DZ: Dizygotic; TG: Triglycerides; TC: Total cholesterol; SBP: Systolic blood pressure; DBP: Diastolic blood pressure; CRP: C-reactive protein.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-021-01486-5.

Additional file 1.

Acknowledgments
The authors would like to thank Jessica Hughes, librarian at the University of Sydney, for assisting with the literature search, Dr. Cheryl McQuire (University of Bristol) and Dr. Tim Watkins (University of New South Wales) who were consulted on the causal inference methods included in our search, as well as the authors of included studies who were generous with their time in assisting with queries.

Authors' contributions
RV, MS, TS and LM conceptualised the study. RV and JW undertook data extraction and curation. All authors contributed to, read, revised, and approved the final manuscript.

Funding
Rachel Visontay is funded by the National Health and Medical Research Council (GNT1190255), as well as the University of Sydney. Publication fees were supported by the National Health and Medical Research Council Centre of Research in Prevention and Early Intervention in Mental Illness and Substance Use (PREMISE; APP11349009). None of these bodies played a role in the design of the study, collection, analysis and interpretation of data, or manuscript preparation.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]The Matilda Centre for Research in Mental Health and Substance Use, The University of Sydney, Level 6, Jane Foss Russell Building, G02, Sydney, NSW 2006, Australia. [2]Centre for Healthy Brain Ageing, University of New South Wales, Level 1, AGSM (G27), Gate 11, Botany Street, Sydney, NSW 2052, Australia.

Received: 12 August 2021 Accepted: 29 November 2021
Published online: 14 January 2022

References

1. Griswold MG, Fullman N, Hawley C, et al. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet. 2018;392(10152):1015–35.
2. Rehm J, Gmel GE Sr, Gmel G, et al. The relationship between different dimensions of alcohol use and the burden of disease—an update. Addiction. 2017;112(6):968–1001.
3. Keyes KM, Calvo E, Ornstein KA, et al. Alcohol consumption in later life and mortality in the United States: results from 9 waves of the health and retirement study. Alcohol Clin Exp Res. 2019;43(8):1734–46.
4. Andreasson S, Chikritzhs T, Dangardt FHH, Naimi T, Stockwell T. Evidence about health effects of "moderate" alcohol consumption: reasons for scepticism and public health implilcations. Alcohol Soc. 2014;6.
5. Naimi TS, Stockwell T, Zhao J, et al. Selection biases in observational studies affect associations between 'moderate' alcohol consumption and mortality. Addiction. 2017;112(2):207–14. https://doi.org/10.1111/add.13451.
6. Callinan S, Chikritzhs T, Livingston M. Consistency of drinker status over time: Drinking patterns of ex-drinkers who describe themselves as lifetime abstainers. J Stud Alcohol Drugs. 2019;80(5):552–6. https://doi.org/10.15288/jsad.2019.80.552.
7. Shaper AG, Wannamethee G, Walker M. Alcohol and mortality in British men: explaining the U-shaped curve. Lancet. 1988;332(8623):1267–73. https://doi.org/10.1016/S0140-6736(88)92890-5.
8. Kerr WC, Lui CK, Williams E, Ye Y, Greenfield TK, Lown EA. Health risk factors associated with lifetime abstinence from alcohol in the 1979 National Longitudinal Survey of youth cohort. Alcohol Clin Exp Res. 2017;41(2):388–98.
9. Naimi TS, Brown DW, Brewer RD, et al. Cardiovascular risk factors and confounders among nondrinking and moderate-drinking U.S. adults. Am J Prev Med. 2005;28(4):369–73. https://doi.org/10.1016/j.amepre.2005.01.011.
10. Wallach JD, Serghiou S, Chu L, et al. Evaluation of confounding in epidemiologic studies assessing alcohol consumption on the risk of ischemic heart disease. BMC Med Res Methodol. 2020;20(1):1–10. https://doi.org/10.1186/s12874-020-0914-6.
11. Brookhart MA, Wyss R, Layton JB, Stürmer T. Propensity score methods for confounding control in nonexperimental research. Circ Cardiovasc Qual Outcomes. 2013;6(5):604–11. https://doi.org/10.1161/CIRCOUTCOMES.113.000359.
12. Thoemmes F, Ong AD. A primer on inverse probability of treatment weighting and marginal structural models. Emerg Adulthood. 2016;4(1):40–59. https://doi.org/10.1177/2167696815621645.
13. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. BMJ. 2016;353:1–6. https://doi.org/10.1136/bmj.i1981.
14. Melberg HO. Does Moderate Alcohol Intake Reduce Mortality? (Jon Elster OG, Moene AH and K, eds.). Oslo Academic Press; 2006.
15. Chu L, Ioannidis JPA, Egilman AC, Vasiliou V, Ross JS, Wallach JD. Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk. Int J Epidemiol. 2020;49(2):608–18. https://doi.org/10.1093/ije/dyz271.
16. Stockwell T, Zhao J, Panwar S, Roemer A, Naimi T, Chikritzhs T. Do "moderate" drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. J Stud Alcohol Drugs. 2016;77(2):185–98.
17. Richmond RC, Al-amin A, Smith GD, Relton CL. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. Early Hum Dev. 2014;90:769–80. https://doi.org/10.1016/j.earlhumdev.2014.08.023.
18. Lipsitch M, Tchetgen ET, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21(3):383.
19. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. Eur J Cardio-thoracic Surg. 2018;53(6):1112–7. https://doi.org/10.1093/ejcts/ezy167.
20. McQuire C, de Vocht F. Methodological advances to mitigate some of the challenges of research on alcohol and all-cause mortality: commentary on Rehm. Drug Alcohol Rev. 2019;38(1):7–8. https://doi.org/10.1111/dar.12898.

21. Watkins TR. Understanding uncertainty and bias to improve causal inference in health intervention research. [PhD thesis]. Sydney, Australia: The University of New South Wales; Published online 2019.
22. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. Int J Epidemiol. 2017;46(2):756–62. https://doi.org/10.1093/ije/dyw323.
23. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. BMJ. 2017;359(October). https://doi.org/10.1136/bmj.j4587.
24. Hernán MA. Robins JM. What If. Chapman & Hill/CRC: Causal Inference; 2020.
25. Williamson T, Ravani P. Marginal structural models in clinical research: When and how to use them? Nephrol Dial Transplant. 2017;32(February):ii84-ii90. doi:https://doi.org/10.1093/ndt/gfw341
26. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. Am J Epidemiol. 2017;185(1):65–73. https://doi.org/10.1093/aje/kww165.
27. Gunasekara FI, Richardson K, Carter K, Blakely T. Fixed effects analysis of repeated measures data. Int J Epidemiol. 2014;43(December 2013):264–9. https://doi.org/10.1093/ije/dyt221.
28. Fergusson DM, Boden JM, Horwood LJ. Psychosocial sequelae of cannabis use and implications for policy : findings from the Christchurch health and development study. Soc Psychiatry Psychiatr Epidemiol. 2015;50(9):1317–26. https://doi.org/10.1007/s00127-015-1070-x.
29. Fergusson DM, Boden JM, Horwood LJ. Tests of causal links between alcohol abuse or dependence and major depression. Arch Gen Psychiatry. 2009;66(3):260–6.
30. Fergusson DM, Boden JM, Horwood LJ. Unemployment and suicidal behavior in a New Zealand birth cohort: a fixed effects regression analysis. Crisis. 2007;28(2):95–101.
31. Allison PD. Fixed effects regression methods for longitudinal data using SAS. SAS Institute Inc; 2005.
32. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol. 2013;42(5):1511–9. https://doi.org/10.1093/ije/dyt127.
33. Pearce N, Lawlor DA. Causal inference — so much more than statistics. Int J Epidemiol. 2017;45(6):1895–903. https://doi.org/10.1093/ije/dyw328.
34. Vanderweele TJ. Mediation Analysis: A Practitioner ' s Guide. Annu Rev Public Health. 2016;37(17–32). doi:https://doi.org/10.1146/annurev-publhealth-032315-021402
35. Dunning T. Natural experiments in the social sciences: a design-based approach: Cambridge University Press; 2012.
36. Gage SH, Munafo MR, Smith GD. Causal inference in developmental origins of health and disease (DOHaD) research. Annu Rev Psychol. 2016;67:567–85. https://doi.org/10.1146/annurev-psych-122414-033352.
37. Davies NM, Holmes M V, Smith GD. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ. 2018;362. doi:https://doi.org/10.1136/bmj.k601
38. Visontay R, Sunderland M, Slade T, Wilson J, Mewton L. Are there nonlinear relationships between alcohol consumption and long-term health? Protocol for a systematic review of observational studies employing approaches to improve causal inference. BMJ Open. 2021;11(3):e043985.
39. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm.
40. Mamluk L, Jones T, Ijaz S, et al. Evidence of detrimental effects of prenatal alcohol exposure on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies. Int J ofEpidemiology, Published online. 2020:1–24. https://doi.org/10.1093/ije/dyz272.
41. Mamluk L, Edwards HB, Savovi J, et al. Low alcohol consumption and pregnancy and childhood outcomes: time to change guidelines indicating apparently ' safe ' levels of alcohol during pregnancy ? A systematic review and meta-analyses. BMJ Open. 2017;7:e015410. https://doi.org/10.1136/bmjopen-2016-015410.
42. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097.
43. Dickerman BA, Coseo S, Markku M, Pukkala E, Mucci LA, Kaprio J. Alcohol intake, drinking patterns, and prostate cancer risk and mortality : a 30-year prospective cohort study of Finnish twins. Cancer Causes Control. 2016;27(9):1049–58. https://doi.org/10.1007/s10552-016-0778-6.
44. Carlsson S, Hammar N, Grill V, Kaprio J. Alcohol consumption and the incidence of type 2 diabetes a 20-year follow-up of the Finnish twin cohort study. Diabetes Care. 2003;26(10):2–7.
45. Peng M, Zhang J, Zeng T, et al. Alcohol consumption and diabetes risk in a Chinese population: a Mendelian randomization analysis. Addiction. 2019;114(3):436–49.
46. Handing EP, Andel R, Kadlecova P, Gatz M, Pedersen NL. Midlife alcohol consumption and risk of dementia over 43 years of follow-Up : a population-based study from the Swedish twin registry. J Gerontol Ser A Biomed Sci Med Sci. 2015;70(10):1248–54. https://doi.org/10.1093/gerona/glv038.
47. Gémes K, Forsell Y, Janszky I, et al. Moderate alcohol consumption and depression-a longitudinal population-based study in Sweden. Acta Psychiatr Scand. Published online 2019.
48. Samuelsson Å, Ropponen A, Alexanderson K, Svedberg P. A prospective cohort study of disability pension due to mental diagnoses : the importance of health factors and behaviors. BMC Public Health. 2013;13(1):1. https://doi.org/10.1186/1471-2458-13-621.
49. Ilomäki J, Hajat A, Kauhanen J, et al. Relationship between alcohol consumption and myocardial infarction among ageing men using a marginal structural model. Eur J Pub Health. 2012;22(6):825–30. https://doi.org/10.1093/eurpub/ckr013.
50. Kadlecová P, Andel R, Mikulík R, Handing EP, Pedersen NL. Alcohol consumption at midlife and risk of stroke during 43 years of follow-up: cohort and twin analyses. Stroke. 2015;46(3):627–33. https://doi.org/10.1161/STROKEAHA.114.006724.
51. Millwood IY, Walters RG, Mei XW, et al. Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. Lancet. 2019;393(10183):1831–42.
52. Ropponen A, Svedberg P. Single and additive effects of health behaviours on the risk for disability pensions among Swedish twins. Eur J Pub Health. 2013;24(4):643–8. https://doi.org/10.1093/eurpub/ckt168.
53. Silverwood RJ, Holmes MV, Dale CE, et al. Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. Int J Epidemiol. 2014;43(6):1781–90. https://doi.org/10.1093/ije/dyu187.
54. Vu KN, Ballantyne CM, Hoogeveen RC, Nambi V. Causal role of alcohol consumption in an improved lipid Profile : the atherosclerosis risk in communities (ARIC) study. PLoS One. 2016;11(2):e0148765. https://doi.org/10.1371/journal.pone.0148765.
55. Sipilä P, Rose RJ, Kaprio J. Drinking and mortality : long-term follow-up of drinking- discordant twin pairs. Addiction. 2016;111(2):245–54. https://doi.org/10.1111/add.13152.
56. Sander PM, Cole SR, Stall RD, et al. Joint effects of alcohol consumption and high-risk sexual behavior on HIV seroconversion among men who have sex with men. AIDS. 2013;27(5):815–23. https://doi.org/10.1097/QAD.0b013e32835cff4b.
57. Pietikäinen S, Silventoinen K, Svedberg P, et al. Health-related and Sociodemographic risk factors for disability pension due to low Back disorders: a 30-year prospective Finnish twin cohort study. J Occup Environ Med. 2011;53(5):488–96. https://doi.org/10.1097/JOM.0b013e31821576dd.
58. Ropponen A, Silventoinen K, Svedberg P, et al. Health-related risk factors for disability pensions due to musculoskeletal diagnoses : a 30-year Finnish twin cohort study. Scand J Public Health. 2011;39(8):839–48. https://doi.org/10.1177/1403494811418283.
59. Rehm J. Why the relationship between level of alcohol-use and all-cause mortality cannot be addressed with meta-analyses of cohort studies. Drug Alcohol Rev. 2019;38(1):3–4. https://doi.org/10.1111/dar.12866.
60. Zhao J, Stockwell T, Roemer A, Chikritzhs T. Is alcohol consumption a risk factor for prostate cancer ? A systematic review and meta – analysis. BMC Cancer. 2016;16(1):1–13. https://doi.org/10.1186/s12885-016-2891-z.
61. Taghdiri S. Association of Alcohol Consumption with Glucose Homeostasis : A Systematic Review and Meta-Analysis. [PhD thesis]. Toronto, Canada: University of Toronto; Published online 2017.
62. Roerecke M, Tobe SW, Kaczorowski J, et al. Sex-specific associations between alcohol consumption and incidence of hypertension: a systematic review and meta-analysis of cohort studies. J Am Heart Assoc. 2018;7(13):e008202.
63. Sayon-Orea C, Martinez-gonzalez MA, Bes-rastrollo M. Alcohol consumption and body weight : a systematic review. Nutr Rev. 2011;69(8):419–31. https://doi.org/10.1111/j.1753-4887.2011.00403.x.

64. Minzer S, Losno RA, Casas R. The effect of alcohol on cardiovascular risk factors: is there new Information ? Nutrients. 2020;12(4):1–22.
65. Larsson SC, Wallin A, Wolk A, Markus HS. Differing association of alcohol consumption with different stroke types : a systematic review and meta-analysis. BMC Med. 2016;14(1). doi:https://doi.org/10.1186/s12916-016-0721-4
66. Yang Y, Liu D-C, Wang Q-M, et al. Alcohol consumption and risk of coronary artery disease : a dose-response meta-analysis of prospective studies. Nutrition. 2016;32(6):637–44. https://doi.org/10.1016/j.nut.2015.11.013.
67. Visontay R, Rao RT, Mewton L. Alcohol use and dementia: new research directions. Curr Opin Psychiatry. 2021;34(2):165–70. https://doi.org/10.1097/yco.0000000000000679.
68. Li J, Wang H, Li M, et al. Effect of alcohol use disorders and alcohol intake on the risk of subsequent depressive symptoms: a systematic review and meta-analysis of cohort studies. Addiction. 2020;115(7):1224–43. https://doi.org/10.1111/add.14935.
69. Rehm J, Probst C, Shield KD, Shuper PA. Does alcohol use have a causal effect on HIV incidence and disease progression ? A review of the literature and a modeling strategy for quantifying the effect. Popul Health Metrics. 2017;15(1):1–7. https://doi.org/10.1186/s12963-017-0121-9.
70. Shuper PA, Neuman M, Kanteres F, Baliunas D, Joharchi N, Rehm J. Causal considerations on alcohol and HIV / AIDS — a systematic review. Alcohol Alcohol. 2010;45(2):159–66. https://doi.org/10.1093/alcalc/agp091.
71. Mukamal KJ, Stampfer MJ, Rimm EB. Genetic instrumental variable analysis : time to call mendelian randomization what it is . The example of alcohol and cardiovascular disease. Eur J Epidemiol. 2020;35(2):93–7.
72. Davey G, Michael S, Davies NM, Ebrahim S. Mendel ' s laws , Mendelian randomization and causal inference in observational data : substantive and nomenclatural issues. Eur J Epidemiol. 2020;35(2):99.
73. Ellison RC, Grønbæk M, Skovenborg E. Using Mendelian randomization to evaluate the effects of alcohol consumption on the risk of coronary heart disease. Drugs and Alcohol Today. Published online 2021.
74. Roerecke M, Rehm J. Alcohol consumption, drinking patterns, and ischemic heart disease: a narrative review of meta-analyses and a systematic review and meta-analysis of the impact of heavy drinking occasions on risk for moderate drinkers. BMC Med. 2014;12(1):182.
75. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing transparency through a multiverse analysis. Perspect Psychol Sci. 2016;11(5):702–12. https://doi.org/10.1177/1745691616658637.
76. Vanderweele TJ, Ding P. Sensitivity analysis in observational Research : introducing the E-value. Ann Intern Med. 2017;167(4):268–74. https://doi.org/10.7326/M16-2607.
77. Sherk A, Gilmore W, Churchill S, Lensvelt E, Stockwell T, Chikritzhs T. Implications of cardioprotective assumptions for national drinking guidelines and alcohol harm monitoring systems. Int J Environ Res Public Health. 2019;16(24):1–17. https://doi.org/10.3390/ijerph16244956.
78. Stockwell T, Room R. Constructing and responding to low-risk drinking guidelines: conceptualisation, evidence and reception. Drug Alcohol Rev. 2012;31(2):121–5. https://doi.org/10.1111/j.1465-3362.2011.00416.x.

Publisher's Note

Appendix B

Protocol paper for chapter 2

Preface

The protocol paper for Chapter 2 was published as **Visontay, R.**, Sunderland, M., Slade, T., Wilson, J., & Mewton, L. (2021). Are there non-linear relationships between alcohol consumption and long-term health?: Protocol for a systematic review of observational studies employing approaches to improve causal inference. *BMJ open, 11*(3), e043985**.**

RV conceptualised the study with assistance from MS, TS, and LM. All authors critically revised the manuscript and approved the final version.

BMJ Open

Are there non-linear relationships between alcohol consumption and long-term health? Protocol for a systematic review of observational studies employing approaches to improve causal inference

Rachel Visontay ,[1] Matthew Sunderland,[1] Tim Slade,[1] Jack Wilson,[1] Louise Mewton[1,2]

To cite: Visontay R, Sunderland M, Slade T, *et al*. Are there non-linear relationships between alcohol consumption and long-term health? Protocol for a systematic review of observational studies employing approaches to improve causal inference. *BMJ Open* 2021;**11**:e043985. doi:10.1136/bmjopen-2020-043985

► Prepublication history and additional material for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-043985).

Received 20 August 2020
Revised 07 February 2021
Accepted 18 February 2021

[1]The Matilda Centre for Research in Mental Health and Substance Use, The University of Sydney, Sydney, New South Wales, Australia
[2]Centre for Healthy Brain Ageing, University of New South Wales, Sydney, New South Wales, Australia

Correspondence to
Dr Louise Mewton;
louisem@unsw.edu.au

ABSTRACT

Introduction There is a substantial literature finding that moderate alcohol consumption is protective against certain health conditions. However, more recent research has highlighted the possibility that these findings are methodological artefacts, caused by confounding and other biases. While modern analytical and study design approaches can mitigate confounding and thus enhance causal inference in observational studies, they are not routinely applied in research assessing the relationship between alcohol use and long-term health outcomes. The purpose of this systematic review is to identify observational studies that employ these analytical/design-based approaches in assessing whether relationships between alcohol consumption and health outcomes are non-linear. This review seeks to evaluate, on a per-outcome basis, what these studies find the strength and form of the relationship between alcohol consumption and health to be.

Methods and analysis Electronic databases (MEDLINE, PsycINFO, Embase and SCOPUS) were searched in May 2020. Study selection will comply with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. Articles will be screened against eligibility criteria intended to capture studies using observational data to assess the relationship between varying levels of alcohol exposure and any long-term health outcome (actual or surrogate), and that have employed at least one of the prespecified approaches to enhancing causal inference. Risk of bias of included articles will be assessed using study design-specific tools. A narrative synthesis of the results is planned.

Ethics and dissemination Formal ethics approval is not required given there will be no primary data collection. The results of the study will be disseminated through published manuscripts, conferences and seminar presentations.

PROSPERO registration number CRD42020185861.

Strengths and limitations of this study

- This systematic review will be the first to identify those observational studies that best promote causal inference when assessing the relationship between alcohol and health outcomes, promising an important contribution to the literature.
- A strength is that we will be searching for a broad, comprehensive range of analytical and design-based approaches to improving causal inference.
- A further strength is the examination of a broad range of long-term health outcomes.
- Given considerable differences among health outcomes and methodological approaches, analysis will be limited to narrative synthesis.

INTRODUCTION

Despite the significant contribution of alcohol to the burden of disease worldwide,[1] there is a substantial literature finding that low-to-moderate alcohol consumption is associated with improved health for a wide range of outcomes, including cardiovascular,[2] cognitive,[3] diabetic[4] and mental health[5] conditions, as well as mortality itself.[6] Here, a 'J-shaped' relationship is common, where moderate consumption represents the nadir for risk, compared with a somewhat elevated risk for those who abstain from alcohol and a greatly increased risk for heavy alcohol drinkers.

However, concerns with this literature have been raised, such as findings of protection against certain outcomes (eg, cirrhosis of the liver) that currently have no known plausible biological mechanism.[7] Such criticisms have prompted scrutiny of biases in observational studies, raising the possibility that these J-shaped relationships are simply

methodological artefacts.[7 8] Methodological limitations in this literature include confounding and reverse causality—issues that plague epidemiological research more generally in efforts to identify and estimate causal relationships—as well as selection biases and measurement errors more specific to this research question.[8–10]

Many of these biases would be obviated by randomised controlled trials (RCT) in which participants are randomly assigned to different consumption levels and followed up on health outcomes. But due to practical and ethical considerations, no long-term RCTs have been conducted. In short-term interventional studies, moderate alcohol consumption has been found to promote beneficial changes in several biomarkers of cardiovascular health.[11] Similarly, a 2-year RCT of diabetics found moderate wine consumption led to reduced cardiometabolic risk.[12] So, while this provides some evidence for the benefits of moderate alcohol consumption, such benefits may only be transitory, and harms not yet apparent. There is also a dearth of RCTs investigating a reduction in alcohol consumption to abstention specifically,[13] particularly in those who are moderate consumers. Those reduction trials that have been conducted tend to be short-term and have limited applicability to the research question at hand. Alcohol consumption can have complex cumulative and or/delayed effects[4]; without long-term follow-up, these cannot be observed.

Instead, recent efforts have focused on mitigating biases in observational designs, with a focus on the impact of methodological decisions regarding data collection and analysis. There is acknowledgement that such decisions—for example, which covariates to measure, how exposure levels are defined and compared, which population to sample, what model to use—can have substantial effects on conclusions about the strength and form of a relationship.[14 15] Tools available in the study planning phase can improve some of these methodological decisions. For example, causal diagrams can be created to identify confounders as well as possible selection bias. Steps implemented post-main analyses can help to determine the robustness of findings across alternative methodological decisions ('sensitivity analyses'), or in the face of potential unmeasured bias ('bias analysis'). Additional checks for robustness across populations with different underlying confounding structures ('cross-cohort comparison') or across different methodological approaches ('triangulation') can also strengthen causal conclusions.[16]

Modern epidemiological developments in data analysis and alternative study designs, however, may have the most potential to mitigate the limitations of existing research, particularly in addressing confounding. Causal inference in conventional observational designs (eg, prospective cohort studies) can be enhanced with modern data analysis methods. These include propensity scores (estimating the exposure mechanism as a data preprocessing step prior to main analyses) used primarily for matching or inverse probability weighting, and 'G-methods' (which can account for time-varying variables). Doubly robust methods incorporate estimation of both the exposure and outcome mechanisms, so only require correct specification of one of these models to produce accurate estimates of an exposure's effect.[17] Importantly, these methods still rely on all relevant confounders being measured, and measured without error (ie, residual confounding is still possible). Other analytical tools that may enhance causal inference include fixed effects regression (focusing exclusively on within-subject variation, thus avoiding confounding) and causal mediation (identifying the mechanism/s through which an exposure may be exerting a causal effect).

Alternate observational study designs represent another approach, often eschewing the need to accurately identify and measure confounders altogether. Natural experiments, including those based on instrumental variables, mimic the random allocation of an RCT, bypassing the problem of confounding, and guarding against reverse causation. Genetic instrumental variable designs such as Mendelian randomisation (MR) show particular promise because of their ability to find proxies for alcohol consumption, and are increasingly popular in this research area. Family based designs can also mitigate unmeasured confounding (to the extent to which relevant covariates are shared), as can negative control designs.[18] Table 1 provides an overview of these analytical and alternative design methods for improving causal inference.

As other sources of bias are addressed with approaches that consist largely of considered, literature-informed researcher decisions, they do not lend themselves to terminology for a systematic database search and are not the primary focus of this review. However, some of the included methods of interest do mitigate other biases in addition to confounding, for example, MR counters reverse causation and measurement error.

Rationale

The functional forms of many of the relationships between alcohol use and specific health outcomes are currently unclear. Several recent pooled analyses and reviews have failed to provide evidence for J-shaped relationships, finding monotonically increasing dose-response relationships with most health outcomes, including most, if not all, cancers.[1 19] Yet certain cardiovascular and diabetic conditions are consistent exceptions,[1 19] and new, rigorous individual studies that consider the impact of methodological decisions continue to find J-shaped relationships for other outcomes, for example, mortality.[6] Whether these findings reflect truly causal relationships remains an open question.

The state of the evidence base may be due to the fact that, while some of the discussed analytical and design-based methods are becoming more popular in health research,[20] they are not frequently employed in this specific research area (with calls for greater implementation[21]). Despite the many reviews focusing on the relationship between alcohol use and various health outcomes, analytical methods to

Table 1 Description of methods to enhance causal inference of interest for this systematic review

Method	Relevant submethods	Description
Analytical methods applied to traditional longitudinal study designs		
Propensity scores (PS)[33 34]	Covariate balancing propensity scores	► The PS is a single value reflecting the probability of exposure for an individual given their values on all relevant covariates. ► PS generation occurs as a data 'preprocessing' step prior to main analysis. ► Usually generated via logistic regression. ► Once generated, the PS can be used for matching, stratification, weighting (using inverse probability of treatment weights) or as a covariate for adjustment in regression.
G-methods[20 35–37]		► A family of methods intended for use with time-dependent variables. ► Developed as a solution to the problem of time-varying covariates affected by past exposure, including those that act as both confounders and mediators over time. ► The three G-methods are the G-formula, marginal structural models and G-estimation, each relying on its own modelling assumptions.
	G-formula (aka G-computation or G-standardisation)[17 20 35 38]	► First models relationship given observed data (using actual exposure for each individual), and then predicts outcomes under counterfactual exposures, with the difference taken as the causal effect. ► Is a generalisation of standardisation (conditioning on covariates and then marginalising) that accounts for dynamic variables by considering covariate distribution over follow-up time.
	Marginal structural models (MSMs)[20 35 38]	► Use weights based on inverse probability of exposure at each time point to create a pseudo-population, where each combination of covariates is equally present in each exposure condition. ► Using these weights, MSMs then estimate the causal effect. ► The most popular of the G-methods.
	G-estimation of structural nested models[35 37 38]	► At each wave assesses the relationship between exposure and likelihood of outcome given covariates, adjusting for exposure and covariate values from past waves, thus accounting for dynamic confounders affected by past exposure. ► Considered semi-parametric in that mean counterfactual outcomes under no exposure are unspecified.
Doubly robust methods[17]	Targeted maximum likelihood estimation Augmented inverse probability weighting	► Incorporates both an estimation of the outcome mechanism (as in G-formula or regression adjustment) and the exposure mechanism (as in propensity scores).
Fixed effects regression[39–41]		► A technique developed in the econometrics literature for use with longitudinal data with repeat outcome measurements, only using information on within-subject variation, thus controlling for all time-invariant sources of confounding. ► Treats time-invariant characteristics that differ between individuals as fixed parameters (unlike in mixed models), allowing estimation of parameters of interest net of stable confounders. ► Each participant serves as own control.

Continued

Table 1 Continued

Method	Relevant submethods	Description
Causal mediation analysis[42–44]		► Integrates traditional mediation analysis (which separately estimates total effect of exposure on outcome, indirect effect via mediators and direct effect unexplained by mediators) with the potential outcomes framework to allow for exposure-mediator interaction and non-linear relationships (ie, is a non-parametric method). ► Uses the concepts of 'controlled direct effect', 'natural direct effect' and 'natural indirect effect'. ► Makes explicit underlying assumptions related to unmeasured confounding, and encourages sensitivity analyses to test robustness to assumption violations.
Alternative observational study designs		
Natural experiments[16 45–47]		► Mimic randomised controlled trials by exploiting exogenous events that are truly randomised/approximate random assignment. ► Differ from true experiments in that exposure is not assigned by the researcher. ► Assignment may be as a result of naturally occurring phenomena (eg, a weather event), or of human intervention implemented for reasons other than the research question (eg, army draft lottery).
	Standard natural experiments	► Natural experiments where individuals are as-if/randomly assigned to exposure and control groups.
	Instrumental variable analysis	► Assesses the relationship between an as-if/ randomly assigned *proxy* for the exposure of interest and the outcome. ► A valid instrumental variable must be associated with the exposure of interest, be independent of confounders of the exposure-outcome relationship and should affect the outcome only via the exposure.
	Genetic instrumental variables	► Subset of instrumental variable analysis using genetic variants as proxies for exposure. ► The most prominent technique is Mendelian Randomisation.
Quasi-experiments[45]		► Like natural experiments, exploit exogenous events to assess relationships between exposures and outcomes, but lack random or as-if random assignment.
Family based designs[16 46]	Twin studies Sibling comparison	► By comparing genetically related participants discordant for the exposure of interest, accounts for confounding from genetic or shared environmental sources.
Negative controls[18 46]	Negative control exposures Negative control outcomes	► Have the same confounding structures as the exposure-outcome relationship of interest, but lack a plausible causal mechanism. ► If association is greater for the relationship of interest than for the negative control, a causal relationship is likely; if not, suggests confounding/other shared biases responsible. ► May take the form of a negative control exposure or a negative control outcome.

counter confounding are rarely a focus. Similarly, reviews of observational research tend to search exclusively for studies with traditional designs (eg, cohort studies), meaning findings from novel study designs are not integrated. As such, the planned review, focusing on those studies that best promote causal inference by employing methods that mitigate confounding, comprises an original and important contribution.

Objective

The objective of this review is therefore to identify all observational studies that have employed the aforementioned

analytical or design-based approaches (see table 1), and to synthesise their findings on the strength and functional form (ie, support for non-linearity) of the relationships between levels of alcohol consumption and various long-term health outcomes (or surrogates) in humans.

METHODS

This protocol complies with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P) statement,[22] the checklist for which can be found in online supplemental table 1. This protocol has been registered with the PROSPERO International Prospective Register of Systematic Reviews of the University of York (CRD42020185861). Reporting of the systematic review will be informed by PRISMA (online supplemental material 1[23]).

Search strategy

Electronic databases (MEDLINE, PsycINFO, Embase and Scopus) were searched in May 2020 for peer-reviewed, English-language journal articles, with no restrictions on date range. Relevant grey literature returned by Scopus will also be considered. The search strategy was designed to capture all studies using observational data to assess the relationship between varying levels of alcohol exposure and any health outcome, and that have employed at least one of the prespecified approaches to improving causal inference.

Searches ran a combination of MeSH or other controlled vocabulary terms as well as free-text words. Five groups of terms are used in the search strategy, relating to: (1) alcohol; (2) levels or patterns of drinking; (3) observational, longitudinal studies; (4) approaches to improve causal inference that are used in conjunction with observational, longitudinal study designs and (5) approaches to improve causal inference that may be considered their own study designs. These concepts are combined with the following logic: 1 and 2 and ((3 and 4) or 5). Search terms were adapted from recent systematic reviews and from keywords/indexing of a set of key eligible papers known to the study authors. These terms were iteratively refined to improve sensitivity by cross-checking results against the set of key papers. The MEDLINE search terms are provided in online supplemental table 2.

Cross-checking of reference lists of relevant, retrieved publications will be used to supplement the database searches.

Eligibility criteria

Eligibility criteria for this review are spread across population, exposure, comparator, outcome and study design components of the publications under consideration (PI/ECOS).

Population

Only human studies will be eligible for inclusion. While there are no other restrictions on eligible populations, alcohol consumption must be assessed in the same individuals to which the health outcome/s accrue (ie, the relationship between an individual's alcohol consumption and the health of others is beyond the scope of this review).

Exposure

The exposure of interest is level of alcohol consumption (volume over a certain period) or level and pattern of consumption (incorporating information on frequency or presence of heavy episodic drinking). Studies must have discretised level of alcohol consumption into categories. This is typical practice for this research question, as analyses involving comparisons between such groups do not make assumptions about functional form (unlike analyses of continuous predictors). Specifically, there must be a non-drinking reference group and at least two other levels of consumption (or alternative structure in instrumental variables designs allowing for tests of non-linear relationships). While it is anticipated that the construction of these categories will vary (eg, different criteria for volume/frequency of consumption, differential use of lifetime and current consumption), all will be eligible for inclusion. Differences between studies in drinking category composition, such as whether former drinkers are categorised separately from lifetime abstainers, will be considered during data synthesis.

Comparator

Reference groups should comprise abstainers/non-drinkers, but reference groups partly or solely comprised occasional drinkers will also be accepted. Non-drinker reference groups that include former drinkers will also be accepted. Differences between studies in reference group composition will be considered during data synthesis.

Outcomes

Studies on the association between alcohol consumption and health outcomes have reported J-shaped relationships for a wide range of long-term conditions and their surrogates (eg, biomarkers). The focus of this review is to examine how approaches to improving causal inference in observational research have been used across the entirety of the literature, regardless of outcome type. Given common targets of study in the broader literature, it is anticipated that papers included in this review will include the following outcomes: cancer (and subtypes), cardiovascular disease (and subtypes), cognition and dementia, mental health conditions, diabetes and cause-specific or all-cause mortality, among others. Short-term or acute conditions such as injury, acute exacerbation of mental illness, sexually transmitted diseases, alcohol-drug interactions and sexual function will be excluded.

Study design

Observational studies assessing the relationship between varying levels of alcohol consumption and any health outcome will be eligible. Comparisons between consumption categories must have allowed for the detection of

non-linear relationships and must have employed at least one of the prespecified analytical/design-based approaches. Incorporating feedback from academics with expertise in causal inference methods, the approaches of explicit interest chosen for this review have been clearly defined a priori and have been recognised and employed in recent years for their potential to enhance causal inference. Specifically, these are: propensity scores, the G family of methods, targeted maximum likelihood estimation and other doubly robust methods, fixed effects regression, causal mediation analysis, standard natural experiments, MR and other instrumental variables, quasi-experiments, family-based methods and negative controls.

Conventional observational studies employing analytical methods to promote causal inference must be cohort or case-control designs and must be longitudinal, with alcohol consumption measured at least once at a time point preceding measurement of health outcome/s (studies employing MR/other natural experiment approaches will be eligible even if not longitudinal). Systematic reviews/meta-analyses, interventional studies and animal studies will be excluded.

Study selection

All citations will be imported into Endnote[24] and deduplicated. Citations will then be uploaded to Covidence,[25] which will be used to facilitate screening. Titles and abstracts returned by the database searches (and any additional papers identified by hand) will be screened independently by one reviewer (RV), with a second reviewer (JW) to screen a random selection of 25% of the titles and abstracts. Full-text articles of potentially eligible articles will be independently assessed by two reviewers (RV and JW). If at any stage discrepancies cannot be resolved through discussion and consensus, a third reviewer (LM) will be consulted.

Data extraction

A standardised prepiloted data extraction form will be used, with extraction performed independently by two reviewers (RV and JW). Study authors will be contacted if further information is required. Data extracted will include:

1. *Study information*: author/s, year of baseline data collection and year of publication.
2. *Participant characteristics*: sample size, study setting, mean age at baseline, relevant eligibility criteria, country/ies of data collection, cohort name (if applicable).
3. *Exposure/comparator characteristics*: number and spread of measurement occasions, number and nature of drinking consumption/abstention categories.
4. *Study design and analysis*: approach/es of interest used (ie, study design of interest, combination of traditional study design and analytical method of interest), any additional statistical tests employed, any additional measures taken to enhance causal inference (ie, causal diagrams or tests of robustness), covariates/confounders included in models or otherwise accounted for.
5. *Outcomes*: health outcome/s assessed and whether these are binary or continuous.
6. *Results*: findings on the strength and form of the relationship between alcohol and health.

Additional information extracted from MR studies will include: whether there is one study population or pooled data, use of one or two sample MR, instrument used and variance in alcohol consumption explained by the instrument, method of MR analysis used and how it is capable of revealing potential non-linearity and whether there is testing of MR assumptions. Additional information extracted from twin studies will include: whether twins are discordant for exposure or outcome, and whether monozygotic twins are compared with dizygotic twins.

Where both conventional and novel (causal inference method of interest) analyses are conducted, information on both will be extracted. Again, if at any stage discrepancies cannot be resolved through discussion and consensus, a third reviewer (LM) will be consulted.

Risk of bias assessment

Assessment of bias depends on study design and methodology. As a range of designs and methodologies are expected in the included studies, several existing risk of bias assessment tools will be used. Quality of cohort and case-control studies (including twin cohorts) will be assessed using the relevant Newcastle-Ottawa Scales.[26] Recently developed risk of bias tools specific to MR studies, natural experiments and family based methods[27] will be used and adapted where appropriate. One reviewer (RV) will apply the assessment tools to all included studies, with a second reviewer (JW) assessing a random 25% of the studies to establish scoring accuracy.

Data synthesis

Descriptive narrative synthesis will be performed. The relationships between alcohol and different health outcomes vary due to mediating causal mechanisms, meaning they are not comparable. As such, in addressing what the included studies find on the relationship between levels of consumption and health, synthesis will be conducted on a per-outcome basis. The heterogeneity of the different analytical approaches means they are not suited to quantitative synthesis. Convergence/lack thereof of findings from each of the various causal inference methods will be assessed in light of the advantages and limitations of each methodology. In evaluating the literature on the effects of moderate alcohol consumption as a whole, the disease burden of each health outcome will be considered.

Assessment of evidence quality

Popular tools to assess the body of evidence gathered, such as Grading of Recommendations, Assessment, Development and Evaluation (GRADE[28]), are inappropriate here. Given initial 'low quality' ratings are automatically assigned to observational studies, GRADE is of limited use for exposures that do not lend themselves to RCTs[29 30] (including alcohol consumption[31]). Additionally, these

tools are intended to rate *bodies* of evidence—such ratings are ill-suited to this review given it is intentionally restricted to a subset of the literature. Consistent with recent reviews focusing on observational studies employing novel causal inference approaches, this review will limit formal assessment of evidence quality to risk of bias.[26 32]

Patient and public involvement

As this is a review of existing published studies, there will be no patient or public involvement in this research.

Ethics and dissemination

Formal ethics approval is not required given there will be no primary data collection. The results of the study will be disseminated through published manuscripts, conferences and seminar presentations.

Acknowledgements The authors would like to thank Jessica Hughes, librarian at the University of Sydney, for assisting with the literature search.

Contributors RV is the guarantor of the review. RV, MS, TS and LM conceptualised the study and developed the study design and protocol. RV wrote the first draft of the manuscript. RV, MS, TS, JW and LM contributed to, read, revised and approved the final manuscript.

Funding RV is funded by the National Health and Medical Research Council (GNT1190255) and the University of Sydney.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Rachel Visontay http://orcid.org/0000-0002-0693-1012

REFERENCES

1 GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet* 2018;392:1015–35.
2 Ronksley PE, Brien SE, Turner BJ, *et al*. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ* 2011;342:d671.
3 Rehm J, Hasan OSM, Black SE, *et al*. Alcohol use and dementia: a systematic scoping review. *Alzheimers Res Ther* 2019;11:1.
4 Kerr WC, Ye Y, Williams E, *et al*. Lifetime alcohol use patterns and risk of diabetes onset in the National alcohol survey. *Alcohol Clin Exp Res* 2019;43:262–9.
5 Gémes K, Forsell Y, Janszky I, *et al*. Moderate alcohol consumption and depression - a longitudinal population-based study in Sweden. *Acta Psychiatr Scand* 2019;139:526–35.
6 Keyes KM, Calvo E, Ornstein KA, *et al*. Alcohol consumption in later life and mortality in the United States: results from 9 waves of the health and retirement study. *Alcohol Clin Exp Res* 2019;43:1734–46.
7 Andreasson S, Chikritzhs T, Dangardt FHH. Evidence about health effect of "moderate" alcohol consumption: Reasons for scepticism and public health implications. In: Alcohol and Society: IOGT–NTO and Swedish Society of MedicineEvidence about health effects of “moderate” alcohol consumption: Reasons for scepticism and public health implilcations. *Alcohol Soc* 2014;6.
8 Naimi TS, Stockwell T, Zhao J, *et al*. Selection biases in observational studies affect associations between 'moderate' alcohol consumption and mortality. *Addiction* 2017;112:207–14.
9 Callinan S, Chikritzhs T, Livingston M. Consistency of drinker status over time: drinking patterns of ex-drinkers who describe themselves as lifetime abstainers. *J Stud Alcohol Drugs* 2019;80:552–6.
10 Shaper AG, Wannamethee G, Walker M. Alcohol and mortality in British men: explaining the U-shaped curve. *Lancet* 1988;2:1267–73.
11 Brien SE, Ronksley PE, Turner BJ, *et al*. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *BMJ* 2011;342:d636.
12 Gepner Y, Golan R, Harman-Boehm I, *et al*. Effects of initiating moderate alcohol intake on cardiometabolic risk in adults with type 2 diabetes: a 2-year randomized, controlled trial. *Ann Intern Med* 2015;163:569–79.
13 Field M, Puddephatt J-A, Goodwin L, *et al*. Benefits of temporary alcohol restriction: a feasibility randomized trial. *Pilot Feasibility Stud* 2020;6:1–15.
14 Chu L, Ioannidis JPA, Egilman AC, *et al*. Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk. *Int J Epidemiol* 2020;49:608–18.
15 Stockwell T, Zhao J, Panwar S, *et al*. Do "Moderate" Drinkers Have Reduced Mortality Risk? A Systematic Review and Meta-Analysis of Alcohol Consumption and All-Cause Mortality. *J Stud Alcohol Drugs* 2016;77:185–98.
16 Richmond RC, Al-Amin A, Smith GD, *et al*. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. *Early Hum Dev* 2014;90:769–80.
17 Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol* 2017;185:65–73.
18 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;21:383.
19 Rehm J, Gmel GE, Gmel G, *et al*. The relationship between different dimensions of alcohol use and the burden of disease-an update. *Addiction* 2017;112:968–1001.
20 Watkins TR. Understanding uncertainty and bias to improve causal inference in health intervention research 2019.
21 McQuire C, de Vocht F. Methodological advances to mitigate some of the challenges of research on alcohol and all-cause mortality: commentary on Rehm. *Drug Alcohol Rev* 2019;38:7–8.
22 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
23 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
24 Clarivate Analytics. Endnote X8, 2016
25 Veritas Health Innovation. Covidence systematic review software. Available: www.covidence.org
26 Wells GA, Shea B, O’Connell D. The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm
27 Mamluk L, Jones T, Ijaz S, *et al*. Evidence of detrimental effects of prenatal alcohol exposure on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies. *Int J Epidemiol* 2021;49:1972–95.
28 Schünemann H, Brożek J, Guyatt G. Grade Handbook for grading quality of evidence and strength of recommendations. updated October 2013. The grade Working group, 2013 2013.
29 Arroyave WD, Mehta SS, Guha N, *et al*. Challenges and recommendations on the conduct of systematic reviews of observational epidemiologic studies in environmental and occupational health. *J Expo Sci Environ Epidemiol* 2021;31:21–30.
30 World Health Organization. Environmental noise guidelines for the European region World Health organization. regional office for Europe 2018.

31 NHMRC Clinical Trials Centre at the University of Sydney. Evaluating the evidence on the health effects of alcohol consumption: evidence evaluation report. Sydney NHMRC Clinical Trials Centre; 2019.
32 Mamluk L, Edwards HB, Savović J, *et al*. Low alcohol consumption and pregnancy and childhood outcomes: time to change guidelines indicating apparently 'safe' levels of alcohol during pregnancy? A systematic review and meta-analyses. *BMJ Open* 2017;7:e015410.
33 Benedetto U, Head SJ, Angelini GD, *et al*. Statistical primer: propensity score matching and its alternatives. *Eur J Cardiothorac Surg* 2018;53:1112–7.
34 Brookhart MA, Wyss R, Layton JB, *et al*. Propensity score methods for confounding control in nonexperimental research. *Circ Cardiovasc Qual Outcomes* 2013;6:604–11.
35 Mansournia MA, Etminan M, Danaei G, *et al*. Handling time varying confounding in observational research. *BMJ* 2017;359:j4587.
36 Naimi AI, Cole SR, Kennedy EH. An introduction to G methods. *Int J Epidemiol* 2017;46:756–62.
37 Hernán MA RJ. *Causal inference: what if*. Boca Raton: Chapman & Hill/CRC, 2020: 171–82.
38 Williamson T, Ravani P. Marginal structural models in clinical research: when and how to use them? *Nephrol Dial Transplant* 2017;32:ii84–90.
39 Gunasekara FI, Richardson K, Carter K, *et al*. Fixed effects analysis of repeated measures data. *Int J Epidemiol* 2014;43:264–9.
40 Fergusson DM, Boden JM, Horwood LJ. Psychosocial sequelae of cannabis use and implications for policy: findings from the Christchurch health and development study. *Soc Psychiatry Psychiatr Epidemiol* 2015;50:1317–26.
41 Fergusson DM, Boden JM, Horwood LJ. Tests of causal links between alcohol abuse or dependence and major depression. *Arch Gen Psychiatry* 2009;66:260–6.
42 Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol* 2013;42:1511–9.
43 Pearce N, Lawlor DA. Causal inference—so much more than statistics. *Int J Epidemiol* 2016;45:1895–903.
44 VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annu Rev Public Health* 2016;37:17–32.
45 Dunning T. *Natural experiments in the social sciences: a design-based approach*. Cambridge: Cambridge University Press, 2012.
46 Gage SH, Munafò MR, Davey Smith G, Smith GD. Causal inference in developmental origins of health and disease (DOHaD) research. *Annu Rev Psychol* 2016;67:567–85.
47 Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018;362:k601.

Appendix C

Sensitivity of the alcohol–inflammation relationship

Preface

This study was published as **Visontay, R.**, Mewton, L., Sunderland, M., Bell, S., Britton, A., Osman, B., North, H., Mathew, N., & Slade, T. (2023). A comprehensive evaluation of the longitudinal association between alcohol consumption and a measure of inflammation: Multiverse and vibration of effects analyses. *Drug and Alcohol Dependence, 247*, 109886.

RV conceptualised the study with assistance from LM, MS, and TS. RV conducted the analyses and drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Drug and Alcohol Dependence 247 (2023) 109886

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

A comprehensive evaluation of the longitudinal association between alcohol consumption and a measure of inflammation: Multiverse and vibration of effects analyses

Rachel Visontay [a,*], Louise Mewton [a,b], Matthew Sunderland [a], Steven Bell [c,d,e], Annie Britton [e], Bridie Osman [a], Hayley North [f], Nisha Mathew [f,g], Tim Slade [a]

[a] The Matilda Centre for Research in Mental Health and Substance Use, Level 6, Jane Foss Russell Building, G02, The University of Sydney, NSW2006, Australia
[b] Centre for Healthy Brain Ageing, Level 1, AGSM (G27), University of New South Wales, Gate 11, Botany Street, Sydney, NSW2052, Australia
[c] Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK
[d] British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK
[e] Research Department of Epidemiology and Public Health, University College London, London, UK
[f] Neuroscience Research Australia, Randwick, NSW2031, Australia
[g] School of Clinical Medicine, UNSW Medicine and Health, University of New South Wales, NSW 2052, Australia

ARTICLE INFO

Keywords:
Alcohol drinking
Inflammation
C-reactive protein
hsCRP
Multiverse
Vibration of effects

ABSTRACT

Background: Moderate alcohol consumption appears to be associated with reduced inflammation. Determining whether this association is robust to common variations in research parameters has wide-reaching implications for our understanding of disease aetiology and public health policy. We aimed to conduct comprehensive multiverse and vibration of effects analyses evaluating the associations between alcohol consumption and a measure of inflammation.

Methods: A secondary analysis of the 1970 British Birth Cohort Study was performed, using data from 1970 through 2016. Measurements of alcohol consumption were taken in early/mid-adulthood (ages 34 and 42), and level of inflammation marker high-sensitivity C-reactive protein (hsCRP) at age 46. Multiverse analyses were applied to comparisons of low-to-moderate consumption and consumption above various international drinking guidelines with an 'abstinent' reference. Research parameters of interest related to: definitions of drinking and reference groups; alcohol consumption measurement year; outcome variable transformation; and breadth of covariate adjustment. After identifying various analytic options within these parameters and running the analysis over each unique option combination, specification curve plots, volcano plots, effect ranges, and variance decomposition metrics were used to assess consistency of results.

Results: A total of 3101 individuals were included in the final analyses, with primary analyses limited to those where occasional consumers served as reference. All combinations of research specifications resulted in lower levels of inflammation amongst low-to-moderate consumers compared to occasional consumers (1st percentile effect: −0.21; 99th percentile effect: −0.04). Estimates comparing above-guidelines drinking with occasional consumers were less definitive (1st percentile effect: −0.26; 99th percentile effect: 0.43).

Conclusions: The association between low-to-moderate drinking and lower hsCRP levels is largely robust to common variations in researcher-defined parameters, warranting further research to establish whether this relationship is causal. The association between above-guidelines drinking and hsCRP levels is less definitive.

1. Introduction

Acute inflammation is a vital and adaptive response to pathogens and tissue injury. However, low-grade chronic inflammation, i.e., inflammation that fails to resolve or is continuously triggered, can negatively affect tissue and organs over time. This pathology ranges in severity, on the higher end associated with inflammatory diseases such as asthma, atherosclerosis, inflammatory bowel disease, and rheumatoid

* Corresponding author.
E-mail address: rachel.visontay@sydney.edu.au (R. Visontay).

https://doi.org/10.1016/j.drugalcdep.2023.109886
Received 17 November 2022; Received in revised form 16 February 2023; Accepted 16 April 2023
Available online 20 April 2023

arthritis (Calder et al., 2013; Furman et al., 2019; Ridker, 2004). Inflammation is also increasingly implicated in a variety of other serious conditions including psychiatric disorders, cancer, type 2 diabetes, and cardiovascular disease (Ansar and Ghosh, 2013; Avan et al., 2018; Furman et al., 2019; Réus et al., 2015).

Of the many putative lifestyle influences on inflammation, there is strong evidence for the role of alcohol consumption. Alcohol use disorder is associated with increased levels of pro-inflammatory cytokines (Adams et al., 2020) as well as neuroinflammation specifically (de Timary et al., 2017; He and Crews, 2008). Harmful drinking is responsible for substantial disease burden globally (Bryazka et al., 2022), and there is evidence that inflammation may mediate the increased risk heavy drinking brings for a range of health outcomes (González-Reimers et al., 2014).

However, as for many conditions to which alcohol consumption is thought to contribute pathophysiologically (Visontay et al., 2022), some research suggests that compared to abstaining, lower levels of drinking may actually be beneficial when it comes to inflammation (Albert et al., 2003; Bell et al., 2017; Pai et al., 2006; Paulson et al., 2018; Volpato et al., 2004). This is typically indexed via inflammatory markers such as high-sensitivity C-reactive protein (hsCRP) – widely used by the clinical and research communities due to its stable expression in response to elevated pro-inflammatory cytokines (Coventry et al., 2009). Reduced inflammation has also been suggested as a partial mediator of the lower risk moderate drinking seemingly affords against conditions such as cardiovascular disease (Pai et al., 2006) and depression (Paulson et al., 2018).

However, there are major inconsistencies in the evidence base. For example, short-term alcohol administration tends to demonstrate no relationship with hsCRP (Brien et al., 2011). Several cohort studies have also failed to find a relationship between drinking and inflammation (Menezes et al., 2019; van't Klooster et al., 2020). A possible explanation for these conflicting findings is that data processing and analysis choices, such as which covariates are controlled for, vary considerably between studies. A growing literature is beginning to uncover the impact these alternative specifications in research parameters, or 'researcher degrees of freedom', (Simmons et al., 2011) can have on effect estimates, challenging implied assumptions that these decisions are innocuous. To this end, three similar analytic frameworks have emerged to quantify sensitivity (or robustness) to alternative specifications: multiverse analysis (Steegen et al., 2016), specification curve analysis (Simonsohn et al., 2020), and vibration of effects (VoE) (Patel et al., 2015). These frameworks are an extension of standard sensitivity analyses, providing methods to comprehensively analyse and summarise the sensitivity of exposure–outcome associations to alternative specifications.

It is well documented that these alternative specifications can influence effect sizes, P-values, and even effect direction for a variety of epidemiological associations. This is becoming increasingly clear in alcohol–health associations specifically. Such parameters include how the abstaining reference group is defined (Stockwell et al., 2016), with increasing calls (but limited uptake) for occasional consumers to be used as the reference given their greater prevalence and more normative risk factor profiles (Naimi et al., 2022). How alcohol intake is categorised, which covariates are adjusted for, and choice of modelling approach have also been identified as factors that may substantially impact findings in this space (Bell et al., 2017; Chu et al., 2020; Stockwell et al., 2016). It is thus concerning that there remains great variation in these parameters in the literature, and that their potential impact is rarely acknowledged in individual studies, where typically only one set of specifications are tested.

The present study aims to address this through systematically applying methods from enhanced sensitivity frameworks to test the alcohol–inflammation association. Using the 1970 British Birth Cohort Study (BCS70) – a source of rich, representative data – we assessed the relationship between alcohol consumption in early/mid-adulthood and hsCRP at age 46. The research parameters of interest explored here comprise: composition of reference group; classification of drinking groups according to various national guidelines ranging from conservative through liberal; measurement year(s) for alcohol consumption; transformation of the hsCRP outcome; and breadth of covariate adjustment. Each unique combination of specifications within these parameters constitutes a possible 'universe' within the analytic multiverse; here alcohol–hsCRP analyses are iterated over each universe and results interpreted on aggregate.

2. Material and methods

Reporting for this study adheres to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (von Elm et al., 2007).

2.1. Data source and participants

BCS70 is a prospective cohort study targeting all individuals born in a single week in 1970 in Great Britain. It has had 11 sweeps to date, collecting diverse information on health, education, lifestyle behaviours, and socio-economic factors (Elliott and Shepherd, 2006). Individuals possessing valid alcohol data for at least one of ages 34 and 42, and an hsCRP measurement at age 46, were considered eligible for the present study.

2.2. Exposure

Self-reported alcohol consumption data at ages 34 (2004) and 42 (2012) were used to generate the exposure variables. In the present study, data on frequency and volume of alcohol consumption were converted to grams of alcohol per week, and then used to categorise individuals as either lifetime abstainers, former consumers, occasional consumers (less than weekly), low-to-moderate consumers (at least once per week), or those consuming above various international drinking guidelines at each of the waves. Additional exposure variables were created re-categorising current consumers according to their average weekly volume across the two ages, as well as their maximum weekly volume. Four separate versions were created for each of these four variables, where the threshold separating low-to-moderate from above-guidelines consumers varied according to the national drinking guidelines of four chosen countries. See Table 1 and Supplementary Methods for further detail.

2.3. Outcome

A non-fasting blood sample was taken at age 46 (2016), with hsCRP concentrations assessed via immunoturbidimetry (Hamer et al., 2020). Prior to analysis, 110 individuals with hsCRP concentrations $> 10\ mg\ L^{-1}$ were excluded, given these concentrations suggest acute inflammation or infection (Giollabhui et al., 2020). Chi-Square tests of independence revealed no relationship between drinking category at age 34 and likelihood of hsCRP being $> 10\ mg\ L^{-1}$, although there was slight evidence of a relationship between drinking category at age 42 and likelihood of hsCRP being $> 10\ mg\ L^{-1}$ (Supplementary Table 1). The hsCRP distribution was positively skewed, so a natural log-transformed version was also created which led to an approximately normal distribution, and universes with both untransformed and log-transformed hsCRP were tested. Normality was assessed via visual inspection.

2.4. Covariates

Covariates were chosen *a prori* based on their relevance to the alcohol–inflammation relationship and were derived from data collected at either birth (for time-invariant variables) or age 30 (this was the

Table 1
Parameters and specifications tested.

Parameter	Specifications	*N* of specifications
Composition of 'abstinence' reference group	(i) Lifetime abstainers only (ii) Lifetime abstainers and former consumers (iii) (iii) Occasional consumers only	3
Cut-point for above-guidelines drinking	(i) Dutch guidelines (>70 g/wk) (Health Council of the Netherlands, 2015) (ii) UK guidelines (>112 g/wk) (Department of Health, 2016) (iii) US guidelines (F: >98 g/wk; M: >196 g/wk) (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020) (iv) Former Spanish guidelines[a] (F: >170 g/wk; M: >280 g/wk) (Spanish Ministry of Health, 2020)	4
Year of alcohol consumption measurement	(i) 2004 (ii) 2012 (iii) Maximum of 2004 and 2012 (iv) Average of 2004 and 2012	4
Transformation of hsCRP outcome	(i) Leave untransformed (ii) Natural log-transform	2
Breadth of covariate adjustment	(i) No covariates adjusted for (ii) Demographics (iii) Demographics + Physical health (iv) Demographics + Mental health (v) Demographics + Lifestyle (vi) Demographics + Social (vii) Demographics + Physical health + Mental health (viii) Demographics + Physical health + Lifestyle (ix) Demographics + Mental health + Lifestyle (x) Demographics + Mental health + Social (xi) Demographics + Physical health + Social (xii) Demographics + Lifestyle + Social (xiii) All covariates adjusted for	13
Total universes		1248

The total number of universes is equal to the product of the five parameters' specification *N*s.
[a] These guidelines are no longer in effect as of 2020, but are indicative of liberal thresholds for 'moderate consumption' adopted by various policymakers and researchers.

nearest measurement occasion preceding age 34 alcohol consumption measurement, chosen to ensure covariates temporally preceded alcohol measurement). Covariates of interest were divided into five groups: demographic/socioeconomic, physical health, lifestyle behaviours, social, and mental health. See Supplementary Methods for further detail.

2.5. Parameters of interest and their specifications

As detailed in Table 1, the research parameters for which alternative specifications were explored were: 1) composition of 'abstinence' reference group (including using occasional consumers instead of true abstainers); 2) the cut-point between low-to-moderate and above-guidelines drinking (based on four different national drinking guidelines ranging from conservative to liberal); 3) measurement year(s) of alcohol consumption predictor; 4) hsCRP outcome transformation; and 5) breadth of covariate adjustment. This produced a total of 1248 universes. Note that specification alternatives often resulted in sample size variation.

2.6. Statistical analysis

A linear regression of continuous hsCRP levels on categorical alcohol consumption was performed for all 1248 universes, producing two comparisons: 'abstinence' vs. low-to-moderate alcohol consumption, and 'abstinence' vs. above-guidelines alcohol consumption. Effect sizes were standardised.

Various approaches from the sensitivity frameworks introduced earlier were employed to interpret results. Specification curve plots allowed for visual inspection of variation in effect sizes and P-values across universes. In these plots, universes are arranged in order of effect size, with an accompanying panel depicting the corresponding specifications. Results were also graphically depicted using volcano plots, which promote evaluation of (in)consistency of effect direction (Klau et al., 2021). Metrics developed in the VoE framework were also calculated: the Range of Betas (RBs) and Range of P-values (RPs). These are simple figures calculated as the range of standardised betas and P-values between the 1st and 99th percentiles, where larger ranges indicate greater variability across universes (Patel et al., 2015).

Additionally, variance in effect estimates were decomposed by parameter, and each specification plotted against effect estimates in box and whisker plots. All analyses were conducted using R version 4.1.3 with packages *multiverse* (Sarma, 2021) and *specr* (Masur and Scharkow, 2020).

3. Results

3.1. Descriptive characteristics

In total, 11 123 individuals provided valid age 34 and/or age 42 alcohol data, with 3211 of these also having age 46 hsCRP data. The small size of the latter figure owed to attrition and low health test completion rates, although drinking status and covariates were largely unassociated with providing hsCRP data (Supplementary Table 2). After removing those with elevated hsCRP, the final sample was 3101.

Compared to other categories, low-to-moderate consumers were more likely to rate their health as excellent, and had lower mean Malaise Inventory and hsCRP values (Table 2 and Supplementary Table 3). There was considerable movement amongst drinking categories between ages 34 and 42, particularly a decrease in individuals drinking above-guidelines and an increase in those drinking occasionally.

3.2. Summary of effects and their heterogeneity

Despite a considerable range of effects (RB=0.46; RP=4.03), results favoured lower levels of hsCRP in low-to-moderate consumers compared to the 'abstainer' reference, with effects at both the 1st (−0.50) and 99th (−0.05) percentiles in this direction. None of the 1248 estimates were consistent with increased hsCRP levels (Supplementary Figs. 1 and 2).

Table 2
Cohort characteristics for selected covariates and hsCRP by age 34 alcohol consumption categories based on UK guidelines.

	Lifetime abstainer (N=52)	Former consumer (N=108)	Occasional consumer (N=621)	Low-to-moderate consumer (N=1160)	Above-guidelines consumer (N=684)	Total (N=2625)
UK standard drinks per week						
Mean (SD)	NA (NA)	NA (NA)	NA (NA)	6.99 (3.68)	29.8 (18.6)	15.5 (16.1)
Median [Min, Max]	NA [NA, NA]	NA [NA, NA]	NA [NA, NA]	6.00 [1.00, 14.0]	24.0 [15.0, 280]	10.0 [1.00, 280]
Missing	52 (100%)	108 (100%)	621 (100%)	0 (0%)	0 (0%)	781 (29.8%)
Sex						
Female	28 (53.8%)	71 (65.7%)	406 (65.4%)	686 (59.1%)	154 (22.5%)	1345 (51.2%)
Male	24 (46.2%)	37 (34.3%)	215 (34.6%)	474 (40.9%)	530 (77.5%)	1280 (48.8%)
Mother's age at birth						
Mean (SD)	28.1 (5.78)	25.1 (4.72)	25.9 (5.23)	26.4 (5.14)	26.1 (5.22)	26.2 (5.19)
Median [Min, Max]	28.0 [17.0, 41.0]	25.0 [16.0, 38.0]	25.0 [15.0, 44.0]	26.0 [15.0, 43.0]	25.0 [16.0, 49.0]	25.0 [15.0, 49.0]
Missing	12 (23.1%)	11 (10.2%)	35 (5.6%)	77 (6.6%)	48 (7.0%)	183 (7.0%)
Self-rated health (Age 30)						
Excellent	16 (30.8%)	25 (23.1%)	185 (29.8%)	411 (35.4%)	200 (29.2%)	837 (31.9%)
Good	25 (48.1%)	60 (55.6%)	322 (51.9%)	590 (50.9%)	355 (51.9%)	1352 (51.5%)
Fair	6 (11.5%)	14 (13.0%)	79 (12.7%)	96 (8.3%)	87 (12.7%)	282 (10.7%)
Poor	2 (3.8%)	5 (4.6%)	13 (2.1%)	6 (0.5%)	5 (0.7%)	31 (1.2%)
Missing	3 (5.8%)	4 (3.7%)	22 (3.5%)	57 (4.9%)	37 (5.4%)	123 (4.7%)
BMI (Age 30)						
Mean (SD)	24.2 (4.27)	24.8 (5.48)	25.2 (4.81)	24.2 (4.09)	25.0 (3.44)	24.7 (4.21)
Median [Min, Max]	24.0 [18.1, 40.2]	23.6 [16.9, 45.2]	24.1 [16.4, 57.2]	23.4 [11.1, 61.2]	24.9 [9.98, 44.9]	24.0 [9.98, 61.2]
Missing	5 (9.6%)	9 (8.3%)	43 (6.9%)	83 (7.2%)	47 (6.9%)	187 (7.1%)
Smoking status (Age 30)						
Never smoked	37 (71.2%)	47 (43.5%)	304 (49.0%)	548 (47.2%)	241 (35.2%)	1177 (44.8%)
Formerly smoked	3 (5.8%)	18 (16.7%)	89 (14.3%)	253 (21.8%)	144 (21.1%)	507 (19.3%)
Occasionally smoke	0 (0%)	9 (8.3%)	34 (5.5%)	94 (8.1%)	69 (10.1%)	206 (7.8%)
Smoke daily	9 (17.3%)	30 (27.8%)	172 (27.7%)	208 (17.9%)	193 (28.2%)	612 (23.3%)
Missing	3 (5.8%)	4 (3.7%)	22 (3.5%)	57 (4.9%)	37 (5.4%)	123 (4.7%)
Weekly exercise (Age 30)						
No	14 (26.9%)	29 (26.9%)	213 (34.3%)	295 (25.4%)	172 (25.1%)	723 (27.5%)
Yes	35 (67.3%)	74 (68.5%)	386 (62.2%)	808 (69.7%)	475 (69.4%)	1778 (67.7%)
Missing	3 (5.8%)	5 (4.6%)	22 (3.5%)	57 (4.9%)	37 (5.4%)	124 (4.7%)
Malaise scale total (Age 30)						
Mean (SD)	3.65 (3.73)	4.23 (3.69)	3.61 (3.48)	2.93 (3.06)	3.30 (3.05)	3.26 (3.22)
Median [Min, Max]	3.00 [0, 15.0]	3.00 [0, 16.0]	3.00 [0, 18.0]	2.00 [0, 21.0]	2.00 [0, 22.0]	2.00 [0, 22.0]
Missing	3 (5.8%)	5 (4.6%)	25 (4.0%)	67 (5.8%)	41 (6.0%)	141 (5.4%)
Ever member of communal organisation (Age 30)						
No	34 (65.4%)	86 (79.6%)	461 (74.2%)	838 (72.2%)	553 (80.8%)	1972 (75.1%)
Yes	15 (28.8%)	17 (15.7%)	138 (22.2%)	265 (22.8%)	94 (13.7%)	529 (20.2%)
Missing	3 (5.8%)	5 (4.6%)	22 (3.5%)	57 (4.9%)	37 (5.4%)	124 (4.7%)
hsCRP (log-transformed) (Age 46)						
Mean (SD)	-0.122 (3.04)	-0.0338 (2.27)	0.0187 (2.12)	-0.312 (2.80)	-0.0637 (1.67)	-0.154 (2.38)
Median [Min, Max]	0.0953 [−20.7, 2.16]	0.182 [−20.7, 2.30]	0.0953 [−20.7, 2.24]	-0.105 [−20.7, 2.29]	-0.105 [−20.7, 2.30]	0 [−20.7, 2.30]

Abbreviations: SD, standard deviation; BMI, body mass index; hsCRP, high-sensitivity C-reactive protein.
See Supplementary Table 3 for characteristics by age 42 drinking categories.

Estimates comparing above-guidelines drinking with 'abstainers' were less definitive (RB=0.86; RP=2.29; Supplementary Figs. 3 and 4), with 136/1248 estimates consistent with increased hsCRP levels, and the effects at the 1st (−0.47) and 99th (0.39) percentiles being in opposite directions. Supplementary Figs. 1 and 3 reveal considerable variation in effects according to reference group composition (lifetime abstainers and former consumers; lifetime abstainers only; occasional consumers), making it difficult to interpret other parameters.

Occasional consumers proved the only appropriate reference given the very small sample size resulting from the other two reference specifications. Analyses with this specification were thereafter considered the primary analyses, to mitigate confounding by sample size and aid interpretability (Figs. 1 and 2; Supplementary Figs. 5 and 6).

The range of effects in these analyses was narrower, with RBs of 0.16 (RP= 4.65) and 0.69 (RP= 2.87) for the low-to-moderate (0/416 estimates consistent with increased hsCRP; 1st percentile effect: −0.21; 99th percentile effect: −0.04) and above-guidelines (85/416 estimates consistent with increased hsCRP; 1st percentile effect: −0.26; 99th percentile effect: 0.43) comparisons respectively.

3.3. Variance decomposition and impact of individual specifications

Using a multi-level model with random effects for each parameter of interest, the highest intraclass correlation coefficient (ICC) for the low-to-moderate consumers vs occasional consumers comparison belonged to breadth of covariate control (0.20), followed by which national guidelines were used to define drinking categories (0.13). The highest ICC for the above-guidelines vs occasional consumers comparison belonged to measurement year that drinking category was based on (0.23), followed by which national guidelines were used to define drinking categories (0.14).

Box and whisker plots (Fig. 3) offer insights into the impact of particular specifications. Mean protective effects were greater when the ceiling for low-to-moderate drinking was lower. Conversely, the Spanish-based categorisation of above-guidelines drinking – i.e., when this group was restricted to those drinking most heavily – resulted in the

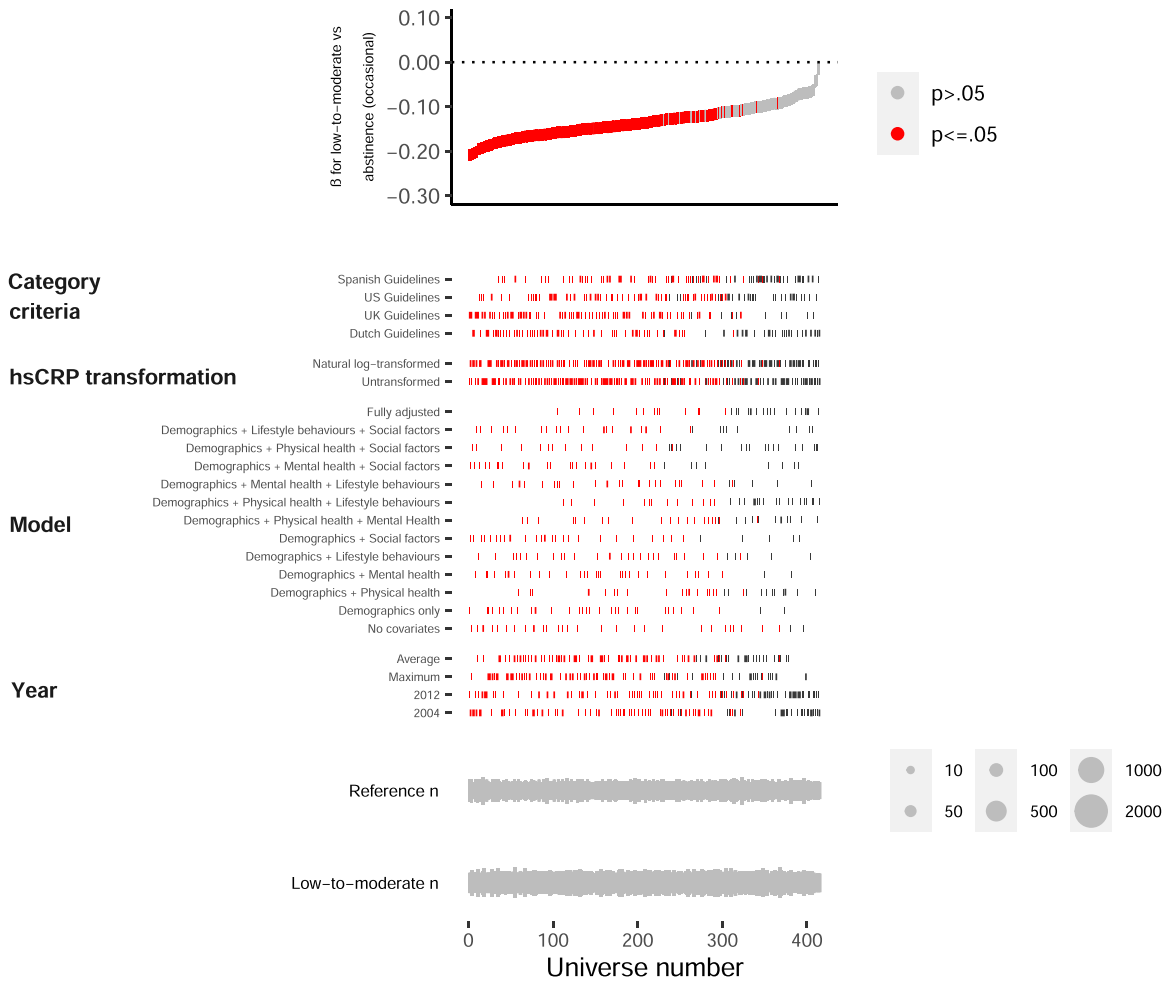


Fig. 1. Specification curve plot of estimates comparing low-to-moderate consumers with a reference group composed of occasional consumers. The top section displays the results associated with each universe, ordered by effect size. Negative values indicate drinking is associated with lower hsCRP. The middle section shows which specifications correspond to each result in the top panel. The bottom section displays the number of individuals in the reference and comparator groups in each universe.

smallest mean protective effect. Fig. 3 also shows that protective effects of above-guidelines consumption appear when considering drinking at age 34 only (12 year before hsCRP measurement) but not drinking at age 42 only (four years before hsCRP measurement).

3.4. Post-hoc sensitivity analyses

The multiverse analyses were re-run on a reduced sample, restricted to individuals with no missing covariate data (n=1358). Results were consistent with the main analyses (see Supplementary Methods and Supplementary Figs. 7–9).

4. Discussion

As inconsistencies in the existing evidence base would suggest, this analysis revealed substantial heterogeneity, or effect vibration, in estimates of the alcohol–hsCRP relationship when research parameter specifications are varied. Composition of the 'abstinence' reference group had the clearest impact on effect estimates. When analyses were restricted to universes with occasional consumers as reference, additional impacts of the breadth of covariate adjustment (for the low-to-moderate drinking comparison), and measurement year and national guidelines used to classify drinking groups (for the above-guidelines drinking comparison), were apparent.

However, despite heterogeneity in effect and P-value size, the direction of results from comparisons of low-to-moderate alcohol consumption with 'abstinence' was robust to parameter specifications. Results were consistent with lower inflammation in the low-to-moderate alcohol consumption group at follow-up. While this does not substantiate a causal role of moderate drinking in reducing inflammation, it fulfils a prerequisite. That is, before determining whether an observed association is causal, that association must be confirmed as a robust one that holds across methodological variations – what has been labelled the 'consistency' criteria for causation (Hill, 1965).

Plausible biological mechanisms that may account for this association have been suggested, largely derived from short-term administration studies in animal and human models, and *in vitro* work. These centre on low-dose ethanol's ability to inhibit pro-inflammatory cytokines and chemokines (Muralidharan et al., 2014; Szabo, 2007). Additionally, given the established impact of chronic stress on inflammation (Rohleder, 2019), moderate drinking may exert a protective effect mediated by stress reduction.

Interestingly, previous research has shown that excessive alcohol use begins to exert opposite effects on these mechanisms, promoting pro-inflammatory responses (Szabo, 2007). As such, it is ostensibly surprising that the above-guidelines comparisons were less definitive. In fact, those effects with smaller P-values belonged to the analyses in which above-guidelines drinking were associated with lower levels of inflammation compared to the reference. Importantly though, the above-guidelines category did not reflect disordered/excessive drinking

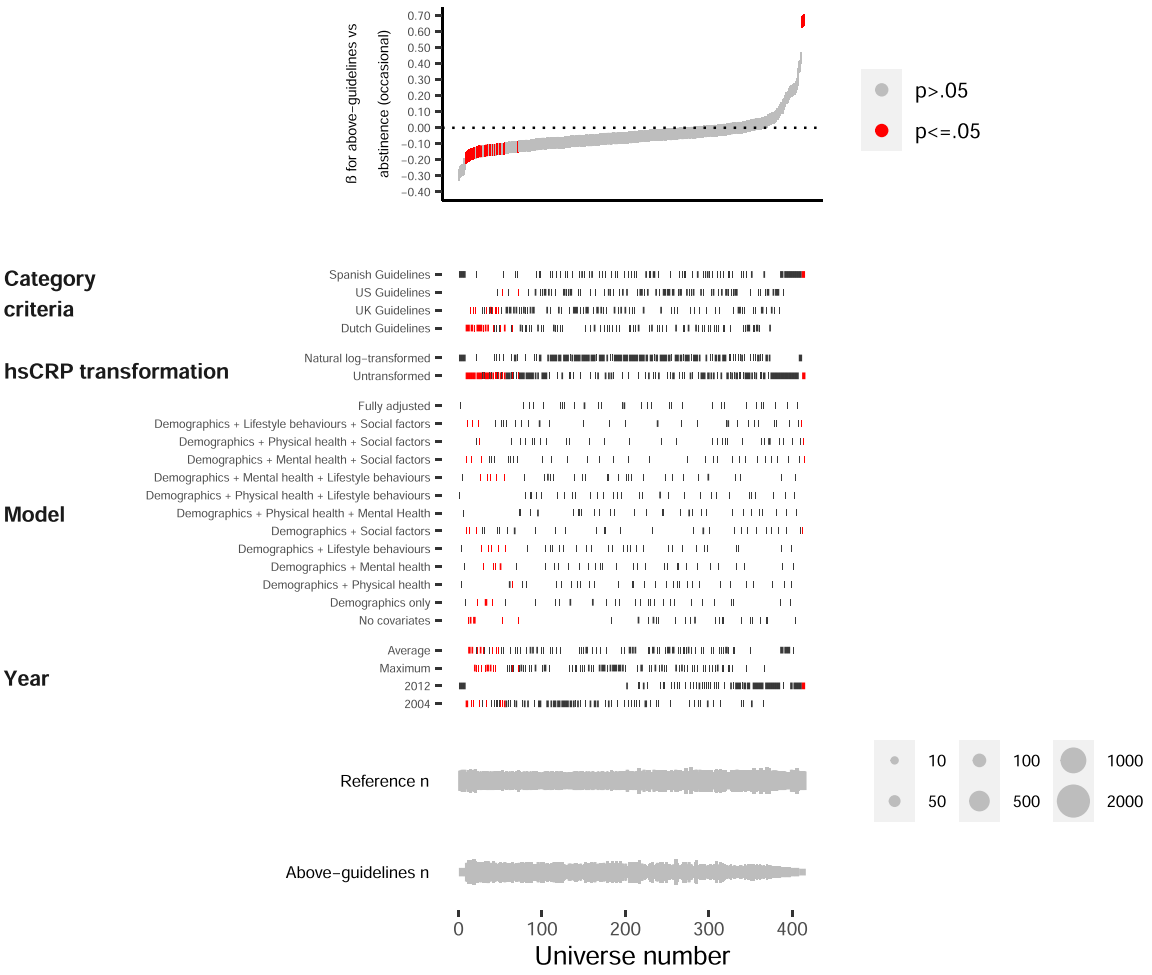


Fig. 2. Specification curve plot of estimates comparing above-guidelines consumers with a reference group composed of occasional consumers. The top section displays the results associated with each universe, ordered by effect size. Negative values indicate drinking is associated with lower hsCRP. The middle section shows which specifications correspond to each result in the top panel. The bottom section displays the number of individuals in the reference and comparator groups in each universe.

specifically, which is difficult to capture in general population samples. Given most individuals in this category were not drinking substantially over guidelines (Supplementary Table 4), they may still have been drinking at levels below those at which ethanol promotes inflammation. Indeed, specifying above-guidelines drinking cut-points based on Spanish guidelines – i.e., restricting this group to those drinking most heavily – resulted in a bimodal split in effect direction (Fig. 2), and the only interquartile effect range *not* situated entirely in the direction of reduced inflammation (Fig. 3).

4.1. Strengths and limitations

To the authors' knowledge, this is the first study to examine the longitudinal association between alcohol consumption and a measure of inflammation via an enhanced sensitivity analysis framework. Previously, sensitivity considerations had only been explored in limited, disparate ways. The focus of the present study was on key parameters that researchers must consider carefully when engaging in study design and analysis – those which existing evidence indicates stand to impact results and for which diverging specifications are typical in the literature. The present findings stand to inform both the evidence base on the alcohol–inflammation relationship as well as future research methodology.

A further strength of this study was the large sample size and the rich covariate selection available, including BMI and pre-existing inflammatory conditions. Age, a key determinant in inflammation, was held constant by design of the cohort. Sensitivity analyses of individuals with full covariate data were consistent with the main analyses, increasing confidence that findings were not confounded by sample variation between universes.

However, a limitation of this work was that the clear effects of reference group composition (Supplementary Figs. 1 and 3) could not be confirmed to be a result of the characteristics of the groups themselves, as opposed to the small number of lifetime abstainers and former consumers. Additionally, because BCS70 tested only a single inflammatory marker, the present analyses were restricted to hsCRP alone as a proxy for inflammation. This limits our ability to generalise results to inflammation more broadly; it is possible that other measures may have shown discrepant results (e.g., differences may be evident between alcohol–CRP and alcohol– interleukin-6 (IL-6) relationships (Silverwood et al., 2014)).

Measurement error may also have impacted the findings; while we removed cases with indications of acute inflammation, promoting a focus on chronic low-grade inflammation, hsCRP was only captured at a single time point. Additionally, the effect of certain covariates on hsCRP levels may have been larger had there been a shorter interval between their relative assessments. Also, most data were collected via self-report, which is problematic for the kinds of health covariates employed here (although it is the chief measurement modality in the literature). Further measurement bias was possible given the considerable movement

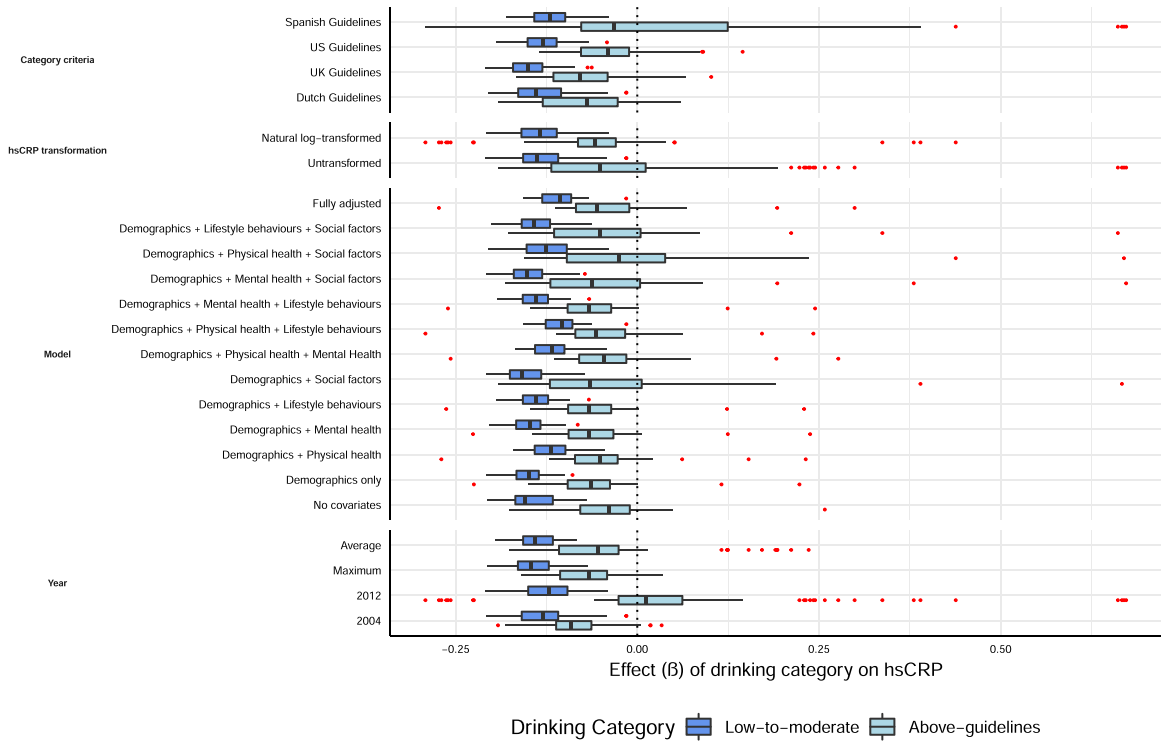


Fig. 3. Box and whisker plot depicting impact of specifications on comparisons between low-to-moderate/above-guidelines alcohol consumption and occasional drinking. Red dots represent outliers.

between drinking categories from age 34 to age 42 – possibly explained by UK policy changes beginning in 2003 (Nicholls, 2012) as well as natural 'maturing out' of drinking – suggesting hidden 'drinker drift' in the interim between ages 42 and hsCRP measurement.

Finally, exploring additional parameters of interest, e.g., predominant beverage consumed, proved difficult within the multiverse framework. Although given suggested protective mechanisms operate via ethanol itself (Krenz and Korthuis, 2012) – common to all alcoholic drinks – this should not have considerably impacted results.

4.2. Implications

Given the present evidence of a robust association between low-to-moderate drinking and reduced hsCRP, future work should investigate whether this relationship is causal. One such avenue is non-linear Mendelian randomisation; to date one study has found a small protective effect for low-level drinking against CRP, but a positive linear relationship between alcohol and IL-6 (Silverwood et al., 2014). Causal mediation analyses should also be undertaken to test the hypothesis that inflammation may mediate the relationship between moderate drinking and reduced cardiovascular disease risk. Were moderate alcohol consumption found to have a causal role in lowering inflammation, this would be important information for clinicians and policymakers, including for the generation of low-risk drinking guidelines. However, these findings would have to be balanced against the known harms to other organ systems/disease processes that accrue at even low-level drinking, such as increased cancer risk (Rehm et al., 2019). While results from universes with and without adjustment for sex did not vary substantially, sex-specific differences in the alcohol–inflammation relationship should also be explored in future work.

The present findings lend support to using occasional/very low volume consumers as the reference group in future work. While this requires the assumption that such infrequent alcohol consumption would be unlikely to have a biological effect, the present study demonstrates clear benefits in the way of increased statistical power and more normative covariate profiles compared with lifetime abstainers and former consumers. Results derived using occasional consumer reference groups are also less likely to be contaminated by 'sick quitters' (those for whom illness precipitates abstinence), former consumers misidentified as lifetime abstainers, or hidden illness belonging to those excluded from the reference group (Naimi et al., 2022).

5. Conclusions

The present study supports a small but robust association between low-to-moderate alcohol consumption and lower levels of the inflammatory marker hsCRP, while the association between drinking at above-guidelines levels and hsCRP demonstrated greater variability. Further research using methodologies that promote causal inference is required. Common variations in research parameters in this field – such as composition of reference group, how drinking groups are defined, and which covariates are controlled for – are *not* innocuous, and must be considered carefully.

Role of the funding source

Nothing declared.

CRediT authorship contribution statement

Rachel Visontay, Louise Mewton, Matthew Sunderland, and Tim Slade conceived of and designed the study. Rachel Visontay analysed the data, with input from Louise Mewton, Matthew Sunderland, and Tim Slade, and with Steven Bell, Annie Britton, Bridie Osman, Hayley North, and Nisha Mathew helping to interpret the results. Rachel Visontay drafted the manuscript, and all authors critically revised it for important intellectual content. All authors read and approved the final version of the manuscript.

Declaration of Competing Interest

No conflict declared.

Acknowledgements

This work was supported by a National Health and Medical Research Council Postgraduate Scholarship (GNT1190255), as well as by the University of Sydney.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.drugalcdep.2023.109886.

References

Adams, C., Conigrave, J.H., Lewohl, J., Haber, P., Morley, K.C., 2020. Alcohol use disorder and circulating cytokines: A systematic review and meta-analysis. Brain Behav. Immun. https://doi.org/10.1016/j.bbi.2020.08.002.

Albert, M.A., Glynn, R.J., Ridker, P.M., 2003. Alcohol consumption and plasma concentration of C-reactive protein. Circulation 107, 443–447. https://doi.org/10.1161/01.CIR.0000045669.16499.EC.

Ansar, W., Ghosh, S., 2013. C-reactive protein and the biology of disease. Immunol. Res. 56, 131–142. https://doi.org/10.1007/s12026-013-8384-0.

Avan, A., Tavakoly Sany, S.B., Ghayour-Mobarhan, M., Rahimi, H.R., Tajfard, M., Ferns, G., 2018. Serum C-reactive protein in the prediction of cardiovascular diseases: overview of the latest clinical studies and public health practice. J. Cell. Physiol. 233, 8508–8525.

Bell, S., Mehta, G., Moore, K., Britton, A., 2017. Ten-year alcohol consumption typologies and trajectories of C-reactive protein, interleukin-6 and interleukin-1 receptor antagonist over the following 12 years: a prospective cohort study. J. Intern. Med. 281, 75–85. https://doi.org/10.1111/joim.12544.

Brien, S.E., Ronksley, P.E., Turner, B.J., Mukamal, K.J., Ghali, W.A., 2011. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. BMJ 342, d636.

Bryazka, D., Reitsma, M.B., Griswold, M.G., Abate, K.H., Abbafati, C., Abbasi-Kangevari, M., Abbasi-Kangevari, Z., Abdoli, A., Abdollahi, M., Abdullah, A.Y.M., Abhilash, E.S., Abu-Gharbieh, E., Acuna, J.M., Addolorato, G., Adebayo, O.M., Adekanmbi, V., Adhikari, K., Adhikari, S., Adnani, Q.E.S., Afzal, S., Agegnehu, W.Y., Aggarwal, M., Ahinkorah, B.O., Ahmad, A.R., Ahmad, S., Ahmad, T., Ahmadi, A., Ahmadi, S., Ahmed, H., Ahmed Rashid, T., Akunna, C.J., al Hamad, H., Alam, M.Z., Alem, D.T., Alene, K.A., Alimohamadi, Y., Alizadeh, A., Allel, K., Alonso, J., Alvand, S., Alvis-Guzman, N., Amare, F., Ameyaw, E.K., Amiri, S., Ancuceanu, R., Anderson, J.A., Andrei, C.L., Andrei, T., Arabloo, J., Arshad, M., Artamonov, A.A., Aryan, Z., Asaad, M., Asemahagn, M.A., Astell-Burt, T., Athari, S.S., Atnafu, D.D., Atorkey, P., Atreya, A., Ausloos, F., Ausloos, M., Ayano, G., Ayanore, M.A. ayanore, Ayinde, O.O., Ayuso-Mateos, J.L., Azadnajafabad, S., Azanaw, M.M., Azangou-Khyavy, M., Azari Jafari, A., Azzam, A.Y., Badiye, A.D., Bagheri, N., Baghernieh, S., Bairwa, M., Bakkannavar, S.M., Bakshi, R.K., Balchut/Bilchut, A.H., Bärnighausen, T.W., Barra, F., Barrow, A., Baskaran, P., Belo, L., Bennett, D.A., Benseñor, I.M., Bhagavathula, A.S., Bhala, N., Bhalla, A., Bhardwaj, N., Bhardwaj, P., Bhaskar, S., Bhattacharyya, K., Bhojaraja, V.S., Bintoro, B.S., Blokhina, E.A.E., Bodicha, B.B.A., Boloor, A., Bosetti, C., Braithwaite, D., Brenner, H., Briko, N.I., Brunoni, A.R., Butt, Z.A., Cao, C., Cao, Y., Cárdenas, R., Carvalho, A.F., Carvalho, M., Castaldelli-Maia, J.M., Castelpietra, G., Castro-de-Araujo, L.F.S., Cattaruzza, M.S., Chakraborty, P.A., Charan, J., Chattu, V.K., Chaurasia, A., Cherbuin, N., Chu, D.-T., Chudal, N., Chung, S.-C., Churko, C., Ciobanu, L.G., Cirillo, M., Claro, R.M., Costanzo, S., Cowden, R.G., Criqui, M.H., Cruz-Martins, N., Culbreth, G.T., Dachew, B.A., Dadras, O., Dai, X., Damiani, G., Dandona, L., Dandona, R., Daniel, B. D., Danielewicz, A., Darega Gela, J., Davletov, K., de Araujo, J.A.P., de Sá-Junior, A. R., Debela, S.A., Dehghan, A., Demetriades, A.K., Derbew Molla, M., Desai, R., Desta, A.A., Dias da Silva, D., Diaz, D., Digesa, L.E., Diress, M., Dodangeh, M., Dongarwar, D., Dorostkar, F., Dsouza, H.L., Duko, B., Duncan, B.B., Edvardsson, K., Ekholuenetale, M., Elgar, F.J., Elhadi, M., Elmonem, M.A., Endries, A.Y., Eskandarieh, S., Etemadimanesh, A., Fagbamigbe, A.F., Fakhradiyev, I.R., Farahmand, F., Farinha, C.S. e S., Faro, Farzadfar, A., Fatehizadeh, F., Fauk, A., Feigin, N.K., Feldman, V.L., Feng, R., Fentaw, X., Ferrero, Z., Ferro Desideri, S., Filip, L., Fischer, I., Francis, F., Franklin, J.M., Gaal, R.C., Gad, P.A., Gallus, M.M., Galvano, S., Ganesan, F., Garg, B., Gebrehiwot, T., Gebremeskel, M.G.D., Gebremichael, T.G., Gemechu, M.A., Getacher, T.R., Getachew, L., Getachew Obsa, M.E., Getie, A., Ghaderi, A., Ghafourifard, A., Ghajar, M., Ghamari, A., Ghandour, S.-H., Ghasemi Nour, L.A., Ghashghaee, M., Ghozy, A., Glozah, S., Glushkova, F.N., Godos, E.V., Goel, J., Goharinezhad, A., Golechha, S., Goleij, M., Golitaleb, P., Greaves, M., Grivna, F., Grosso, M., Gudayu, G., Gupta, T.W., Gupta, B., Gupta, R., Gupta, S., Gupta, V.B., Hafezi-Nejad, V.K., Haj-Mirzaian, N., Hall, A., Halwani, B.J., Handiso, R., Hankey, T.B., Hariri, G.J., Haro, S., Hasaballah, J.M., Hassanian-Moghaddam, A.I., Hay, H., Hayat, S.I., Heidari, K., Heidari, G., Hendrie, M., Herteliu, D., Heyi, C., Hezam, D.Z., Hlongwa, K., Holla, M.M., Hossain, R., Hossain, M.M., Hosseini, S., hosseinzadeh, S.K., M, Hostiuc, M., Hostiuc, S., Hu, G., Huang, J., Hussain, S., Ibitoye, S.E., Ilic, I.M., Ilic, M.D., Immurana, M., Irham, L.M., Islam, M.M., Islam, R.M., Islam, S.M.S., Iso, H., Itumalla, R., Iwagami, M., Jabbarinejad, R., Jacob, L., Jakovljevic, M., Jamalpoor, Z., Jamshidi, E., Jayapal, S.K., Jayarajah, U.U., Jayawardena, R., Jebai, R., Jeddi, S.A., Jema, A.T., Jha, R.P., Jindal, H.A., Jonas, J.B., Joo, T., Joseph, N., Joukar, F., Jozwiak, J.J., Jürisson, M., Kabir, A., Kabthymer, R.H., Kamble, B.D., Kandel, H., Kanno, G.G., Kapoor, N., Karaye, I.M., Karimi, S.E., Kassa, B.G., Kaur, R.J., Kayode, G.A., Keykhaei, M., Khajuria, H., Khalilov, R., Khan, I.A., Khan, M.A., Kim, H., Kim, J., Kim, M.S., Kimokoti, R.W., Kivimäki, M., Klymchuk, V., Knudsen, A.K.S., Kolahi, A.-A., Korshunov, V.A., Koyanagi, A., Krishan, K., Krishnamoorthy, Y., Kumar, G.A., Kumar, Narinder, Kumar, Nithin, Lacey, B., Lallukka, T., Lasrado, S., Lau, J., Lee, S., Lee, W.-C., Lee, Y.H., Lim, L.-L., Lim, S.S., Lobo, S.W., Lopukhov, P.D., Lorkowski, S., Lozano, R., Lucchetti, G., Madadizadeh, F., Madureira-Carvalho, Á.M., Mahjoub, S., Mahmoodpoor, A., Mahumud, R.A., Makki, A., Malekpour, M.-R., Manjunatha, N., Mansouri, B., Mansournia, M.A., Martinez-Raga, J., Martinez-Villa, F.A., Matzopoulos, R., Maulik, P.K., Mayeli, M., McGrath, J.J., Meena, J.K., Mehrabi Nasab, E., Menezes, R. G., Mensink, G.B.M., Mentis, A.-F.A., Meretoja, A., Merga, B.T., Mestrovic, T., Miao Jonasson, J., Miazgowski, B., Micheletti Gomide Nogueira de Sá, A.C., Miller, T.R., Mini, G., Mirica, A., Mirijello, A., Mirmoeeni, S., Mirrakhimov, E.M., Misra, S., Moazen, B., Mobarakabadi, M., Moccia, M., Mohammad, Y., Mohammadi, E., Mohammadian-Hafshejani, A., Mohammed, T.A., Moka, N., Mokdad, A.H., Momtazmanesh, S., Moradi, Y., Mostafavi, E., Mubarik, S., Mullany, E.C., Mulugeta, B.T., Murillo-Zamora, E., Murray, C.J.L., Mwita, J.C., Naghavi, M., Naimzada, M.D., Nangia, V., Nayak, B.P., Negoi, I., Negoi, R.I., Nejadghaderi, S.A., Nepal, S., Neupane, S.P.P., Neupane Kandel, S., Nigatu, Y.T., Nowroozi, A., Nuruzzaman, K.M., Nzoputam, C.I., Obamiro, K.O., Ogbo, F.A., Oguntade, A.S., Okati-Aliabad, H., Olakunde, B.O., Oliveira, G.M.M., Omar Bali, A., Omer, E., Ortega-Altamirano, D. v, Otoiu, A., Otstavnov, S.S., Oumer, B., M, P.A., Padron-Monedero, A., Palladino, R., Pana, A., Panda-Jonas, S., Pandey, Anamika, Pandey, Ashok, Pardhan, S., Parekh, T., Park, E.-K., Parry, C.D.H., Pashazadeh Kan, F., Patel, J., Pati, S., Patton, G.C., Paudel, U., Pawar, S., Peden, A.E., Petcu, I.-R., Phillips, M.R., Pinheiro, M., Plotnikov, E., Pradhan, P.M.S., Prashant, A., Quan, J., Radfar, A., Rafiei, A., Raghav, P.R., Rahimi-Movaghar, V., Rahman, A., Rahman, M. M., Rahman, M., Rahmani, A.M., Rahmani, S., Ranabhat, C.L., Ranasinghe, P., Rao, C.R., Rasali, D.P., Rashidi, M.-M., Ratan, Z.A., Rawaf, D.L., Rawaf, S., Rawal, L., Renzaho, A.M.N., Rezaei, N., Rezaei, S., Rezaeian, M., Riahi, S.M., Romero-Rodríguez, E., Roth, G.A., Rwegerera, G.M., Saddik, B., Sadeghi, E., Sadeghian, R., Saeed, U., Saeedi, F., Sagar, R., Sahebkar, A., Sahoo, H., Sahraian, M.A., Saif-Ur-Rahman, K., Salahi, S., Salimzadeh, H., Samy, A.M., Sanmarchi, F., Santric-Milicevic, M.M., Sarikhani, Y., Sathian, B., Saya, G.K., Sayyah, M., Schmidt, M.I., Schutte, A.E., Schwarzinger, M., Schwebel, D.C., Seidu, A.-A., Senthil Kumar, N., SeyedAlinaghi, S., Seylani, A., Sha, F., Shahin, S., Shahraki-Sanavi, F., Shahrokhi, S., Shaikh, M.A., Shaker, E., Shakhmardanov, M.Z., Shams-Beyranvand, M., Sheikhbahaei, S., Sheikhi, R.A., Shetty, A., Shetty, J.K., Shiferaw, D.S., Shigematsu, M., Shiri, R., Shirkoohi, R., Shivakumar, K.M., Shivarov, V., Shobeiri, P., Shrestha, R., Sidemo, N.B., Sigfusdottir, I.D., Silva, D.A.S., Silva, N.T. da, Singh, Singh, J.A., Skryabin, S., Skryabina, V.Y., Sleet, A.A., Solmi, D.A., SOLOMON, M., Song, Y., Song, S., Sorensen, Y., Soshnikov, R.J.D., Soyiri, S., Stein, I.N., Subba, D.J., Szócska, S.H., Tabarés-Seisdedos, M., Tabuchi, R., Taheri, T., Tan, M., Tareke, K.-K., Tarkang, M., Temesgen, E.E., Temesgen, G., Temsah, W.A., Thankappan, M.-H., Thapar, K.R., Thomas, R., Tiruneh, N.K., Todorovic, C., Torrado, J., Touvier, M., Tovani-Palone, M., Tran, M.R., Trias-Llimós, M.T.N., Tripathy, S., Vakilian, J.P., Valizadeh, A., Varmaghani, R., Varthya, M., Vasankari, S.B., Vos, T.J., Wagaye, T., Waheed, B., Walde, Y., Wang, M.T., Wang, C., Wang, Y., Westerman, Y.-P., Wickramasinghe, R., Wubetu, N.D., Xu, A.D., Yamagishi, S., Yang, K., Yesera, L., Yigit, G.E.E., Yiğit, A., Yimaw, V., Yon, A.E.A.E., Yonemoto, D.K., Yu, N., Zadey, C., Zahir, S., Zare, M., Zastrozhin, I., Zastrozhina, M.S., Zhang, A., Zhong, Z.-J., Zmaili, C., Zuniga, M., Gakidou, Y.M.H., E, 2022. Population-level risks of alcohol consumption by amount, geography, age, sex, and year: a systematic analysis for the Global Burden of Disease Study 2020. Lancet 400, 185–235. https://doi.org/10.1016/S0140-6736(22)00847-9.

Calder, P.C., Ahluwalia, N., Albers, R., Bosco, N., Bourdet-Sicard, R., Haller, D., Holgate, S.T., Jönsson, L.S., Latulippe, M.E., Marcos, A., Moreines, J., Mrini, C., Müller, M., Pawelec, G., van Neerven, R.J.J., Watzl, B., Zhao, J., 2013. A consideration of biomarkers to be used for evaluation of inflammation in human nutritional studies. Br. J. Nutr. 109. https://doi.org/10.1017/S0007114512005119.

Chu, L., Ioannidis, J.P.A., Egilman, A.C., Vasiliou, V., Ross, J.S., Wallach, J.D., 2020. Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk. Int. J. Epidemiol. 49, 608–618. https://doi.org/10.1093/ije/dyz271.

Coventry, B.J., Ashdown, M.L., Quinn, M.A., Markovic, S.N., Yatomi-Clarke, S.L., Robinson, A.P., 2009. CRP identifies homeostatic immune oscillations in cancer patients: a potential treatment targeting tool? J. Transl. Med. 7, 1–8.

de Timary, P., Stärkel, P., Delzenne, N.M., Leclercq, S., 2017. A role for the peripheral immune system in the development of alcohol use disorders? Neuropharmacology 122, 148–160.

Department of Health, 2016. UK Chief Medical Officers' Low Risk Drinking Guidelines.

Elliott, J., Shepherd, P., 2006. Cohort profile: 1970 British Birth Cohort (BCS70). Int. J. Epidemiol. 35, 836–843. https://doi.org/10.1093/ije/dyl174.

Furman, D., Campisi, J., Verdin, E., Carrera-Bastos, P., Targ, S., Franceschi, C., Ferrucci, L., Gilroy, D.W., Fasano, A., Miller, G.W., Miller, A.H., Mantovani, A., Weyand, C.M., Barzilai, N., Goronzy, J.J., Rando, T.A., Effros, R.B., Lucia, A., Kleinstreuer, N., Slavich, G.M., 2019. Chronic inflammation in the etiology of disease across the life span. Nat. Med. 25, 1822–1832. https://doi.org/10.1038/s41591-019-0675-0.

Giollabhui, N.M., Ellman, L.M., Coe, C.L., Byrne, M.L., Abramson, L.Y., Alloy, L.B., 2020. To exclude or not to exclude: considerations and recommendations for C-reactive protein values higher than 10 mg/L. Brain Behav. Immun. 87, 898.

González-Reimers, E., Santolaria-Fernández, F., Martín-González, M.C., Fernández-Rodríguez, C.M., Quintero-Platt, G., 2014. Alcoholism: a systemic proinflammatory condition. World J. Gastroenterol. https://doi.org/10.3748/wjg.v20.i40.14660.

Hamer, M., O'Donovan, G., Batty, G.D., Stamatakis, E., 2020. Estimated cardiorespiratory fitness in childhood and cardiometabolic health in adulthood: 1970 British Cohort Study. Scand. J. Med. Sci. Sports 30, 932–938. https://doi.org/10.1111/sms.13637.

He, J., Crews, F.T., 2008. Increased MCP-1 and microglia in various regions of the human alcoholic brain. Exp. Neurol. 210, 349–358. https://doi.org/10.1016/j.expneurol.2007.11.017.

Health Council of the Netherlands, 2015. Dutch Dietary Guidelines 2015.

Hill, A.B., 1965. The environment and disease: association or causation? Proc. R. Soc. Med. 58, 295–300.

Klau, S., Hoffmann, S., Patel, C.J., Ioannidis, J.P.A., Boulesteix, A.L., 2021. Examining the robustness of observational associations to model, measurement and sampling uncertainty with the vibration of effects framework. Int. J. Epidemiol. 50, 266–278. https://doi.org/10.1093/ije/dyaa164.

Krenz, M., Korthuis, R.J., 2012. Moderate ethanol ingestion and cardiovascular protection: from epidemiologic associations to cellular mechanisms. J. Mol. Cell Cardiol. 52, 93–104.

Masur, P.K., Scharkow, M., 2020. specr: Conducting and Visualizing Specification Curve Analyses.

Menezes, A.M.B., Oliveira, P.D., Wehrmeister, F.C., Assunção, M.C.F., Oliveira, I.O., Tovo-Rodrigues, L., Ferreira, G.D., Gonçalves, H., 2019. Association of modifiable risk factors and IL-6, CRP, and adiponectin: findings from the 1993 birth cohort, Southern Brazil. PLoS One 14. https://doi.org/10.1371/journal.pone.0216202.

Muralidharan, S., Ambade, A., Fulham, M.A., Deshpande, J., Catalano, D., Mandrekar, P., 2014. Moderate alcohol induces stress proteins HSF1 and hsp70 and inhibits proinflammatory cytokines resulting in endotoxin tolerance. J. Immunol. 193, 1975–1987. https://doi.org/10.4049/jimmunol.1303468.

Naimi, T., Chikritzhs, T., Stockwell, T., 2022. Commentary on Di Castelnuovo et al.: implications of using low volume drinkers instead of never drinkers as the reference group. Addiction 117, 327–329.

Nicholls, J., 2012. Time for reform? Alcohol policy and cultural change in England since 2000. Br. Polit. 7, 250–271.

Pai, J.K., Hankinson, S.E., Thadhani, R., Rifai, N., Pischon, T., Rimm, E.B., 2006. Moderate alcohol consumption and lower levels of inflammatory markers in US men and women. Atherosclerosis 186, 113–120. https://doi.org/10.1016/j.atherosclerosis.2005.06.037.

Patel, C.J., Burford, B., Ioannidis, J.P.A., 2015. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J. Clin. Epidemiol. 68, 1046–1058. https://doi.org/10.1016/j.jclinepi.2015.05.029.

Paulson, D., Shah, M., Herring, D., Scott, R., Herrera, M., Brush, D., Bassett, R., 2018. The relationship between moderate alcohol consumption, depressive symptomatology, and C-reactive protein: the Health and Retirement Study. Int. J. Geriatr. Psychiatry 33, 316–324. https://doi.org/10.1002/gps.4746.

Rehm, J., Soerjomataram, I., Ferreira-Borges, C., Shield, K.D., 2019. Does alcohol use affect cancer risk? Curr. Nutr. Rep. 8, 222–229.

Réus, G.Z., Fries, G.R., Stertz, L., Badawy, M., Passos, I.C., Barichello, T., Kapczinski, F., Quevedo, J., 2015. The role of inflammation and microglial activation in the pathophysiology of psychiatric disorders. Neuroscience. https://doi.org/10.1016/j.neuroscience.2015.05.018.

Ridker, P.M., 2004. Inflammation in atherothrombosis: how to use high-sensitivity C-reactive protein (hsCRP) in clinical practice. Am. Heart Hosp. J. 2, 4–9.

Rohleder, N., 2019. Stress and inflammation – the need to address the gap in the transition between acute and chronic stress effects. Psychoneuroendocrinology. https://doi.org/10.1016/j.psyneuen.2019.02.021.

Sarma, M.A., 2021. Package 'Multiverse.'

Silverwood, R.J., Holmes, M. v, Dale, C.E., Lawlor, D.A., Whittaker, J.C., Smith, G.D., Leon, D.A., Palmer, T., Keating, B.J., Zuccolo, L., 2014. Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. Int. J. Epidemiol. 1781–1790. https://doi.org/10.1093/ije/dyu187.

Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22, 1359–1366. https://doi.org/10.1177/0956797611417632.

Simonsohn, U., Simmons, J.P., Nelson, L.D., 2020. Specification curve analysis. Nat. Hum. Behav. 4, 1208–1214. https://doi.org/10.1038/s41562-020-0912-z.

Spanish Ministry of Health, 2020. Low Risk Alcohol Consumption Thresholds. Update on the Risks Related to Alcohol Consumption Levels, Consumption Patterns and Type of Alcoholic Beverages.

Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing transparency through a multiverse analysis. Perspect. Psychol. Sci. 11, 702–712. https://doi.org/10.1177/1745691616658637.

Stockwell, T., Zhao, J., Panwar, S., Roemer, A., Naimi, T., Chikritzhs, T., 2016. Do "moderate" drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. J. Stud. Alcohol Drugs 77, 185–198.

Szabo, G., 2007. Moderate drinking, inflammation, and liver disease. Ann. Epidemiol. 17. https://doi.org/10.1016/j.annepidem.2007.01.012.

U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2020. Dietary Guidelines for Americans, 2020-2025. ninth edition.

van't Klooster, C.C., van der Graaf, Y., Ridker, P.M., Westerink, J., Hjortnaes, J., Sluijs, I., Asselbergs, F.W., Bots, M.L., Kappelle, L.J., Visseren, F.L.J., 2020. The relation between healthy lifestyle changes and decrease in systemic inflammation in patients with stable cardiovascular disease. Atherosclerosis 301, 37–43. https://doi.org/10.1016/j.atherosclerosis.2020.03.022.

Visontay, R., Sunderland, M., Slade, T., Wilson, J., Mewton, L., 2022. Are there non-linear relationships between alcohol consumption and long-term health? A systematic review of observational studies employing approaches to improve causal inference. BMC Med. Res. Methodol. 22, 1–28.

Volpato, S., Pahor, M., Ferrucci, L., Simonsick, E.M., Guralnik, J.M., Kritchevsky, S.B., Fellin, R., Harris, T.B., 2004. Relationship of alcohol intake with inflammatory markers and plasminogen activator inhibitior-1 in well-functioning older adults: the health, aging, and body composition study. Circulation 109, 607–612. https://doi.org/10.1161/01.CIR.0000109503.13955.00.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P., Initiative, S., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann. Intern. Med. 147, 573–577.

Appendix D

Alcohol in early-to-middle adulthood and midlife depression

Preface

This study was published as **Visontay, R.**, Mewton, L., M., Slade, T., Aris, I. M., & Sunderland, M. (2023). Moderate alcohol consumption and depression: A marginal structural model approach promoting causal inference. *American Journal of Psychiatry, 180*(3), 209-17.

RV conceptualised the study with assistance from LM, TS, and MS. RV conducted the analyses and drafted the manuscript. All authors critically revised the manuscript and approved the final version.

Moderate Alcohol Consumption and Depression: A Marginal Structural Model Approach Promoting Causal Inference

Rachel Visontay, B.A.(Hons), Louise Mewton, Ph.D., Tim Slade, Ph.D., Izzuddin M. Aris, Ph.D., Matthew Sunderland, Ph.D.

Objective: Moderate alcohol consumption is associated with decreased risk for depression, but it remains unclear whether this is a causal relationship or a methodological artifact. To compare the effects of consistent abstinence and occasional, moderate, and above-guideline alcohol consumption throughout early to middle adulthood on depression at age 50, the authors conducted a secondary analysis of the National Longitudinal Survey of Youth 1979 cohort and employed a marginal structural model (MSM) approach.

Methods: Baseline was set at 1994, when individuals were ages 29–37. The MSM incorporated measurements of alcohol consumption in 1994, 2002, and 2006, baseline and time-varying covariates, and repeated measurements with the Center for Epidemiologic Studies Depression Scale–Short Form (CES-D-SF). A total of 5,667 eligible participants provided valid data at baseline, 3,593 of whom provided valid outcome data. The authors used all observed data to predict CES-D-SF means and rates of probable depression for hypothetical trajectories of consistent alcohol consumption.

Results: The results approximated J-curve relationships. Specifically, both consistent occasional and consistent moderate drinkers were predicted to have reduced CES-D-SF scores and rates of probable depression at age 50 compared with consistent abstainers (CES-D-SF scores: b=−0.84, 95% CI=−1.47, −0.11; probable depression: odds ratio=0.58, 95% CI=0.36, 0.88 for consistent occasional drinkers vs. abstainers; CES-D-SF scores: b=−1.08, 95% CI=−1.88, −0.20; probable depression: odds ratio=0.59, 95% CI=0.26, 1.13 for consistent moderate drinkers vs. consistent abstainers). Consistent above-guideline drinkers were predicted to have slightly increased risk compared with consistent abstainers, but this was not significant. In sex-stratified analyses, results were similar for females and males.

Conclusions: This study contributes preliminary evidence that associations between moderate alcohol consumption and reduced risk for depression may reflect genuine causal effects. Further research using diverse methodologies that promote causal inference is required.

AJP in Advance (doi: 10.1176/appi.ajp.22010043)

Depression contributes substantially to disease burden, particularly in people in the 25–49 age range, where depressive disorders constitute the sixth most burdensome condition globally (1). This burden is likely to be exacerbated as incidence increases, and cases rose by roughly 50% worldwide from 1990 to 2017 (2). On an individual level, depression impairs quality of life, heightens suicide risk (3), and increases risk for later-life conditions such as cardiovascular disease and associated morbidity and mortality (4). Prevention of depressive symptoms and disorders is therefore a public health priority. However, this relies on the accurate identification of modifiable risk factors for the condition—an area where current understanding is limited (5).

Alcohol consumption is one candidate worthy of further study, given the well-established comorbidity between alcohol use disorders and depression (6, 7). Specifically, alcohol use disorders precede major depression (6, 8, 9), with some additional evidence for bidirectionality (10) (i.e., depression also being associated with subsequent alcohol use disorders), consistent with the "self-medication" hypothesis (11, 12). However, for non-disordered drinking at a population level, the nature of the alcohol-depression relationship is less clear. There are indications that light to moderate consumption is associated with lower risk for depressive conditions when compared with abstinence, with risk increasing again for heavier drinkers, resulting in a J-shaped relationship—a finding supported by a recent meta-analysis (9). However, the extent to which these protective effects are genuine causal relationships, as opposed to biased associations driven by methodological limitations, has not been established.

One such methodological bias may be confounding, given the possibility that background environmental (e.g., socioeconomic) factors may explain demonstrated alcohol-depression associations (13). Most studies employ a "conditional" approach when controlling for confounding, that is, adjusting for covariates in a regression model. To be effective, this approach requires accurate model specification, including the shape of the relationship of each covariate with the outcome (e.g., linear, exponential, etc.) (14–16). Even then, this approach is limited when there are few individuals for certain variable combinations (15, 17, 18), and it may be biased when there exists a time-varying factor that can act as both a confounder and a mediator over time (18).

For example, income at some time during the exposure period may be a common cause of later alcohol consumption and depression and may itself be affected by past alcohol consumption. Conditional approaches that control for income at various waves of a study will give biased estimates by removing (i.e., adjusting away) the effects of prior alcohol consumption mediated through income. On the other hand, simply not adjusting for income would lead to results biased by confounding. See Figure S1 in the online supplement for a visualization of these complex longitudinal relationships.

In addition to confounding, it is also difficult to isolate effects in a single direction, with reverse causation often present in the form of "sick quitters"—those for whom illness precipitates abstinence from or a reduction in alcohol consumption (19). Additionally, selection biases may be at play, such as healthier drinkers being overrepresented in recruited cohorts (20), and/or those with poorer outcomes being more likely to drop out over the follow-up period. Finally, categorization of alcohol consumption is often based on a single baseline measurement, which ignores changes in consumption ("drinker drift") during the period to outcome measurement (21).

Modern statistical methods for causal inference, such as marginal structural models (MSMs), overcome these limitations. MSMs attempt to fully adjust for measured confounders (both time-invariant and time-varying) through a "marginal" approach, that is, balancing confounders across exposure groups (i.e., different levels of alcohol consumption) before regression modeling, accomplished by assigning weights to each observation. These weights reflect the extent to which each observation is under- or overrepresented in the sample, with respect to a target population in which potential confounders are balanced across exposure groups. Applying these weights to the sample creates a "pseudo population" in which the distribution of confounders between exposure groups at each individual wave is balanced, while still allowing for longitudinal mediation via these confounders (22).

Thus, relying on certain assumptions (see section 6 in the online supplement), the pseudo population should be free of confounding and can be used to approximate randomized controlled trials—the gold standard for assessing causality—in contexts where their implementation is not feasible (23). This approach can address many of the limitations of conventional cohort studies (see Table S1 in the online supplement). Our recent review (24) identified a single study applying MSMs to the alcohol-depression relationship. In a sample ages 20–64, Gemes et al. (25) found a U-shaped relationship with major depression after 7–9 years of follow-up.

However, to determine whether the Gemes et al. findings are truly robust to the methodological biases outlined above, an extension of the application of MSMs is required. First, to utilize the full potential of MSMs and acknowledge drinker drift, multiple waves of exposure and covariate information should be included. Additionally, it is vital that a substantial portion of participants' drinking (or abstinence) histories are incorporated, lest the complex cumulative and/or delayed effects of alcohol use (26) or the effects of sick quitters adulterate estimates of the targeted exposure's effect. Similarly, given that the nature of alcohol-depression relationships may change over the lifespan (27), estimates should be derived from a sample of a fairly narrow age range so as not to mask underlying heterogeneity. Finally, exposure weights should only incorporate time-varying covariates that are known to temporally precede exposure.

Questions about the protective effects of moderate consumption tend to imply a stable level of consumption (e.g., "Does moderate consumption confer benefits over abstinence?"), allowing for simple, interpretable public health messages. Therefore, it is important to estimate the effects of *consistent* levels of consumption. With a secondary analysis of the National Longitudinal Survey of Youth 1979 (NLSY79), a cohort offering rich longitudinal data, our aim in the present study was to employ a robust MSM approach in addressing the following research question: Compared with consistent abstinence from alcohol, what is the effect of consistent occasional, moderate, and above-guideline alcohol consumption throughout early to middle adulthood on depression at age 50? Given evidence that methodological biases are responsible for previously identified J-shaped relationships, it was hypothesized that the use of a sophisticated MSM approach—reducing the impact of these biases—would result in a positive, linear relationship.

METHODS

Study Design and Participants

The NLSY79 is a nationally representative U.S. cohort of 12,686 individuals born between 1957 and 1964 (the sampling procedure has been described elsewhere [28]). Participants were first interviewed in 1979, at ages 14–22, and data on depressive symptoms (as indexed by the Center for Epidemiologic Studies Depression Scale–Short Form [CES-D-SF]) were first collected in 1992. Given the need to control for pre-baseline depressive symptoms, baseline for the present study was set at the following measurement occasion, which was in 1994. Eligible individuals were those who had valid baseline alcohol data and both time-fixed and pre-baseline (1992) time-varying covariate data (including CES-D-SF values),

TABLE 1. Criteria for alcohol consumption categories

Category	Criteria
Abstinence	No alcohol consumption
Occasional consumption	Average frequency <1 day/week; no heavy episodic drinking
Moderate consumption	Average frequency ≥1 day/week; average weekly drinks, ≤7 for females, ≤14 for males; no heavy episodic drinking
Above-guideline consumption	Average frequency ≥1 day/week; average weekly drinks, ≥7 for females, ≥14 for males; and/or heavy episodic drinking

resulting in 5,667 participants. All NLSY79 participants provided informed consent.

Guided by the timing of CES-D-SF readministration, which changed after 1994 from year-based measurement to age intervals, the MSM incorporated alcohol consumption in 1994, 2002, and 2006, covariates and CES-D-SF values in 1992, in 1994, and at age 40, and a final CES-D-SF outcome measured at age 50 (see section 1 and Figure S1 in the online supplement for more detail and a visualization, respectively, of what data were used from which assessment points). Thus, the alcohol consumption variables span ages 29–37 through 41–49, corresponding with early to middle adulthood. In addition to those who dropped out of the study entirely, individuals who were missing any alcohol consumption, covariate, or CES-D-SF values in 2002 or 2006, or CES-D-SF values at age 50, were considered censored at the earliest of those assessments.

Variable Derivation

Alcohol consumption. Participants were asked about their current drinking frequency (number of drinking days in the previous month), volume (number of drinks per drinking day, with a drink explicitly defined as "the equivalent of a can of beer, a glass of wine, or a shot glass of hard liquor"), and heavy episodic drinking (defined as at least six drinks per occasion by the NLSY79 questionnaire). For this study, alcohol consumption was separately categorized for 1994, 2002, and 2006, with categorization based on methods (29) incorporating frequency, volume, and heavy episodic drinking, modified such that volume criteria were concordant with current U.S. guidelines (see Table 1 for category criteria) (30). For every pre-baseline occasion with adequate information, individuals were similarly categorized, and wave-specific categorizations were condensed into a single historical variable (see section 2 in the online supplement).

Covariates and CES-D-SF scores. The CES-D was designed to measure depressive symptoms in the general population and asks respondents to indicate the frequency with which they experienced a range of symptoms during the previous week, with higher scores suggesting greater symptoms (31). The CES-D-SF is a reduced, seven-item version of the CES-D and has demonstrated strong psychometric properties (32). Two versions of the analyses were performed: one in which CES-D-SF scores were analyzed in their continuous form and another in which the scale was dichotomized into non-depression (scores <8) or probable depression (scores ≥8) based on suggested cutoffs for probable depression (32). The distribution of the outcome variable was not sufficiently skewed to warrant transformation (see section 3 in the online supplement) (33).

Covariate selection was guided by a recent meta-analysis (9). The baseline time-fixed variables incorporated were historical alcohol consumption, sex, ever smoked, ever used illicit drugs, pre-baseline self-esteem, pre-baseline frequent religious service attendance, race, educational attainment, and average parental educational attainment. The time-varying variables incorporated were CES-D-SF scores, age, self-reported health limitations, marital status, smoking status (not available at age 40), illicit drug use (not available at age 40), body mass index, employment status, health insurance status, household size, urban/rural residence, welfare receipt, and income. Note that the incorporation of CES-D-SF scores prior to the final outcome measurement means that the model specifically isolates the effect of alcohol consumption on age-50 depression. Further detail on covariate selection and derivation is presented in sections 2, 4, and 5 and Figure S2 in the online supplement.

Statistical Analysis

Marginal structural models (MSMs). The first step in constructing MSMs is to sequentially estimate inverse probability of treatment weights (IPTWs) for each wave—that is, the probability of an individual being in a given treatment (or exposure) group given how their covariate profile compares to the typical profile for that group. Individuals whose covariate profiles are atypical of their exposure category are assigned larger weights than those whose covariate profiles align more closely with their group's typical profile. Inverse probability of censoring weights (IPCWs) are calculated at each wave too, such that those whose covariate profiles are not typical of those who remained in the study are upweighted. Combined with the use of a survey weight, IPCWs allow MSMs to be representative of the target population. The final weight for each individual is the product of all time-specific IPTWs, IPCWs, and the survey weight.

In the second step, the weighted pseudo population is statistically analyzed, comparing hypothetical joint interventions (i.e., exposure trajectories of interest) in their effect on an outcome. Such an analysis provides marginal estimates that apply to the whole population because these estimates are not conditional on other variables (i.e., they are not based on models that hold covariates constant) (34).

Weight generation. Stabilized IPTWs were created for alcohol consumption at the 1994, 2002, and 2006 waves, as were IPCWs (relating to attrition status in 2002 and 2006 and at age 50). The final MSM weight was trimmed at the 98th percentiles to mitigate the effect of extreme weights (35). Further information on weight calculation, distribution, and

covariate balance is presented in section 7 and Figure S3 in the online supplement.

Linear regression and contrasts. Using these final weights, weighted multivariable regression models regressed CES-D-SF scores (linear model) and probable depression (logistic model) at age 50 on 1994, 2002, and 2006 alcohol consumption. The first set of analyses, using a continuous CES-D-SF outcome, provides predicted mean depressive symptoms, while the second, using a binary outcome, provides predicted probabilities of probable depression. In results focusing on stable trajectories of alcohol consumption, drinking level "values" were substituted into a model using the parameters derived from the whole sample to generate predictions about *hypothetical* drinking trajectories of interest. This involved using linear contrasts to estimate the predicted difference in outcome between hypothetical trajectories of consistent consumption. Specifically, consistent occasional, consistent moderate, and consistent above-guideline drinking were compared against a consistent abstinence reference category. We selected these stable drinking categories because they are particularly informative in terms of formulating national alcohol guidelines. For the continuous outcome, this is equivalent to summing the relevant wave-specific effect estimates for each category of alcohol consumption from the initial regression, and then taking the difference with the reference group. Results were bootstrapped using 500 bootstrap replicates to account for the additional sources of uncertainty introduced by weighting.

Sensitivity analyses. Despite incorporating a wide range of potential confounders, E-values were calculated for the contrast estimates to assess robustness to unmeasured confounding. E-values are a recently developed metric intended to represent the effect size that an excluded confounder must have on both exposure and outcome to account for the observed association (36). They are calculated using the standardized effect estimates and their standard errors/confidence intervals for an observed association. A small E-value suggests that only a small amount of unmeasured confounding would be needed to account for the association. Conversely, as E-values get larger, the likelihood that unmeasured confounding fully accounts for the association becomes increasingly unlikely. The main analyses were additionally performed in men and women separately.

All analyses were conducted in R, version 4.1.0 (see section 8 in the online supplement for packages and code used). We considered p values <0.05 statistically significant, but we also interpreted results with respect to strength of the effect estimates and confidence intervals.

RESULTS

Baseline Sample Characteristics

The sample was 50.5% female, and in 1992 had a mean age of 30.81 years. At baseline, 1,947 participants abstained from alcohol, 1,259 consumed occasionally, 660 consumed moderately, and 1,801 consumed above guidelines. The mean CES-D-SF score in 1992 was 4.16 (SD=4.01), with baseline abstainers having the highest mean (4.55; SD=4.36) and baseline moderate drinkers the lowest (3.17; SD=3.33). Compared with abstainers and above-guideline drinkers, occasional and moderate drinkers had a higher mean income at baseline, were less likely to be Black, and were more likely to be employed and to be married. Additional baseline sample characteristics are presented in Table 2 (note that IPTW weighting effectively nullified baseline differences between exposure groups; see Figure S3 in the online supplement).

Primary Analyses

A total of 3,593 individuals provided valid outcome data at age 50. The results of the regression models are presented in Table 3. For both the continuous and binary outcomes, the largest protective effects for both occasional drinking (b=−0.61, bootstrap 95% CI=−1.19, −0.01; odds ratio=0.70, bootstrap 95% CI=0.46, 1.04) and moderate drinking (b=−0.82, bootstrap 95% CI=−1.46, −0.13; odds ratio=0.58, bootstrap 95% CI=0.29, 1.00) corresponded to 1994 consumption, while the greatest detrimental effects for above-guideline drinking corresponded to 2006 consumption (b=0.55, bootstrap 95% CI=−0.98, 0.40; odds ratio=1.17, bootstrap 95% CI=0.72, 1.96).

Results from the contrast analyses approximated a J-shape (see Figures 1 and 2): after bootstrapping, consistent occasional drinkers (b=−0.84, 95% CI=−1.47, −0.11) and consistent moderate drinkers (b=−1.08, 95% CI=−1.88, −0.20) were both predicted to have statistically significantly lower mean CES-D-SF values at age 50 than consistent abstainers. Consistent above-guideline drinkers had greater predicted CES-D-SF values than consistent abstainers, but the difference was not statistically significant (b=0.34, 95% CI=−0.47, 1.25). Similar results were obtained for the binary outcome analyses, with consistent occasional drinkers predicted to have significantly lower odds of depression than consistent abstainers at age 50 (odds ratio=0.58, 95% CI=0.36, 0.88) and consistent above-guideline drinkers having marginally higher odds (odds ratio=1.06, 95% CI=0.66, 1.72). Consistent moderate drinkers had reduced odds of developing probable depression, comparable to that of occasional drinkers, but this was not significant after bootstrapping (odds ratio=0.59, 95% CI=0.26, 1.13).

See Tables S2 and S3 in the online supplement for full statistical output from contrast analyses.

Sensitivity analyses. E-values for the consistent abstinence versus consistent moderate consumption contrast in the continuous outcome analyses were 1.83 (for a standardized β of −0.25) and 1.30 (for the confidence interval). This indicates that to fully account for the observed association, one or more unmeasured confounders would need to almost double both the probability of an individual belonging to the moderate exposure group and the probability of being high (compared to low) on

TABLE 2. Selected baseline characteristics of eligible participants in a nationally representative U.S. cohort, by alcohol consumption at baseline[a]

Characteristic	Abstainer (N=1,947)		Occasional (N=1,259)		Moderate (N=660)		Above Guidelines (N=1,801)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CES-D-SF score	4.55	4.36	3.83	3.75	3.17	3.33	4.32	3.93
Age (years)	30.81	2.24	30.90	2.22	31.03	2.30	30.67	2.21
Income (US$)	16,005	12,938	20,745	13,213	24,084	15,567	17,541	13,408
Body mass index	26.48	5.49	25.65	5.03	24.72	4.32	26.18	4.66
	N	%	N	%	N	%	N	%
Sex (female)	1,187	61.0	778	61.8	305	46.2	592	32.9
Historical alcohol consumption								
Lifetime abstainer	80	4.1	1	0.1	0	0.0	0	0.0
Past occasional drinker	502	25.8	238	18.9	30	4.5	53	2.9
Past moderate drinker	196	10.1	201	16.0	134	20.3	59	3.3
Past above-guideline drinker	1,169	60.0	819	65.1	496	75.2	1,689	93.8
Race								
Black	617	31.7	298	23.7	144	21.8	476	26.4
Hispanic	376	19.3	256	20.3	77	11.7	324	18.0
Non-Black and non-Hispanic	954	49.0	705	56.0	439	66.5	1,001	55.6
Current smoking	615	31.6	357	28.4	178	27.0	885	49.1
Unemployed	579	29.7	241	19.1	81	12.3	367	20.4
Marital status								
Married	1,077	55.3	778	61.8	396	60.0	826	45.9
Separated, divorced, or widowed	340	17.5	162	12.9	97	14.7	334	18.5
Never married	530	27.2	319	25.3	167	25.3	641	35.6

[a] Baseline alcohol categorization is based on 1994 values; time-varying characteristics are based on 1992 values. CES-D-SF=Center for Epidemiologic Studies Depression Scale–Short Form.

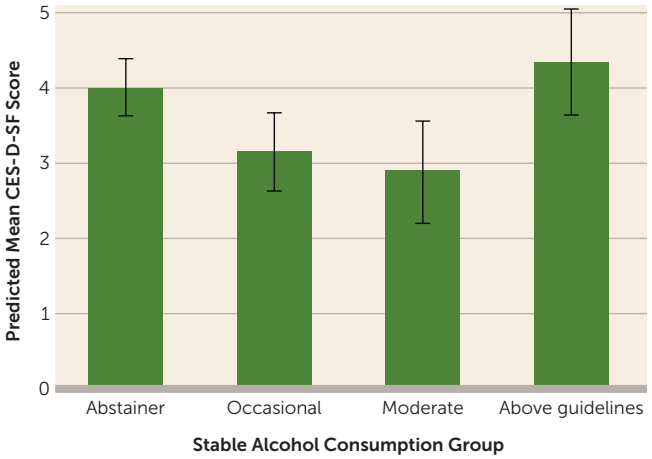
TABLE 3. Results of outcome regressions of age-50 CES-D-SF score on alcohol category predictors in a nationally representative U.S. cohort

	Mean CES-D-SF Score Analyses				Probable Depression Analyses			
Predictor	Estimate (b)[a]	95% CI	p	Bootstrap 95% CI	Estimate (Odds Ratio)[a]	95% CI	p	Bootstrap 95% CI
Intercept	3.99	3.70, 4.28	<0.001	3.63, 4.39	0.24	0.20, 0.28	<0.001	0.18, 0.30
1994_occasional	−0.61	−1.01, −0.21	0.003	−1.19, −0.01	0.70	0.55, 0.89	0.004	0.46, 1.04
1994_moderate	−0.82	−1.33, −0.32	0.001	−1.46, −0.13	0.58	0.42, 0.79	0.001	0.29, 1.00
1994_above-guideline	−0.11	−0.52, 0.30	0.596	−0.71, 0.50	0.95	0.75, 1.19	0.630	0.66, 1.32
2002_occasional	−0.01	−0.40, 0.39	0.977	−0.61, 0.62	0.99	0.78, 1.26	0.957	0.64, 1.43
2002_moderate	−0.01	−0.49, 0.47	0.974	−0.67, 0.76	1.34	1.01, 1.76	0.042	0.80, 2.15
2002_above-guideline	−0.09	−0.59, 0.41	0.718	−0.88, 0.80	0.96	0.71, 1.29	0.788	0.55, 1.57
2006_occasional	−0.22	−0.61, 0.16	0.259	−0.73, 0.34	0.83	0.65, 1.04	0.111	0.55, 1.25
2006_moderate	−0.25	−0.71, 0.21	0.290	−0.98, 0.40	0.77	0.58, 1.01	0.065	0.45, 1.29
2006_above-guideline	0.55	0.04, 1.06	0.036	−0.24, 1.43	1.17	0.87, 1.55	0.296	0.72, 1.96

[a] The reference group for each measurement year is abstinence. The analysis was based on 3,593 observations. For the CES-D-SF score analyses, R^2=0.011, adjusted R^2=0.009, p<0.0001. CES-D-SF=Center for Epidemiologic Studies Depression Scale–Short Form.

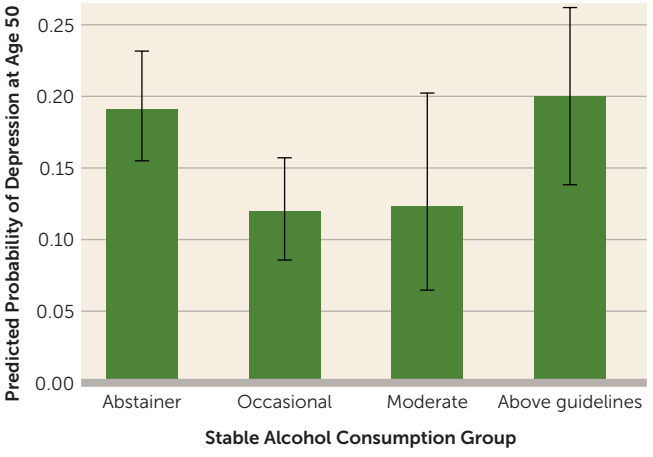
CES-D-SF scores (with a smaller effect required to alter the confidence interval such that it includes the null). The E-value was similar for the binary outcome analyses (1.93). For the consistent abstinence versus consistent occasional consumption contrast, the E-values for mean depression score were 1.68 and 1.19 for a standardized β of −0.20 and confidence interval, respectively, and for the binary outcome were 1.95 for the effect and 1.33 for the confidence interval.

FIGURE 1. Predicted mean CES-D-SF scores at age 50 for individuals consistently abstaining or drinking occasionally, moderately, or above guidelines over the 1994, 2002, and 2006 measurement waves in a representative U.S. sample[a]



[a] Groups do not possess a specific number of subjects as they represent hypothetical trajectories. Error bars indicate bootstrapped 95% confidence intervals. CES-D-SF=Center for Epidemiologic Studies Depression Scale–Short Form. To view predicted means for all hypothetical trajectories, please visit the R Shiny applications that accompany this article (males: https://rachelvisontay.shinyapps.io/Men_continuous/; females: https://rachelvisontay.shinyapps.io/Women_continuous/).

FIGURE 2. Predicted probability of probable depression at age 50 for individuals consistently abstaining or drinking occasionally, moderately, or above guidelines over the 1994, 2002, and 2006 measurement waves in a representative U.S. sample[a]



[a] Groups do not possess a specific number of subjects as they represent hypothetical trajectories. Error bars indicate bootstrapped 95% confidence intervals. To view predicted probabilities for all hypothetical trajectories, please visit the R Shiny applications that accompany this article (males: https://rachelvisontay.shinyapps.io/Men_probable_depression/; females: https://rachelvisontay.shinyapps.io/Women_probable_depression/).

Women were less likely to be above-guideline drinkers than men (see Table 2) and had higher depressive symptom levels overall: at age 50 the mean CES-D-SF score for women was 4.20 (SD=4.48), compared with 3.05 (SD=3.91) for men. However, in stratified analyses, the pattern of results from the overall sample replicated for both men and women (see section 9 and Figures S4 and S5 in the online supplement). This was true of both the continuous and binary outcome analyses, although for the latter, consistent moderate drinking resulted in the lowest predicted probability for males rather than being equivalent with consistent occasional drinking, as it was for females.

DISCUSSION

This MSM analysis of the relationship between alcohol consumption in early to middle adulthood and depression at age 50 replicated the classic J-shaped relationship reported in the conventional epidemiological literature. In the present study, consistent occasional or moderate drinking was predicted to result in lower CES-D-SF mean scores and reduced odds of probable depression than consistent abstinence.

This finding was contrary to the hypothesis that an MSM using methods that address potential biases would result in a positive, linear relationship between alcohol use and depression. That protective effects of low to moderate consumption were found even after accounting for common biases offers preliminary evidence that these effects may be causal. Unlike the harmful effects of above-guideline consumption, which were largest at the measurement wave just prior to outcome measurement, alcohol consumption at baseline contributed most strongly to the protective effects of occasional and moderate drinking. (While moderate drinking at the 2002 wave was associated with *increased* risk in the binary outcome analyses, this estimate had wide confidence intervals after bootstrapping.) This suggests that the benefits of low-level drinking accrue over time. Various mechanisms have been suggested for these benefits, including possible biological mediation via both GABAergic (37, 38) and dopaminergic effects (39). Moderate alcohol consumption is also associated with increased levels of brain-derived neurotrophic factor as well as lower levels of inflammatory biomarkers (39–41), which are both implicated in depression (42). However, given that moderate drinking is known to reflect or facilitate healthy social interaction (43)—an established protective factor for depression (44)—this should also be acknowledged as a key pathway for observed benefits (25).

Analyses using continuous CES-D-SF scores maximized statistical power and offer population-level interpretability. While applying effect size heuristics to psychological research is imperfect given that small or very small effects are typical (45), the protective effects found in the continuous outcome analyses were small (46). At first glance, these may not represent a clinically meaningful benefit, but given evidence that increased depressive symptom severity is associated with risk for subsequent physical health conditions (4), changes in mean scores—even if below clinical thresholds—are important to study. Further, these effects are consequential at the population level, particularly when they contribute to modeling that informs national drinking

guidelines, where the incorporation of evidence of alcohol's protective effects can substantially alter recommendations (47). Indeed, the present findings offer support for the current U.S. drinking guidelines (30), at least with respect to depression.

The second set of analyses, using a binary outcome, provided predicted probabilities of probable depression and maximize clinical interpretability. These analyses also supported protective effects for stable occasional and moderate consumption (although due to reduced statistical power, the moderate drinker–abstainer contrast was no longer statistically significant). However, any evidence for protective effects must be balanced against that demonstrating increasing dose-response relationships between alcohol consumption and other health outcomes, including several cancers (48).

While consistent above-guideline drinking was associated with increased depressive symptoms and risk for probable depression, this increase was modest and not statistically significant. At higher levels, alcohol begins to exert opposite effects on the same pathways through which lower consumption may offer protection (39). As such, it is ostensibly surprising that consistent above-guideline drinkers did not have greater depression risk. However, recent meta-analytic evidence has suggested that the increased risk of depression among heavy drinkers may be largely driven by confounding (9). Our findings here are consistent with this; only modest increases in depression were found among the above-guideline drinkers after rigorously controlling for available confounders. Importantly, evidence still indicates that disordered drinking more specifically, that is, a compulsive pattern of drinking associated with significant physical and social harms, is associated with increased depression risk (9).

That the results of the primary analyses replicated when the sample was stratified by sex is consistent with previous findings that despite sex-based differences in the epidemiology of alcohol consumption and depression more generally, the nature of the alcohol-depression relationship is similar (25, 49).

Strengths

The application of causal inference methods in this area is a nascent field, and the present study contributes preliminary evidence that associations between moderate alcohol consumption and reduced depression may reflect genuine causal effects. The key strength of MSMs in this context is in generating effect estimates that account for the dynamic nature of alcohol consumption and confounders over time. Studies that ignore drinker drift are likely to produce biased estimates, and although using latent class growth models (LCGMs) to characterize trajectories of consumption avoids this problem, it relies on modeled classes rather than observed patterns in the actual data and cannot answer questions about stable levels of exposure. This also applies to emerging methods combining LCGMs and MSMs (50).

In this study, we controlled for a wide range of variables commonly considered to confound the alcohol-depression relationship. This included pre-baseline alcohol consumption, which was key to mitigating sick-quitter bias given that 60% of baseline abstainers were previous above-guideline drinkers. Robust pre-baseline data and the incorporation of multiple waves of alcohol consumption measures over follow-up represent considerable improvements over the only other extant MSM analysis addressing the alcohol-depression relationship (25).

Limitations

MSMs, like any statistical model, depend on the quality and breadth of confounder measurement. The breakdown of exposure group by covariates in Table 2 illustrates that individual covariates—excepting prior alcohol consumption—tended to associate only modestly with baseline drinker category. This may indicate that the identified protective effects are unlikely to be solely a result of unmeasured confounding. That said, those trends that were evident, such as occasional and moderate drinkers having higher income and employment levels, constitute important social determinants of health. It is plausible that one of, or a combination of, similar unmeasured covariates may have effects on exposure and outcome exceeding the generated E-values. In particular, the present study lacked data on parental alcohol use disorder, as well as sophisticated measures of social support and interaction, which may confound (and mediate) the alcohol-depression relationship in a longitudinal context. Residual confounding may also be possible given that weighting for the 2002 and 2006 exposure groups did not achieve balance for all covariates within the conservative 0.1 standardized mean difference heuristic (see Figure S3 in the online supplement). Moreover, the desired interpretation of the weighted estimates rests on strong assumptions (discussed in section 6 in the online supplement and at length elsewhere [22, 51]) that, at best, approximately hold.

Future Directions

More research applying novel approaches to the alcohol-depression relationship, such as flexible, nonlinear Mendelian randomization (which eschews confounding entirely), is needed to answer remaining questions about causality. Alongside this, triangulation of evidence across various observational designs and statistical methods is key. Due consideration must also be given to the manifold data processing and analysis decisions researchers make, the impacts of which can be quantified using multiverse analysis (52).

AUTHOR AND ARTICLE INFORMATION

Matilda Centre for Research in Mental Health and Substance Use, University of Sydney, Sydney, Australia (Visontay, Mewton, Slade, Sunderland); Centre for Healthy Brain Ageing, University of New South Wales, Sydney, Australia (Mewton); Division of Chronic Disease Research Across the Lifecourse, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston (Aris).

Send correspondence to Ms. Visontay (rachel.visontay@sydney.edu.au).

Ms. Visontay is funded by the National Health and Medical Research Council (GNT1190255), as well as the University of Sydney.

The authors thank Dr. Noah Greifer for his advice on the construction of inverse probability weights and on the use of his R package *WeightIt*, as well as Dr. Alexa Yakubovich for her advice on the use of marginal structural models.

The authors report no financial relationships with commercial interests.

Received January 14, 2022; revisions received April 16 and July 25, 2022; accepted August 16, 2022.

REFERENCES

1. GBD 2019 Diseases and Injuries Collaborators: Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020; 396:1204–1222
2. Liu Q, He H, Yang J, et al: Changes in the global burden of depression from 1990 to 2017: findings from the Global Burden of Disease study. J Psychiatr Res 2020; 126:134–140
3. Patel V, Chisholm D, Parikh R, et al: Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. Lancet 2016; 387:1672–1685
4. Seligman F, Nemeroff CB: The interface of depression and cardiovascular disease: therapeutic implications. Ann N Y Acad Sci 2015; 1345:25–35
5. Choi KW, Stein MB, Nishimi KM, et al: An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression. Am J Psychiatry 2020; 177: 944–954
6. Boden JM, Fergusson DM: Alcohol and depression. Addiction 2011; 106:906–914
7. Brière FN, Rohde P, Seeley JR, et al: Comorbidity between major depression and alcohol use disorder from adolescence to adulthood. Compr Psychiatry 2014; 55:526–533
8. Fergusson DM, Boden JM, Horwood LJ: Tests of causal links between alcohol abuse or dependence and major depression. Arch Gen Psychiatry 2009; 66:260–266
9. Li J, Wang H, Li M, et al: Effect of alcohol use disorders and alcohol intake on the risk of subsequent depressive symptoms: a systematic review and meta-analysis of cohort studies. Addiction 2020; 115: 1224–1243
10. Pacek LR, Martins SS, Crum RM: The bidirectional relationships between alcohol, cannabis, co-occurring alcohol and cannabis use disorders with major depressive disorder: results from a national sample. J Affect Disord 2013; 148:188–195
11. Turner S, Mota N, Bolton J, et al: Self-medication with alcohol or drugs for mood and anxiety disorders: a narrative review of the epidemiological literature. Depress Anxiety 2018; 35:851–860
12. Khantzian EJ: The self-medication hypothesis of addictive disorders: focus on heroin and cocaine dependence. Am J Psychiatry 1985; 142:1259–1264
13. Keyes KM, Allel K, Staudinger UM, et al: Alcohol consumption predicts incidence of depressive episodes across 10 years among older adults in 19 countries. Int Rev Neurobiol 2019; 148:1–38
14. Brookhart MA, Wyss R, Layton JB, et al: Propensity score methods for confounding control in nonexperimental research. Circ Cardiovasc Qual Outcomes 2013; 6:604–611
15. Thoemmes F, Ong AD: A primer on inverse probability of treatment weighting and marginal structural models. Emerg Adulthood 2016; 4:40–59
16. Benedetto U, Head SJ, Angelini GD, et al: Statistical primer: propensity score matching and its alternatives. Eur J Cardiothorac Surg 2018; 53:1112–1117
17. Greenland S, Mansournia MA, Altman DG: Sparse data bias: a problem hiding in plain sight. BMJ 2016; 352:i1981
18. Melberg HO: Does moderate alcohol intake reduce mortality? in Understanding Choice, Explaining Behaviour: Essays in Honour of Ole-Jorgen Kkog. Edited by Jon Elster OG, Moene AHandK. Oslo, Oslo Academic Press, 2006, pp 191–210
19. Shaper AG, Wannamethee G, Walker M: Alcohol and mortality in British men: explaining the U-shaped curve. Lancet 1988; 2: 1267–1273
20. Naimi TS, Stockwell T, Zhao J, et al: Selection biases in observational studies affect associations between "moderate" alcohol consumption and mortality. Addiction 2017; 112:207–214
21. Chikritzhs T, Fillmore K, Stockwell T: A healthy dose of scepticism: four good reasons to think again about protective effects of alcohol on coronary heart disease. Drug Alcohol Rev 2009; 28:441–444
22. Robins JM, Hernán MÁ, Brumback B: Marginal structural models and causal inference in epidemiology. Epidemiology 2000; 11: 550–560
23. Vandecandelaere M, Vansteelandt S, de Fraine B, et al: Time-varying treatments in observational studies: marginal structural models of the effects of early grade retention on math achievement. Multivariate Behav Res 2016; 51:843–864
24. Visontay R, Sunderland M, Slade T, et al: Are there non-linear relationships between alcohol consumption and long-term health? A systematic review of observational studies employing approaches to improve causal inference. BMC Med Res Methodol 2022; 22:16
25. Gémes K, Forsell Y, Janszky I, et al: Moderate alcohol consumption and depression: a longitudinal population-based study in Sweden. Acta Psychiatr Scand 2019; 139:526–535
26. Kerr WC, Ye Y, Williams E, et al: Lifetime alcohol use patterns and risk of diabetes onset in the National Alcohol Survey. Alcohol Clin Exp Res 2019; 43:262–269
27. Alati R, Lawlor DA, Najman JM, et al: Is there really a "J-shaped" curve in the association between alcohol consumption and symptoms of depression and anxiety? Findings from the Mater-University Study of Pregnancy and its outcomes. Addiction 2005; 100:643–651
28. Rothstein DS, Carr D, Cooksey E: Cohort profile: the National Longitudinal Survey of Youth 1979 (NLSY79). Int J Epidemiol 2019; 48:22–22e
29. Calvo E, Allel K, Staudinger UM, et al: Cross-country differences in age trends in alcohol consumption among older adults: a cross-sectional study of individuals aged 50 years and older in 22 countries. Addiction 2021; 116:1399–1412
30. US Department of Agriculture and US Department of Health and Human Services: Dietary Guidelines for Americans, 2020–2025, 9th ed. Washington, DC, US Department of Agriculture, 2020
31. Radloff LS: The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977; 1: 385–401
32. Levine SZ: Evaluating the seven-item Center for Epidemiologic Studies Depression Scale short-form: a longitudinal US community study. Soc Psychiatry Psychiatr Epidemiol 2013; 48:1519–1526
33. Kim HY: Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 2013; 38:52–54
34. Williamson T, Ravani P: Marginal structural models in clinical research: when and how to use them? Nephrol Dial Transplant 2017; 32:ii84–ii90
35. Aris IM, Sordillo JE, Rifas-Shiman SL, et al: Childhood patterns of overweight and wheeze and subsequent risk of current asthma and obesity in adolescence. Paediatr Perinat Epidemiol 2021; 35:569–577
36. VanderWeele TJ, Ding P: Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017; 167:268–274
37. Steffensen SC, Walton CH, Hansen DM, et al: Contingent and non-contingent effects of low-dose ethanol on GABA neuron activity in the ventral tegmental area. Pharmacol Biochem Behav 2009; 92:68–75
38. Bellos S, Skapinakis P, Rai D, et al: Longitudinal association between different levels of alcohol consumption and a new onset of

depression and generalized anxiety disorder: results from an international study in primary care. Psychiatry Res 2016; 243:30–34
39. Tizabi Y, Getachew B, Ferguson CL, et al: Low vs high alcohol: central benefits vs detriments. Neurotox Res 2018; 34:860–869
40. Imhof A, Froehlich M, Brenner H, et al: Effect of alcohol consumption on systemic markers of inflammation. Lancet 2001; 357: 763–767
41. Bell S, Mehta G, Moore K, et al: Ten-year alcohol consumption typologies and trajectories of C-reactive protein, interleukin-6, and interleukin-1 receptor antagonist over the following 12 years: a prospective cohort study. J Intern Med 2017; 281:75–85
42. Milaneschi Y, Kappelmann N, Ye Z, et al: Association of inflammation with depression and anxiety: evidence for symptom specificity and potential causality from UK Biobank and NESDA cohorts. Mol Psychiatry 2021; 26:7393–7402
43. Peele S, Brodsky A: Exploring psychological benefits associated with moderate alcohol use: a necessary corrective to assessments of drinking outcomes? Drug Alcohol Depend 2000; 60:221–247
44. Cruwys T, Dingle GA, Haslam C, et al: Social group memberships protect against future depression, alleviate depression symptoms, and prevent depression relapse. Soc Sci Med 2013; 98:179–186
45. Owens MM, Potter A, Hyatt CS, et al: Recalibrating expectations about effect size: a multi-method survey of effect sizes in the ABCD Study. PLoS One 2021; 16:e0257535
46. Cohen J: Statistical Power Analysis for the Behavioral Sciences. New York, Routledge, 2013
47. Australian Guidelines to Reduce Health Risks from Drinking Alcohol. Canberra, Australia, National Health and Medical Research Council, 2020. https://www.nhmrc.gov.au/about-us/publications/australian-guidelines-reduce-health-risks-drinking-alcohol
48. Rehm J, Soerjomataram I, Ferreira-Borges C, et al: Does alcohol use affect cancer risk? Curr Nutr Rep 2019; 8:222–229
49. Liang L, Hua R, Tang S, et al: Low-to-moderate alcohol intake associated with lower risk of incidental depressive symptoms: a pooled analysis of three intercontinental cohort studies. J Affect Disord 2021; 286:49–57
50. Diop A, Sirois C, Guertin JR, et al: Marginal structural models with latent class growth modeling of treatment trajectories. arXiv, May 26, 2021. https://doi.org/10.48550/arXiv.2105.12720
51. VanderWeele TJ, Hernan MA: Causal inference under multiple versions of treatment. J Causal Inference 2013; 1:1–20
52. Steegen S, Tuerlinckx F, Gelman A, et al: Increasing transparency through a multiverse analysis. Perspect Psychol Sci 2016; 11:702–712

Appendix 1

Supplementary materials for Chapter 2

Appendix 1.1. Ovid MEDLINE Search Strategy

1	exp Alcoholic Beverages/
2	(alcohol* or spirit* or beer* or wine*).mp.
3	exp Alcohol Drinking/
4	1 or 2 or 3
5	alcoholic intoxication/ or binge drinking/
6	Alcohol Abstinence/
7	Alcohol Dehydrogenase/
8	aldehyde oxidoreductases/ or aldehyde dehydrogenase/ or aldehyde dehydrogenase 1/ or aldehyde dehydrogenase, mitochondrial/
9	(alcohol dehydrogenase* or acetaldehyde dehydrogenase* or aldehyde dehydrogenase or ADH1B* or ALDH2* or ADH1C*).mp.
10	((former or current) adj drinker*).mp.
11	(abstain* or abstinence or "ex drinker*" or "non drinker*" or "non drinking" or exdrinker* or nondrinker* or nondrinking*).mp.
12	((level* or life-course or lifetime or moderate or regular or low or lower or lowest or light or lighter or lightest or occasional or hazardous or risky or heavy or heavier or heaviest or bing* or episodic or pattern*) adj3 (drinking or drinker* or consum* or intake or "use" or "usage" or habit*)).mp.
13	5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14	exp case-control studies/
15	exp Cohort Studies/
16	(cohort or case control or longitudinal or prospective or retrospective or follow-up or baseline).mp.
17	14 or 15 or 16
18	(causal adj2 (effect* or inference or model* or mediat*)).mp. (9656)
19	(mediat* adj2 (pathway* or mechanism*)).mp.
20	Propensity Score/
21	propensity scor*.mp.
22	(inverse probability or probability weight* or IPW or IPTW).mp.
23	(marginal structural model* or marginal method* or pseudopopulation or pseudo population).mp.
24	(generali#ed method* or G formula* or G estimat* or G computation or G method* or general formula* or structural nested model* or (standardi* and caus*)).mp.
25	(targeted maximum likelihood estimat* or TMLE).mp.
26	Fixed effect* regression.mp.
27	(potential outcome* or doubly robust).mp.
28	or/18-27
29	negative control.mp.
30	Twin study/
31	(family-based design or co-relative design or sibship or sibling or twin or genetic epidemiology or genetic variant* or genetically informed method*).mp.
32	Mendelian Randomization Analysis/
33	(instrumental variable* or instrumental analys* or mendelian randomi*).mp.
34	(Natural experiment or quasiexperiment* or quasi experiment or pre post or "before and after study" or "before and after studies" or before after stud* or difference in differences or interrupted time series or regression discontinuity).mp.
35	or/29-34
36	35 or (17 and 28)
37	4 and 13 and 36
38	limit 37 to english language

Appendix 1.2. Studies contacted for additional information (n=10)

First author (year)	Number of contact attempts, dates, author contacted	Contact with authors made?	Information requested	Information received
Included articles				
Carlsson (2003)	1. 22.12.20; lead author	Y; response received 27.12.20	To confirm that, and ask for reason why, individuals consuming 5-15g/day of alcohol were excluded from discordant twin analyses	Confirmed; in order to ensure twins were truly exposure discordant
Peng (2019)	1. 26.12.20; corresponding author 2. 18.01.21; corresponding author 3. 02.02.21; joint lead author	N	To ask whether Table 4 reports on results adjusted for age only, and to confirm that the LATE method was not applied to diabetes risk itself	N/A
Samuelsson (2013)	1. 04.01.21; corresponding author	Y; response received 05.01.21	To confirm that reported mean age represents baseline age, and to ask for results from analyses reported as conducted excluding those with baseline mental illness	Confirmed; these analyses were only conducted for the pooled cohort (not discordant twins) and data is not accessible due to COVID
Kadlecová (2015)	1. 11.01.21; corresponding author (bounce back) 2. 11.01.21; other author with accessible contact details – message forwarded to corresponding author	Y; response received 13.01.21	To confirm that reported twin results are adjusted analyses	Confirmed
Silverwood (2014)	1. 02.02.21; lead author	Y; response received 03.02.21	To request mean age at baseline for the overall sample	Pointed to another paper (Holmes et al.) containing the same cohorts, from which this statistic can be calculated

First author (year)	Number of contact attempts, dates, author contacted	Contact with authors made?	Information requested	Information received
Ropponen (2013)	1. 06.02.21; lead author	Y; response received 08.02.21	To request mean age at baseline for the sample, and to ask whether twin analyses were adjusted for covariates	Mean age was 53.7 (SD 5.7); twin analyses adjusted for sex and age only
Sipilä (2016)	1. 20.11.20; corresponding author	Y; response received 20.11.20	To request information on numbers of drinking discordant pairs	Excel file provided with number of discordant pairs per discordance combination
Gémes (2019)	1. 20.02.21; lead author	Y; response received 20.02.21	To ask which of two discrepant estimates reported for non-drinker RR is correct	Told that the statistics reported in Table 2 (as opposed to body of text) were correct

First author (year)	Number of contact attempts, dates, author contacted	Contact with authors made?	Information requested	Information received
Articles deemed ineligible on basis of author response/non-response				
Korhonen (2015)	1. 06.10.20; lead author 2. 27.10.20; lead author	N	To request more information on twin analyses	N/A
Virta (2010)	1. 06.10.20; lead author 2. 27.10.20; lead author	Y; response received 30.10.20	To request more information on results of twin analyses	Data no longer accessible

Appendix 1.3. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #[a]
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Table S1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6

Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	6-7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	7
Section/topic	**#**	**Checklist item**	**Reported on page #**
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7; Fig. 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	11; Table S6
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A

Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11-12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	1

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

[a] Page numbers refer to the published journal article (see Appendix A).

Appendix 1.4. Exclusion reasons for key ineligible studies

Study	Reason for exclusion
Mendelian Randomization studies	
Beasley, M., Freidin, M. B., Basu, N., Williams, F. M., & Macfarlane, G. J. (2019). What is the effect of alcohol consumption on the risk of chronic widespread pain? A Mendelian randomisation study using UK Biobank. *Pain*, *160*(2), 501-507.	Could not detect non-linearity.
Au Yeung, S. L. A., Jiang, C., Cheng, K. K., Cowling, B. J., Liu, B., Zhang, W., ... & Schooling, C. M. (2013). Moderate alcohol use and cardiovascular disease from Mendelian randomization. *PloS one*, *8*(7), e68054.	Sensitivity analyses excluding heavy drinkers not adequate to determine functional form.
Au Yeung, S. L., Jiang, C. Q., Cheng, K. K., Liu, B., Zhang, W. S., Lam, T. H., ... & Schooling, C. M. (2012). Evaluation of moderate alcohol use and cognitive function among men using a Mendelian randomization design in the Guangzhou biobank cohort study. *American journal of epidemiology*, *175*(10), 1021-1028.	Sensitivity analyses excluding heavy drinkers not adequate to determine functional form.
Au Yeung, S. L., Jiang, C., Long, M., Cheng, K. K., Liu, B., Zhang, W., ... & Schooling, C. M. (2015). Evaluation of moderate alcohol use with QT Interval and heart rate using Mendelian randomization analysis among older Southern Chinese men in the Guangzhou Biobank Cohort Study. *American journal of epidemiology*, *182*(4), 320-327.	Sensitivity analyses excluding heavy drinkers not adequate to determine functional form.
Andrews, S. J., Goate, A., & Anstey, K. J. (2019). Association between alcohol consumption and Alzheimer's disease: A Mendelian Randomization Study. *Alzheimer's & Dementia*.	Could not detect non-linearity.
Holmes, M. V., Dale, C. E., Zuccolo, L., Silverwood, R. J., Guo, Y., Ye, Z., ... & Langenberg, C. (2014). Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ*, *349*, g4164.	Did not conduct formal IV analysis/provide estimates in terms of genetically-predicted alcohol consumption.
Kumari, M., Holmes, M. V., Dale, C. E., Hubacek, J. A., Palmer, T. M., Pikhart, H., ... & Bobak, M. (2014). Alcohol consumption and cognitive performance: a Mendelian randomization study. *Addiction*, *109*(9), 1462-1471.	Sensitivity analyses excluding heavy drinkers not adequate to determine functional form.
Lawlor, D. A., Nordestgaard, B. G., Benn, M., Zuccolo, L., Tybjaerg-Hansen, A., & Davey Smith, G. (2013). Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian randomization study in the Copenhagen General Population Study. *European heart journal*, *34*(32), 2519-2528.	Could not detect non-linearity.
Tabara, Y., Arai, H., Hirao, Y., Takahashi, Y., Setoh, K., Kawaguchi, T., ... & Nagahama Study Group. (2017). The causal effects of alcohol on lipoprotein subfraction and triglyceride levels using a Mendelian randomization analysis: The Nagahama study. *Atherosclerosis*, *257*, 22-28.	Could not detect non-linearity.
Cohort studies	
Koch, M., Costanzo, S., Fitzpatrick, A. L., Lopez, O. L., DeKosky, S., Kuller, L. H., ... & Mukamal, K. J. (2020). Alcohol Consumption, Brain Amyloid-β Deposition, and Brain Structural Integrity Among Older Adults Free of Dementia. *Journal of Alzheimer's Disease*, (Preprint), 1-11.	IPWs used for attrition only.
Paul, K. C., Chuang, Y. H., Shih, I. F., Keener, A., Bordelon, Y., Bronstein, J. M., & Ritz, B. (2019). The association between lifestyle factors and Parkinson's disease progression and mortality. *Movement Disorders*, *34*(1), 58-66.	IPWs used for attrition only.

Marti, C. N., Choi, N. G., DiNitto, D. M., & Choi, B. Y. (2015). Associations of lifetime abstention and past and current alcohol use with late-life mental health: a propensity score analysis. *Drug and alcohol dependence*, *149*, 245-251.	Cross-sectional.
Kerr, W. C., & Ye, Y. (2010). Relationship of life-course drinking patterns to diabetes, heart problems, and hypertension among those 40 and older in the 2005 US National Alcohol Survey. *Journal of studies on alcohol and drugs*, *71*(4), 515-525.	Cross-sectional.
Beulens, J. W., van der Schouw, Y. T., Moons, K. G., Boshuizen, H. C., & Groenwold, R. H. (2013). Estimating the mediating effect of different biomarkers on the relation of alcohol consumption with the risk of type 2 diabetes. *Annals of epidemiology*, *23*(4), 193-197.	Exposure was not categorised, and so study could not detect non-linearity.

Appendix 1.5. Risk of Bias in cohort studies as assessed by the Newcastle-Ottawa Scale

Study first author	Selection 1	Selection 2	Selection 3	Selection 4	Comparability	Outcome 1	Outcome 2	Outcome 3	Total /9
Carlsson	✸	✸		✸	✸✸	✸	✸	✸	8
Dickerman	✸	✸			✸✸	✸	✸	✸	7
Gémes	✸	✸		✸	✸✸		✸	✸	7
Handing	✸	✸		✸	✸✸	[a]	✸	✸	7
Ilomäki	✸	✸			✸✸	✸	✸	✸	7
Kadlecová	✸	✸		✸	✸✸	✸	✸	✸	8
Pietkäinen	✸	✸		✸	✸✸	✸	✸	✸	8
Ropponen (2011)	✸	✸		✸	✸✸	✸	✸	✸	8
Ropponen (2013)	✸	✸	✸	✸	✸✸	✸	✸	✸	9
Samuelsson	✸	✸	✸	✸	✸✸	✸	✸	✸	9
Sander	✸	✸	[b]	✸	✸✸	✸	✸	✸	8
Sipilä	✸	✸		✸	✸✸	✸	✸	✸	8

Selection 1 = representativeness of the exposed cohort [star for truly or somewhat representative]

Selection 2 = selection of the non-exposed cohort [star for drawn from same community as exposed cohort]

Selection 3 = ascertainment of exposure [star for secure record or structured interview]

Selection 4 = demonstration that outcome of interest was not present at start of study [star for yes]

Comparability = comparability of cohorts on the bases of the design or analysis [up to two stars for controlling for the most important factor and any additional important factor]

Outcome 1 = assessment of outcome [star for independent blind assessment or record linkage]

Outcome 2 = was follow-up long enough for outcomes to occur [star for yes]

Outcome 3 = adequacy of follow-up cohorts [star for complete follow-up or subjects lost to follow-up unlikely to introduce bias]

[a] Study did use record linkage but had poor sensitivity.

[b] Study used a mix of self-report and structured interview, but for some participants only self-report was conducted (depending on time of enrolment).

Appendix 1.6. Risk of Bias in Mendelian Randomization studies using Mamluk et al. (2020) tool

Study first author	Weak instrument bias	Genetic confounding	'Non-genetic confounding'[a]	Pleiotropy	Selection Bias
Silverwood	Moderate risk	Moderate risk[b]	Low risk	Moderate risk	Moderate risk
Millwood	Moderate risk	Low risk	Low risk[c]	Low risk	Low risk[d]
Peng	Low risk	Low risk	Low risk	Low risk	Moderate risk
Vu	Moderate risk[e]	Low risk	Low risk	Low risk	Low risk

[a] Inclusion of covariates other than age or sex in analyses.

[b] Although simulations demonstrated robustness of relationship detection to confounding.

[c] Further analyses did adjust for education, income and smoking, but we did not extract these given the potential for introducing bias.

[d] While the study did not adjust for principal components, it did stratify by areas within China.

[e] Some F values slightly above and some slightly below 10.

Appendix 1.7. Covariates adjusted/stratified for in each study

Study first author	In main causal analyses	In main conventional analyses (if applicable)	In secondary/sensitivity causal (if applicable)	In secondary/sensitivity other (if applicable)
Dickerman	Twin: none	Intrapair correlations (for SEs/CIs only), BMI, smoking, social class, education, physical activity	N/A	N/A
Carlsson	Twin: none	Age, BMI	N/A	Smoking, socioeconomic status, physical activity, intrapair correlations
Peng	MR: age, sex	N/A	Age, sex, BMI, waist circumference, physical activity, and education (results of these analyses not extracted in this review as controlling for additional confounders in MR studies can introduce bias)	N/A
Handing	Twin: age, sex, education, smoking, physical activity, diabetes, and hypertension	Age, sex, education, smoking, physical activity, diabetes, and hypertension	N/A	N/A
Gémes	MSM: alcohol consumption pre-baseline, symptoms of depression, stressful life events, financial stress, social support, and having small children measured both pre-baseline and at baseline considered as time-varying confounders; age, sex, education, country of birth, childhood adversities, smoking included as time-fixed	Non-MSM: age, sex, education, being born in Sweden, cohabiting with small children, smoking, stressful life events, financial stress, childhood adversities, and baseline MDI	MSM without those with baseline depression: living with small children, stressful life events, financial stress, social support, symptoms of depression and previous alcohol consumption as time-varying covariates; age, sex, education, country of birth, smoking and childhood adversities as time-stable covariates	Non-MSM without those with baseline depression: age, sex, education, being born in Sweden, cohabiting with small children, smoking, stressful life events, financial stress and childhood adversities

	confounders (assessed only at baseline)			
Samuelsson	Twins: age and sex	Age, sex, education, marital status, disease severity, physical activity, BMI, tobacco use, and self-rated health	N/A	N/A
Ilomäki	MSM: time-invariant (age, marital status, rurality, employment status, education, family history of CVD, cigarettes per day, years of smoking, and diabetes) and time-varying covariates (BMI, smoking status, HDL-C, serum insulin, serum fibrinogen, resting SBP, and history of CVD)	Various – see Table 1 in paper proper	N/A	N/A
Kadlecová	Twin: age, sex, preferred alcohol, current smoking, BMI, physical activity, stress reactivity, and a range of other illnesses	Age, sex, preferred alcohol, current smoking, BMI, physical activity, stress reactivity, and a range of other illnesses	Age, sex, preferred alcohol, ever smoking, BMI, physical activity, stress reactivity, and a range of other illnesses	Age, sex, preferred alcohol, ever smoking, BMI, physical activity, stress reactivity, and a range of other illnesses
Millwood	MR: area and age	Area, age, education, household income, and smoking	Area, age, education, income, and smoking (results of these analyses not extracted in this review as controlling for additional confounders in MR studies can introduce bias)	N/A
Ropponen (2014)	Twins: age and sex	Stratified by sex and clustered on twin-pair identify; adjusted for age,	N/A	N/A

		BMI, other diseases, education, marital status, and health behaviours		
Silverwood	MR: age, sex, and study	N/A	N/A	N/A
Vu	MR: sex, age, and genetic principal components	None reported	N/A	N/A
Sipilä	Twin: age, gender, marital status, smoking status, physical activity, obesity, education, and social class	Age, gender, marital status, smoking status, physical activity, obesity, education, and social class	N/A	N/A
Sander	MSM: race, ethnicity, age, enrollment city, and education as time-fixed covariates; depressive symptoms, gonorrhea/chlamydia, smoking, and illicit drug use as time-dependent covariates lagged one visit	Baseline covariates: age (spline), race, ethnicity, education; time-dependent covariates lagged one visit: illicit drug use, cigarette smoking, depression, sexually transmitted infection, and multiple unprotected receptive anal intercourse partners	N/A	N/A
Pietkäinen	Twin: age	Age	N/A	N/A
Ropponen (2011)	Twin: none	Age, socio-demographic factors, education, social class, health-related factors, BMI, chronic disease, number locations musculoskeletal pain, use of analgesics, smoking status, and clustered on twin pair identity	N/A	N/A

Supplementary References

Korhonen, T., Smeds, E., Silventoinen, K., Heikkilä, K., & Kaprio, J. (2015). Cigarette smoking and alcohol use as predictors of disability retirement: a population-based cohort study. *Drug and alcohol dependence*, *155*, 260-266.

Mamluk, L., Jones, T., Ijaz, S., Edwards, H. B., Savović, J., Leach, V., ... & Zuccolo, L. (2020). Evidence of detrimental effects of prenatal alcohol exposure on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies. *International Journal of Epidemiology*, *49*(6), 1972-1995.

Virta, J. J., Järvenpää, T., Heikkilä, K., Perola, M., Koskenvuo, M., Räihä, I., ... & Kaprio, J. (2010). Midlife alcohol consumption and later risk of cognitive impairment: a twin follow-up study. *Journal of Alzheimer's disease*, *22*(3), 939-948.

Appendix 2

Supplementary materials for Chapter 3

Appendix 2.1. Results of chi-square tests of independence for > 10 mg L^{-1} exclusion by drinking categories (using categorisations based on UK guidelines).

	hsCRP concentration				Chi-squared p value
	≤ 10 mg L^{-1}		> 10 mg L^{-1}		
	N	%	N	%	
Drinking group (2004)					0.12
Lifetime abstainers	52	95	3	5	
Former drinkers	108	95	7	5	
Occasional drinkers	621	95	30	5	
Low-to-moderate drinkers	1160	97	32	3	
Above-guidelines drinkers	684	96	26	4	
Drinking group (2012)					0.04
Current abstainers[a]	218	94	14	6	
Occasional drinkers	303	97	10	3	
Low-to-moderate drinkers	1056	98	26	2	
Above-guidelines drinkers	1186	96	44	4	

[a] As noted in Chapter 3.2.2, there was no information on lifetime abstention at 2012.

Appendix 2.2. Supplementary Methods

Covariates: The demographic/socioeconomic group

Sex was recorded as female or male, and for the present study was composed of information recorded at either age 34 (2004), or if that was missing, age 42 (2012). The ethnicity variable was based on responses at age 16 (1986) collapsed into White and other (due to small numbers of individuals reporting non-White ethnicities), or if that was missing, age 34 (2004).

Variables collected at birth included in this group were: mother's age at birth, parent's marital status at birth (collapsed into married and unmarried), number of years mother spent in education, father's job (collapsed into manual, non-manual, does not support mother), mother's social status based on job (collapsed into manual, non-manual, and housewife).

Variables collected at age 30 included in this group were: current employment status (collapsed into employed, unemployed/in education/sick or disabled, looking after home/family), self-assessment of financial situation (collapsed into difficult, just about getting by, doing alright, and living comfortably), experience of homelessness (yes/no), marital status at age 30 (collapsed into married/cohabiting, single, and separated/divorced/widowed)

Covariate	**Missing *N* (/3101)**
Sex	0
Ethnicity	170
Mother's age at birth	232
Parents' marital status at birth	221
Mother education at birth	236
Father's job at birth	310
Mother's job at birth/social status	451
Marital status	335
Experience of homelessness	495
Financial self-assessment	304
Employment status	340
TOTAL (missing at least one of the above)	1027 (33%)

Covariates: The physical health group

Delivery method (vaginal or caesarean) was recorded at birth. The other variables included in this group were recorded at age 30: self-rated health (excellent, good, fair, or poor), presence of any disability, past-year doctor visitation for ongoing conditions, self-reported history (yes/no) of asthma, eczema, psoriasis, irritable bowel syndrome (IBS), ulcerative colitis, Crohn's disease, cancer, diabetes, and high blood pressure, and body mass index (BMI) derived from weight and height variables.

Covariate	Missing *N* (/3101)
Delivery method	225
Self-rated health	304
Any disability	304
Past-year doctor visitation	304
History of asthma	304
History of eczema	304
History of psoriasis	304
History of IBS	562
History of ulcerative colitis	304
History of Crohn's disease	304
History of cancer	305
History of diabetes	304
History of high blood pressure	304
BMI	374
TOTAL (missing at least one of the above)	801 (26%)

Covariates: The lifestyle behaviours group

All variables in this group were recorded at age 30: smoking status (never smoked, former smoker, occasional smoker, daily smoker), employed in night shift work (yes/no) derived from two individual variables, experience of sleeping difficulties (yes/no), fruit consumption (collapsed into at least daily and less than daily), vegetable consumption (derived from data on cooked vegetables and salad consumption; collapsed into at least daily and less than daily), sugar consumption (derived from data on cakes and sweets consumption; collapsed into at least daily and less than daily), weekly exercise (yes/no), weekly moderate-to-vigorous physical activity (MVPA; yes/no).

Covariate	Missing *N* (/3101)
Smoking status	306
Night shift work	304
Sleeping difficulties	322
Fruit consumption	304
Veg consumption	306

Sugar consumption	306
Weekly exercise	306
Weekly MVPA	306
TOTAL (missing at least one of the above)	323 (10%)

Covariates: The social group

All variables in this group were recorded at age 30: whether participant has someone they can turn to for emotional support (yes/no), at least monthly religious attendance (yes/no), history of membership in a communal organisation (yes/no).

Covariate	**Missing *N* (/3101)**
Support person	304
Religious attendance	1026
Communal organisation	306
TOTAL (missing at least one of the above)	1026 (33%)

Covariates: The mental health group

All variables in this group were recorded at age 30: total Malaise Inventory score (a scale demonstrating good psychometric properties; McGee et al., 1986), self-reported history of eating disorders (yes/no), visiting a medical practitioner for mental health disorders (yes/no).

Covariate	**Missing *N* (/3101)**
Malaise Inventory score	324
History of eating disorders	304
Visiting practitioner for mental health	304
TOTAL (missing at least one of the above)	324 (10%)

Post-hoc sensitivity analyses

Results of post-hoc sensitivity analyses run on a reduced sample, restricted to individuals with no missing covariate data (n=1358), were similar to those of the main analyses: compared to occasional drinking, none of the 416 estimates for low-to-moderate drinking were consistent with increased hsCRP (RB: 0.24; RP: 4.20), while 130/416 estimates for above-guidelines drinking were consistent with increased hsCRP (RB: 0.89; RP: 2.90). Specification curve plots are depicted in Appendices 2.10 and 2.11.

Variance decomposition effects were similar to those of the main analyses but stronger: the highest ICC for the low-to-moderate drinkers vs occasional drinkers comparison belonged to breadth of covariate control (0.38), followed by measurement year that drinking category was based on (.17). The highest ICC for the above-guidelines vs occasional drinkers comparison belonged to measurement year that drinking category was based on (.41), followed by which national guidelines were used to define drinking categories (0.26). These are also evident in stronger individual specification effects (Appendix 2.12).

Appendix 2.3. Results from regressing a variable reflecting whether an individual provided hsCRP data on drinking status and covariate values.

(a) For those with age 34 drinking category data.

	Providing hsCRP data		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.59	0.39 – 0.88	**0.010**
Drinking category (Age 34)			
Lifetime abstainer	*Reference*		
Former drinker	1.09	0.96 – 1.23	0.189
Occasional drinker	0.88	0.59 – 1.28	0.509
Low-to-moderate drinker	1.06	0.81 – 1.36	0.672
Above-guidelines drinker	1.13	0.98 – 1.31	0.093
Mother's age at birth	1.00	0.99 – 1.01	0.586
Sex			
Female	*Reference*		
Male	1.02	0.92 – 1.13	0.714
Self-rated health (Age 30)			
Excellent	*Reference*		
Good	1.03	0.92 – 1.15	0.590
Fair	1.01	0.85 – 1.21	0.872
Poor	0.70	0.45 – 1.06	0.105
BMI (Age 30)	0.99	0.97 – 1.00	**0.012**
Smoking status (Age 30)			
Never smoker	*Reference*		

Former smoker	1.01	0.89 – 1.15	0.873
Occasional smoker	1.04	0.86 – 1.26	0.653
Daily smoker	0.81	0.72 – 0.92	**0.001**
Weekly exercise (Age 30)			
No	*Reference*		
Yes	0.94	0.84 – 1.05	0.268
Malaise scale total (Age 30)	0.99	0.98 – 1.01	0.459
Observations	7755		
R^2 Tjur	0.004		

Note: drinking categories are based on UK guidelines.

(b) For those with age 42 drinking category data.

	Providing hsCRP data		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.80	0.52 – 1.25	0.328
Drinking category (Age 42)			
Abstainer	*Reference*		
Occasional drinker	0.89	0.70 – 1.13	0.331
Low-to-moderate drinker	0.98	0.82 – 1.17	0.832
Above-guidelines drinker	1.05	0.88 – 1.26	0.582
Mother's age at birth	1.00	0.99 – 1.01	0.382
Sex			
Female	*Reference*		
Male	1.04	0.94 – 1.16	0.435

Self-rated health (Age 30)			
Excellent	*Reference*		
Good	1.07	0.95 – 1.20	0.277
Fair	1.04	0.87 – 1.25	0.645
Poor	0.84	0.54 – 1.27	0.413
BMI_(Age 30)	0.98	0.97 – 0.99	**0.001**
Smoking status (Age 30)			
Never smoker	*Reference*		
Former smoker	1.02	0.89 – 1.17	0.809
Occasional smoker	1.04	0.85 – 1.26	0.690
Daily smoker	0.85	0.75 – 0.97	**0.015**
Weekly exercise (Age 30)			
No	*Reference*		
Yes	0.96	0.86 – 1.08	0.474
Malaise scale total (Age 30)	0.99	0.98 – 1.01	0.548
Observations	6652		
R^2 Tjur	0.004		

Note: drinking categories are based on UK guidelines.

Appendix 2.4. Cohort characteristics for selected covariates and hsCRP by age 42 alcohol consumption categories based on UK guidelines.

	Abstainer (N=218)	**Occasional drinker (N=1186)**	**Low-to-moderate drinker (N=1056)**	**Above-guidelines drinker (N=303)**	**Total (N=2763)**
UK standard drinks per week					
Mean (SD)	NA (NA)	NA (NA)	7.44 (3.22)	23.3 (8.49)	5.46 (7.99)
Median [Min, Max]	NA [NA, NA]	NA [NA, NA]	8.25 [3.75, 13.8]	19.3 [18.8, 59.1]	0 [0, 59.1]
Missing	218 (100%)	1186 (100%)	0 (0%)	0 (0%)	28 (1.0%)
Sex					
Female	136 (62.4%)	688 (58.0%)	531 (50.3%)	85 (28.1%)	1440 (52.1%)
Male	82 (37.6%)	498 (42.0%)	525 (49.7%)	218 (71.9%)	1323 (47.9%)
Mother's age at birth					
Mean (SD)	26.3 (5.28)	25.8 (5.17)	26.4 (5.35)	26.3 (5.51)	26.1 (5.29)
Median [Min, Max]	26.0 [16.0, 44.0]	25.0 [15.0, 44.0]	25.0 [15.0, 49.0]	26.0 [16.0, 42.0]	25.0 [15.0, 49.0]
Missing	21 (9.6%)	84 (7.1%)	64 (6.1%)	28 (9.2%)	197 (7.1%)
Self-rated health (Age 30)					
Excellent	60 (27.5%)	330 (27.8%)	355 (33.6%)	83 (27.4%)	828 (30.0%)
Good	106 (48.6%)	576 (48.6%)	503 (47.6%)	151 (49.8%)	1336 (48.4%)
Fair	28 (12.8%)	140 (11.8%)	83 (7.9%)	32 (10.6%)	283 (10.2%)
Poor	7 (3.2%)	18 (1.5%)	5 (0.5%)	4 (1.3%)	34 (1.2%)

	Abstainer (N=218)	Occasional drinker (N=1186)	Low-to-moderate drinker (N=1056)	Above-guidelines drinker (N=303)	Total (N=2763)
Missing	17 (7.8%)	122 (10.3%)	110 (10.4%)	33 (10.9%)	282 (10.2%)
BMI (Age 30)					
Mean (SD)	24.5 (4.59)	24.9 (4.43)	24.1 (3.72)	24.8 (3.81)	24.6 (4.13)
Median [Min, Max]	23.6 [16.9, 45.2]	24.1 [11.1, 57.2]	23.7 [9.98, 61.2]	24.5 [17.4, 41.7]	24.0 [9.98, 61.2]
Missing	26 (11.9%)	151 (12.7%)	131 (12.4%)	38 (12.5%)	346 (12.5%)
Smoking status (Age 30)					
Never smoker	106 (48.6%)	538 (45.4%)	451 (42.7%)	83 (27.4%)	1178 (42.6%)
Former smoker	34 (15.6%)	177 (14.9%)	210 (19.9%)	68 (22.4%)	489 (17.7%)
Occasional smoker	9 (4.1%)	74 (6.2%)	101 (9.6%)	22 (7.3%)	206 (7.5%)
Daily smoker	52 (23.9%)	274 (23.1%)	184 (17.4%)	96 (31.7%)	606 (21.9%)
Missing	17 (7.8%)	123 (10.4%)	110 (10.4%)	34 (11.2%)	284 (10.3%)
Weekly exercise (Age 30)					
No	67 (30.7%)	327 (27.6%)	234 (22.2%)	82 (27.1%)	710 (25.7%)
Yes	134 (61.5%)	736 (62.1%)	712 (67.4%)	187 (61.7%)	1769 (64.0%)
Missing	17 (7.8%)	123 (10.4%)	110 (10.4%)	34 (11.2%)	284 (10.3%)
Malaise scale total (Age 30)					
Mean (SD)	4.01 (4.11)	3.35 (3.16)	2.99 (3.01)	3.25 (3.00)	3.26 (3.19)

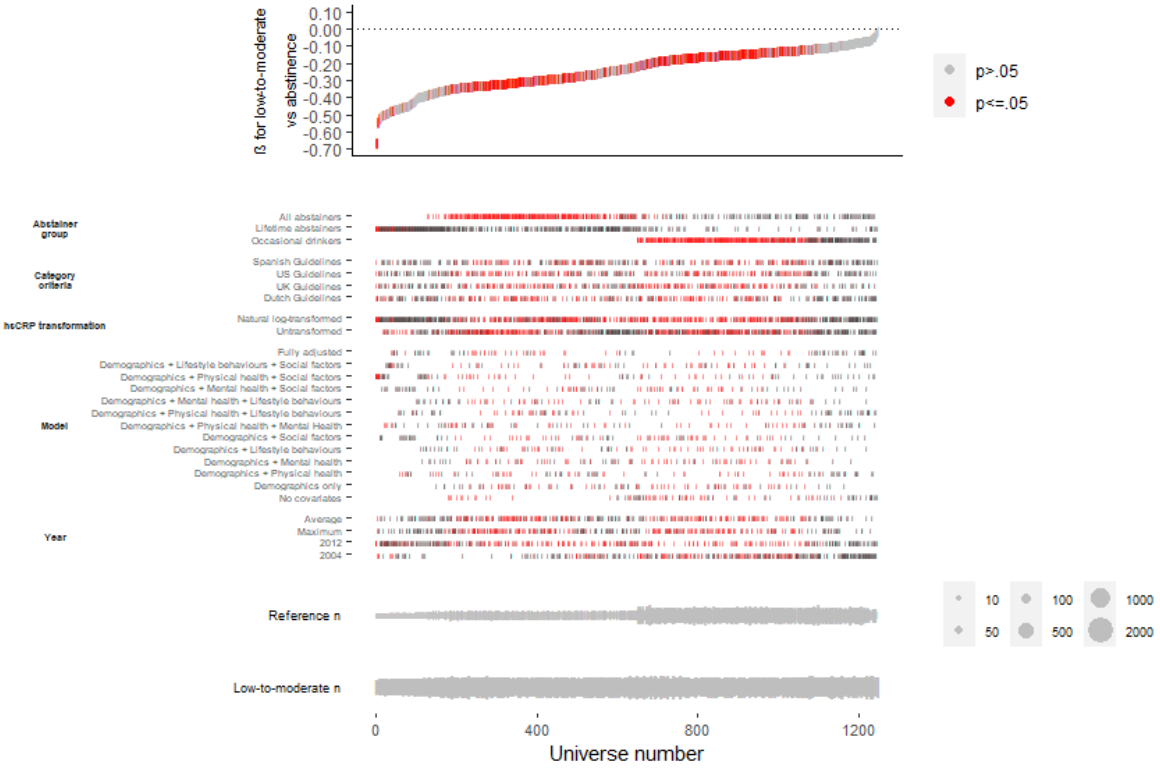
	Abstainer (N=218)	Occasional drinker (N=1186)	Low-to-moderate drinker (N=1056)	Above-guidelines drinker (N=303)	Total (N=2763)
Median [Min, Max]	3.00 [0, 18.0]	3.00 [0, 17.0]	2.00 [0, 22.0]	2.00 [0, 14.0]	2.00 [0, 22.0]
Missing	19 (8.7%)	132 (11.1%)	114 (10.8%)	36 (11.9%)	301 (10.9%)
Ever member of communal organisation (Age 30)					
No	155 (71.1%)	833 (70.2%)	747 (70.7%)	225 (74.3%)	1960 (70.9%)
Yes	46 (21.1%)	230 (19.4%)	199 (18.8%)	44 (14.5%)	519 (18.8%)
Missing	17 (7.8%)	123 (10.4%)	110 (10.4%)	34 (11.2%)	284 (10.3%)
hsCRP (log-transformed) (Age 46)					
Mean (SD)	0.00148 (2.66)	-0.106 (2.46)	-0.350 (2.70)	0.00413 (1.96)	-0.178 (2.53)
Median [Min, Max]	0.182 [-20.7, 2.30]	0.0953 [-20.7, 2.29]	-0.223 [-20.7, 2.30]	0 [-20.7, 2.26]	0 [-20.7, 2.30]

Abbreviations: SD, standard deviation; BMI, body mass index; hsCRP, high-sensitivity C-reactive protein

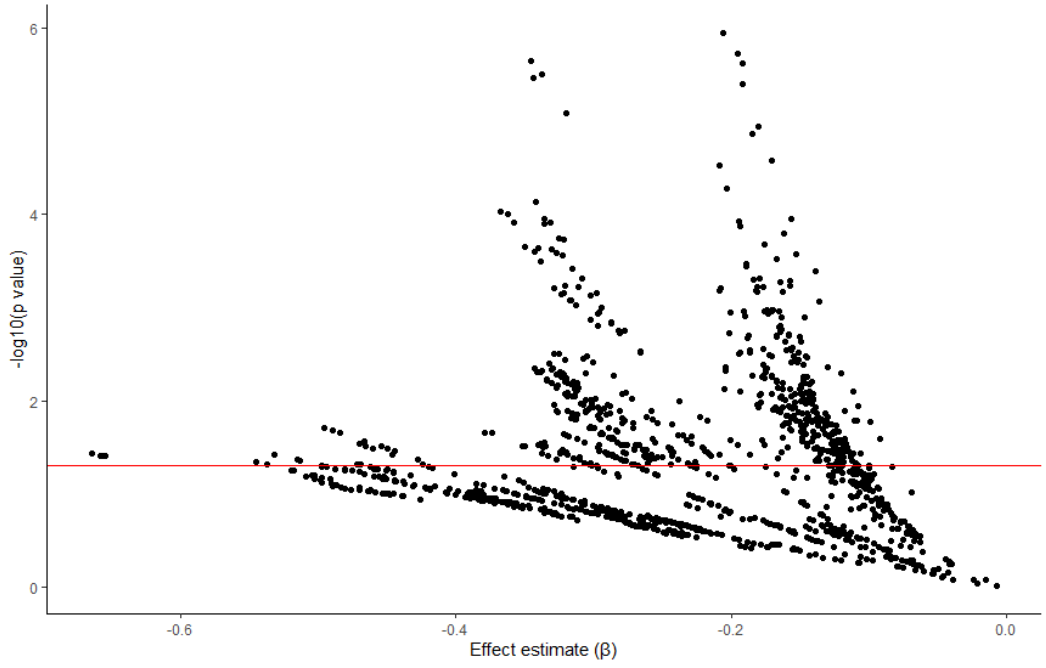
Appendix 2.5. Descriptive statistics for (UK) standard drinks per week consumed by those in above-guidelines drinking categories.

Guidelines	Age 34 above-guidelines drinkers		Age 42 above guidelines drinkers	
	N	Median drinks per week (IQR)	N	Median drinks per week (IQR)
Dutch	1073	18 (16)	431	19.25 (5.5)
UK	684	24 (15)	303	19.25 (7.63)
US	478	30 (20.75)	190	19.25 (11.5)
Spanish	199	45 (20)	38	41.25 (11)

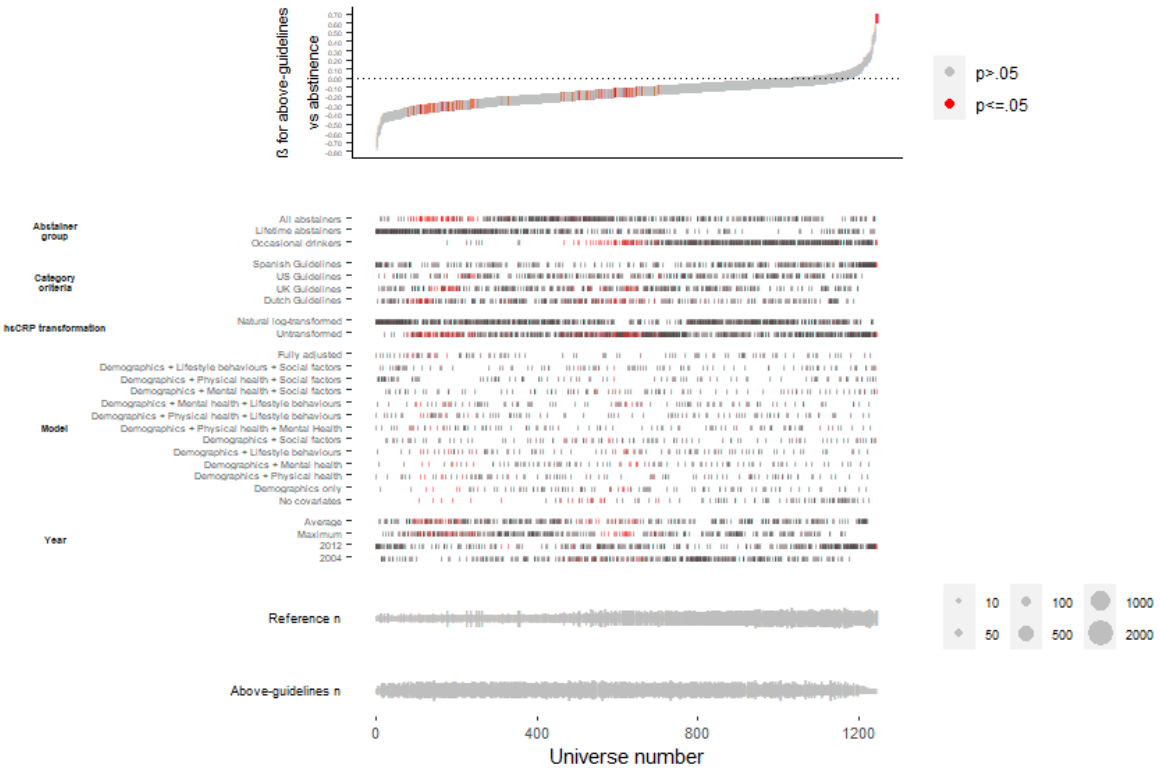
Appendix 2.6. Specification curve plot of estimates comparing low-to-moderate drinkers with an abstaining reference group, incorporating data from all three reference group specifications.



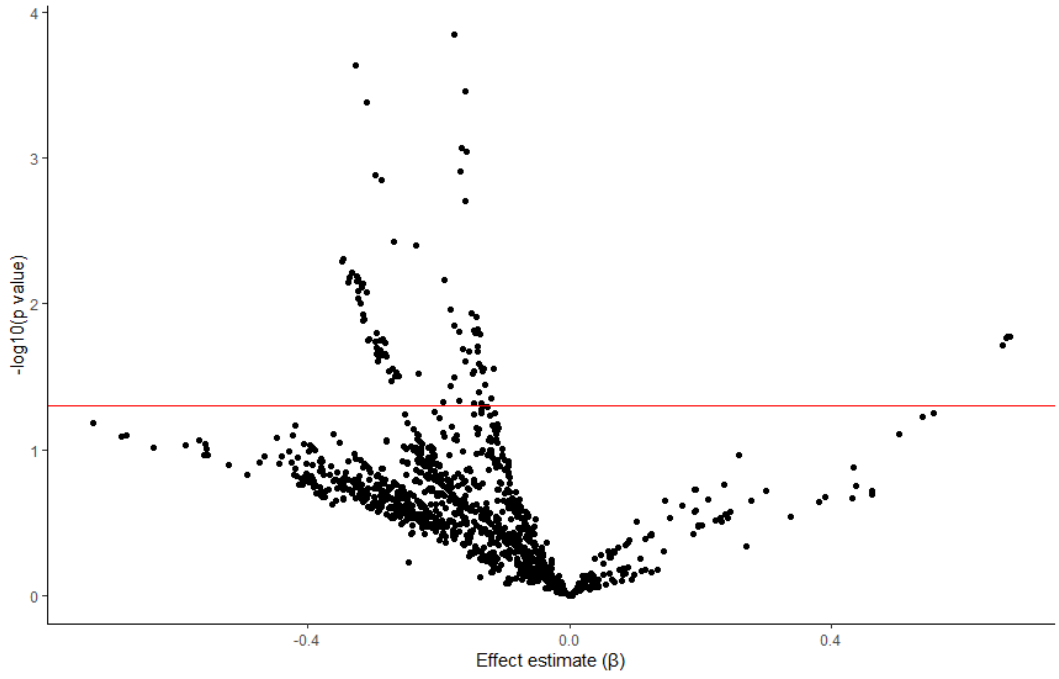
Appendix 2.7. Vibration of effects volcano plot depicting effects of low-to-moderate drinking and their *p* values compared with an abstaining reference group, incorporating data from all three reference group specifications. The red line indicates the log10-transformed number corresponding to a *p* value of .05.



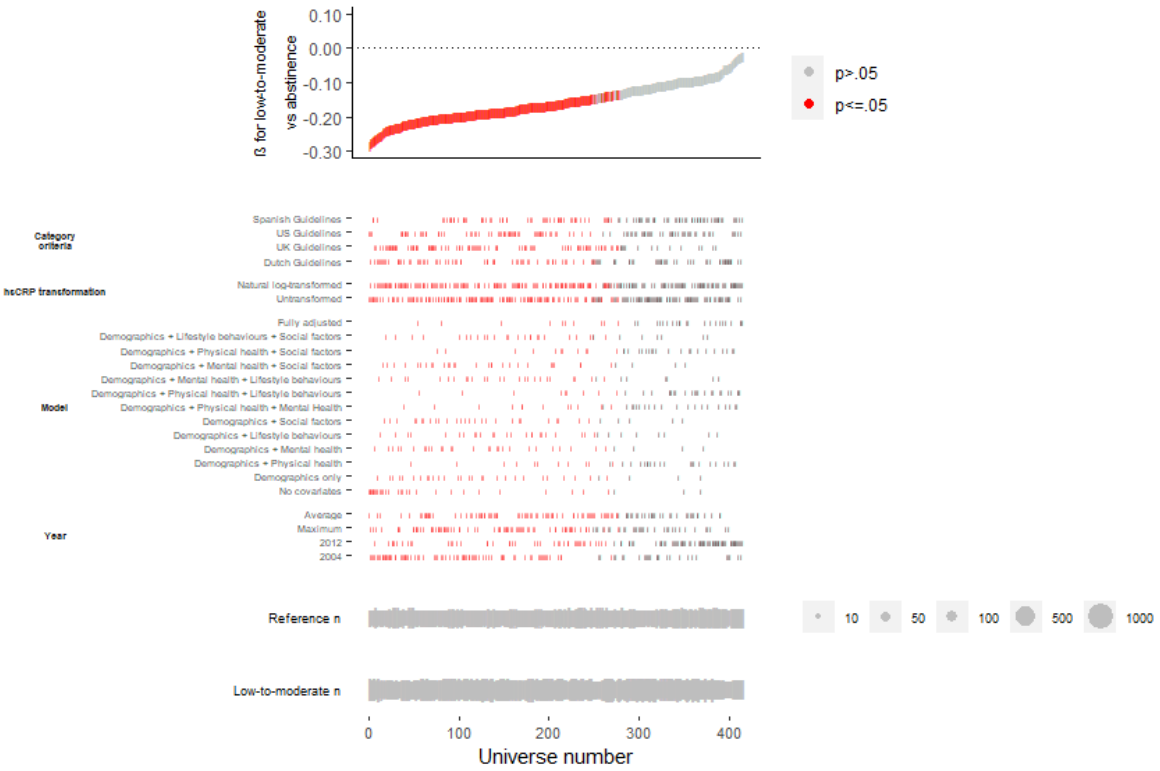
Appendix 2.8. Specification curve plot of estimates comparing above-guidelines drinkers with an abstaining reference group, incorporating data from all three reference group specifications.



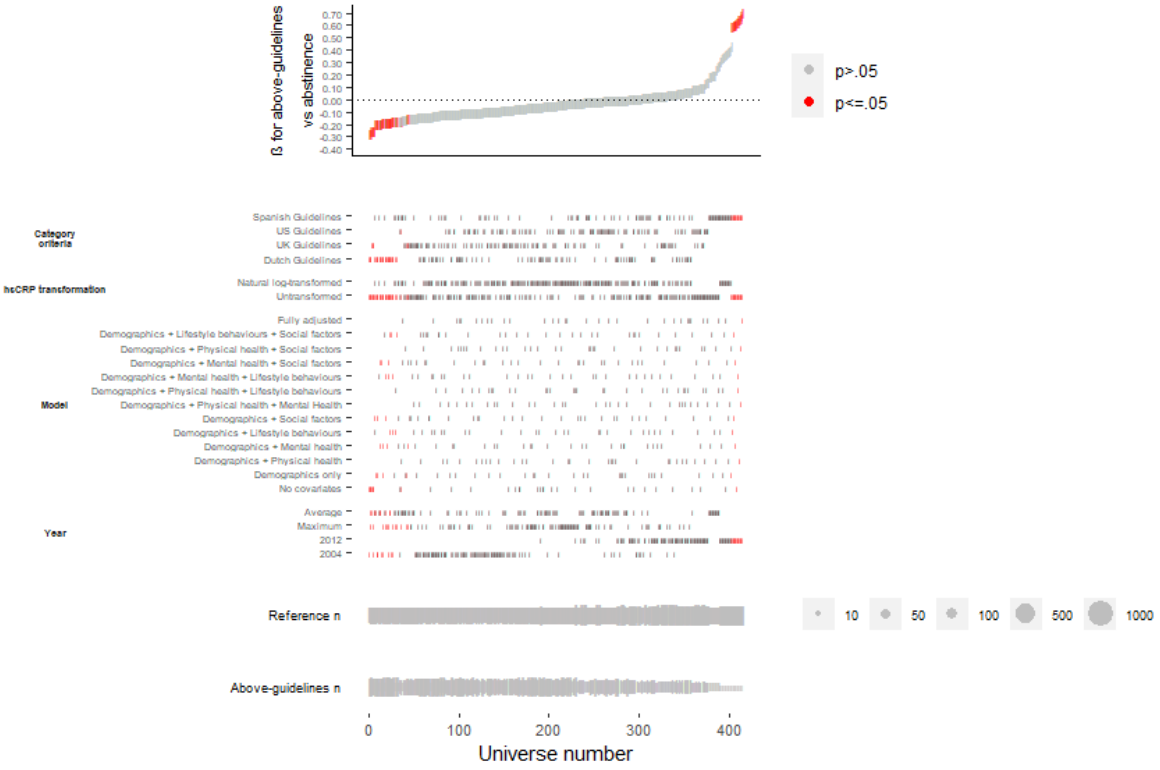
Appendix 2.9. Vibration of effects volcano plot depicting effects and their p values of above-guidelines drinking compared with an abstaining reference group, incorporating data from all three reference group specifications. The red line indicates the log10-transformed number corresponding to a p value of .05.



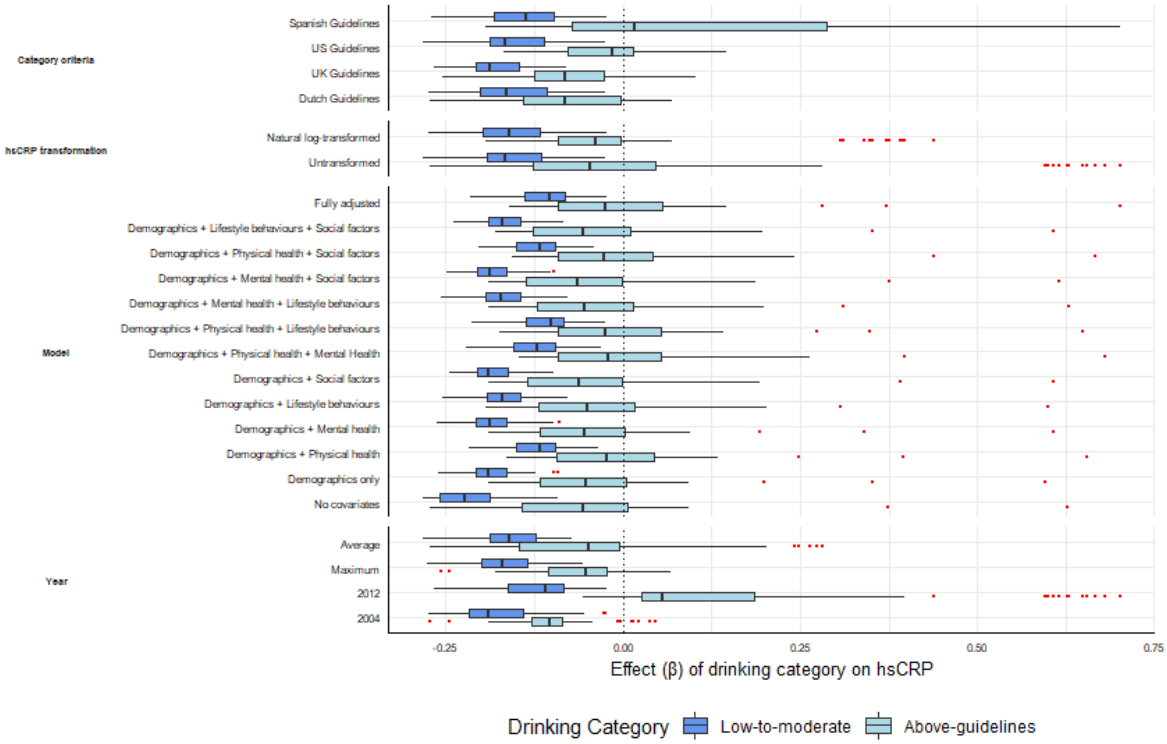
Appendix 2.10. Specification curve plot of estimates comparing low-to-moderate drinkers with a reference group composed only of occasional drinkers, from sensitivity analyses excluding those with any missing covariate data.



Appendix 2.11. Specification curve plot of estimates comparing above-guidelines drinkers with a reference group composed only of occasional drinkers, from sensitivity analyses excluding those with any missing covariate data.



Appendix 2.12. Box and whisker plot reflecting the impact of particular specifications on estimates comparing hsCRP between low-to-moderate/above-guidelines alcohol consumption and occasional drinking, from sensitivity analyses excluding those with any missing covariate data.



Red dots represent outliers.

Supplementary References

McGee, R., Williams, S., & Silva, P. A. (1986). An evaluation of the Malaise Inventory. *Journal of Psychosomatic Research*, *30*(2), 147–152.

Appendix 3

Supplementary materials for Chapter 4

Appendix 3.1. Supplementary Methods

Measurement occasions

The NLSY-79 introduced a suite of additional measurements to be completed by participants at ages 40 and 50, i.e., at the scheduled wave coinciding with the year when they reach the relevant age. While the CES-D-SF was initially tied to year-based measurement i.e., the 1992 and 1994 surveys, it was then moved to the suite of health measurements at ages 40 (conducted between 1998 and 2006) and 50 (conducted between 2008 and 2016). Because CES-D-SF score is one of the time-varying covariates in our model, we decided to change all time-varying covariates after 1994 to be those measured at waves coinciding with age 40 measurement (as opposed to a particular calendar year). To mitigate the impacts of this change, age was included as a covariate in the IPTW models for 1994 and 2002 alcohol consumption, and participants were considered censored if their designated 'age 40' or 'age 50' questionnaire occurred when their actual age at measurement deviated from the target by more than one year (i.e., only those aged 39-41, and 49-51 respectively were eligible).

Distribution of CES-D-SF scores at age 50

Given a large number of 0 values on the CES-D-SF (i.e., complete absence of depressive symptoms), a Shapiro-Wilk test of normality in the baseline sample was performed. This revealed a non-normal distribution ($W=0.81$, $p<0.001$), but this skew value is not considered large enough to warrant transformation or a non-parametric statistical test (Kim, 2013).

Variable derivation: baseline time-fixed variables

Time-fixed historical alcohol consumption was based on pre-baseline measurement occasions at 1983, 1984, 1985, 1988, 1989 and 1992, and condensed into a single four-category variable consisting of: some past occasional drinking, some past moderate drinking, some past above-guidelines drinking, or lifetime abstention (abstinent at all waves *and* reported no past lifetime consumption). Individuals missing more than one wave of pre-1994 data, and with no confirmed above-guidelines drinking, were excluded. No information on heavy episodic drinking was available at the 1992 wave.

Sex was provided at the 1979 initial interview, with male and female the only options. The binary ever-smoked variable was based on 1992 measurement, as was the ever-use of illicit drugs variable (derived from answers to separate questions about cocaine, cannabis, and crack cocaine. Frequent previous religious service attendance was also a binary variable, coded as yes if the individual attended services at least monthly in either 1979 or 1982. Race, as assessed in 1979, was a categorical variable with three options: Hispanic, Black, or Non-Black and Non-Hispanic. Educational attainment was a continuous variable based on response in 1992, and average parental educational attainment was a continuous variable averaging mother and father's years of education based on response in 1992.

Variable derivation: time-varying variables

The age variables at each wave were calculated based on age given in 1981, as recommended by the NLSY79 itself. Self-reported health limitation was a binary variable reflecting subjective assessment of whether a health limitation impacted the individual's working

ability (excluding cases for whom this limitation is pregnancy only). The marital status variables had three categories: never married, married (both of these options were present in the original question), and then a condensed category of separated/divorced/widowed. Smoking status (N/A at age 40) was a binary current smoking or not variable, as was illicit drug use (derived from answers to separate questions about cocaine, cannabis, and crack cocaine). BMI was a continuous variable and was winsorised at the 95th percentile. Employment status was a binary variable (currently employed or unemployed), as was health insurance status (currently possess or not), urban/rural-dwelling, and current receipt of any form of welfare. Household size was a continuous variable indicating number of individuals in the household (including the participant). Income was a continuous variable representing the individual's income, or if they had a spouse, an average of combined income, and was winsorised at the 95th percentile.

R packages and scripts

The key R packages used for these analyses were *WeightIt* (for generating weights; Cole & Hernán, 2008; Greifer, 2020; Imai & Ratkovic, 2014), *cobalt* (for generating plots of covariate balance before and after weighting) (Greifer, 2017), *multcomp* (for running contrast analyses; Hothorn et al., 2008), *boot* (for bootstrapping results; Canty & Ripley, 2017; Davison & Hinkley, 1997), and *EValue* (for generating E-values; Mathur & VanderWeele, 2019; Smith & VanderWeele, 2019; VanderWeele & Ding, 2017). An example script for generating IPTWs is provided below. Code for any other steps discussed in this paper is available from the authors on request.

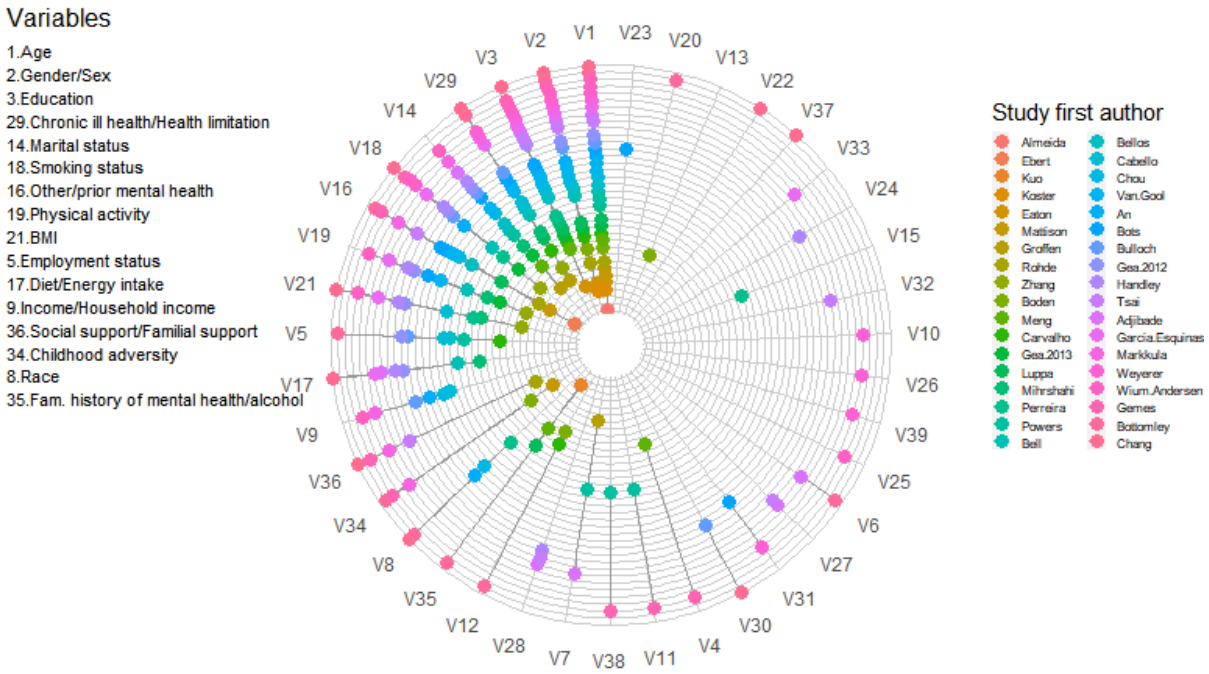
```{r}

*#Making propensity score formulae for use in weight generation*

fixedVars<- c ("Historical_alcoholCategory", "Sex", "EverSmoked", " Everused_illicit", " Self_esteem_historical", "
Relgious_attendance_historical", "RACE"," Education_years", "Parental_education_years")

Vars92<- c ("CESD_92", "Age92", "HealthLimit92", "Marital92", "SMOKE92", "Illicit92", "BMI92", "Labour92", "Insurance92",
"Household_size92", "Urban92", "Welfare92", "Income92", "I(CESD_92^2)","I(Household_size92^2)")

Vars94<- c ("alcoholCategory_94","CESD_94","Age94", "HealthLimit94","Marital94","SMOKE94",
"Illicit94","BMI94","Labour94","Insurance94","Household_size94","Welfare94","Urban94", "Income94",
"I(CESD_94^2)","I(Household_size94^2)")

Vars40<- c ("alcoholCategory_02", "CESD_Age40",
"HealthLimit40","Marital40","BMI40","Labour40","Insurance40","Household_size40", "Welfare40", "Urban40","Income40",
"I(CESD_Age40^2)","I(Household_size40^2)")

fixedVars <- paste(fixedVars, collapse ="+")
Vars92 <- paste(Vars92, collapse ="+")
Vars94 <- paste(Vars94, collapse ="+")
Vars40 <- paste(Vars40, collapse ="+")

full94_IPTW<-paste("alcoholCategory_94 ~ ",paste(fixedVars,Vars92,sep="+"), collapse="")
full02_IPTW<-paste("alcoholCategory_02 ~ ",paste(fixedVars,Vars92,Vars94,sep="+"), collapse="")
full06_IPTW<-paste("alcoholCategory_06 ~ ",paste(fixedVars,Vars92,Vars94,Vars40,sep="+"), collapse="")

*#Generating exposure weights for first wave*

Tweights94_surv <- weightit(as.formula(full94_IPTW),
                    data=IPTW_df,
                    method="cbps",
                    over=FALSE,
                    stabilize=TRUE,
                    s.weights=IPTW_df$NORMSurvWeight)

*#Joining these weights back to the data frame*

IPTW_df<-bind_cols(IPTW_df, Tweights94_surv[["weights"]])

*#Generating exposure weights for second wave*
```

```
IPTW_df_uncensored1<-IPTW_df %>% filter(CENSOR1==0)
IPTW_df_censored1<-IPTW_df %>% filter(CENSOR1==1)

full02_IPTW_list<-list(as.formula(full02_IPTW))
Tweights02_surv <- weightitMSM((full02_IPTW_list),
                  data=IPTW_df_uncensored1,
                  method="cbps",
                  over=FALSE,
                  stabilize=TRUE,
                  num.formula=~alcoholCategory_94,
                  s.weights=IPTW_df_uncensored1$NORMSurvWeight)

#Joining these weights back to the data frame
IPTW_df_censored1<-IPTW_df_censored1 %>% mutate(Tweights02_surv=0)
IPTW_df_uncensored1<-bind_cols(IPTW_df_uncensored1, Tweights02_surv[["weights"]])
IPTW_df<-bind_rows(IPTW_df_censored1, IPTW_df_uncensored1)

#Generating exposure weights for third wave
IPTW_df_uncensored2<-IPTW_df %>% filter(CENSOR2==0)
IPTW_df_censored2<-IPTW_df %>% filter(CENSOR2==1)

full06_IPTW_list<-list(as.formula(full06_IPTW))
Tweights06_surv <- weightitMSM((full06_IPTW_list),
                  data=IPTW_df_uncensored2,
                  method="cbps",
                  over=FALSE,
                  stabilize=TRUE,
                  num.formula=~alcoholCategory_94+alcoholCategory_02,
                  s.weights=IPTW_df_uncensored2$NORMSurvWeight)

#Joining these weights back to the data frame
IPTW_df_censored2<-IPTW_df_censored2 %>% mutate(Tweights06_surv=0)
#now bind weights from above to IPTW_df_censored using col bing
IPTW_df_uncensored2<-bind_cols(IPTW_df_uncensored2, Tweights06_surv[["weights"]])
#Then bind the 2 dfs together so you have both censored and uncensored together again
IPTW_df<-bind_rows(IPTW_df_censored2, IPTW_df_uncensored2)
```

Sensitivity analyses: Stratification by sex

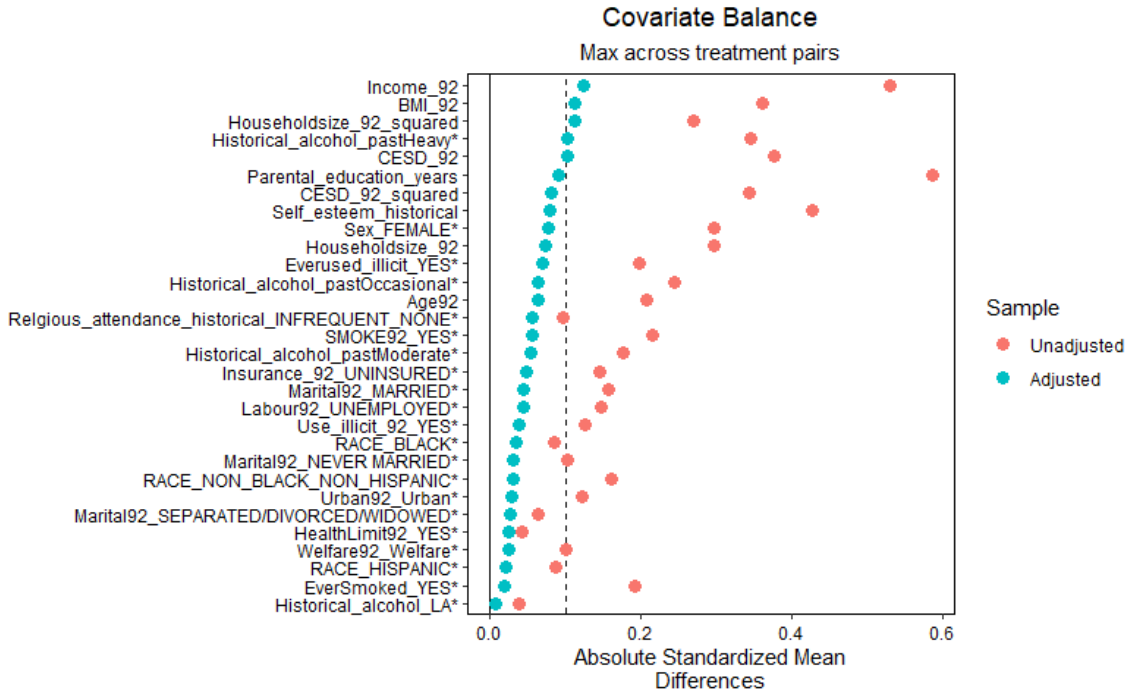
Analyses were re-performed separately for the female and male sub-samples, beginning with separate weight generation in each sub-sample. Unlike for the main analysis, sex was not included as a covariate in IPTW/IPCW models, and survey weights not incorporated, as subsamples are not intended to be representative of the general population. For females, weights for the final analysed sample ranged from 0.11 to 10.92, with a mean of 1.55 and standard deviation of 2.00. For males, weights in the final analysed sample ranged from 0.13 to 7.69, with a mean of 1.34 and standard deviation of 1.51.

Appendix 3.2. Covariates controlled for in studies included in the Li et al. meta-analysis

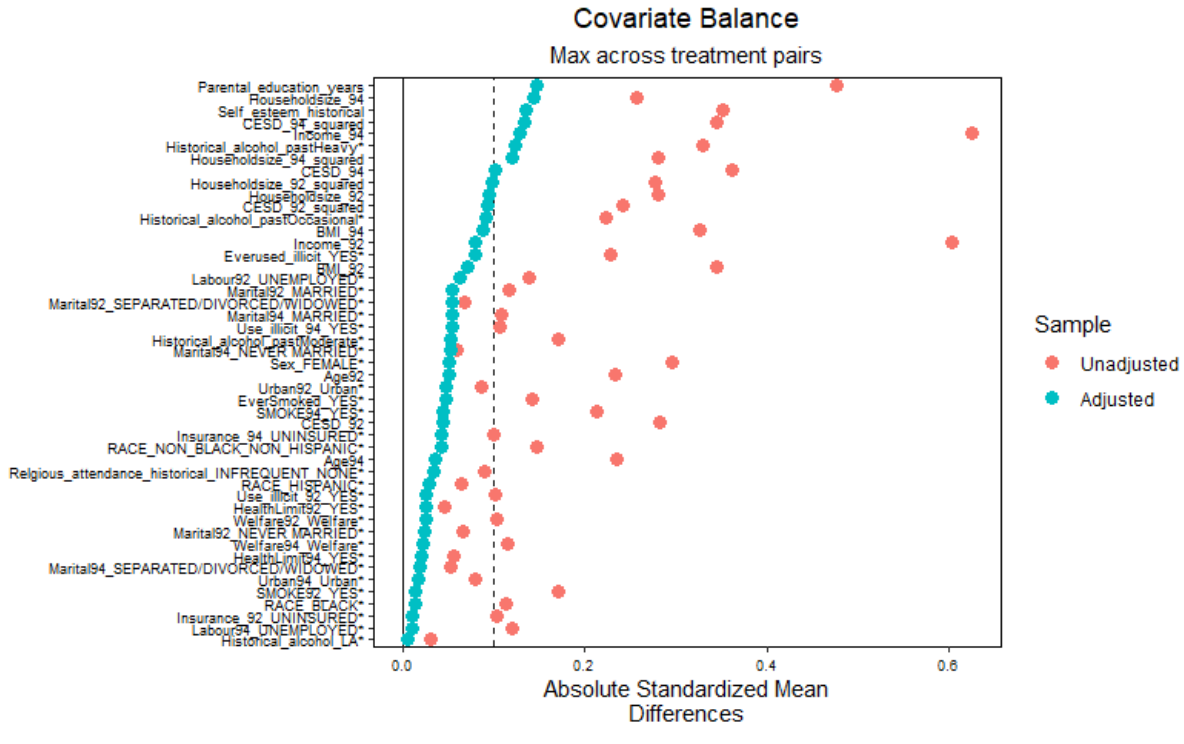


Appendix 3.3. Covariate balance between exposure groups at 1994, 2002, and 2006 before and after weighting by the final weight[a]

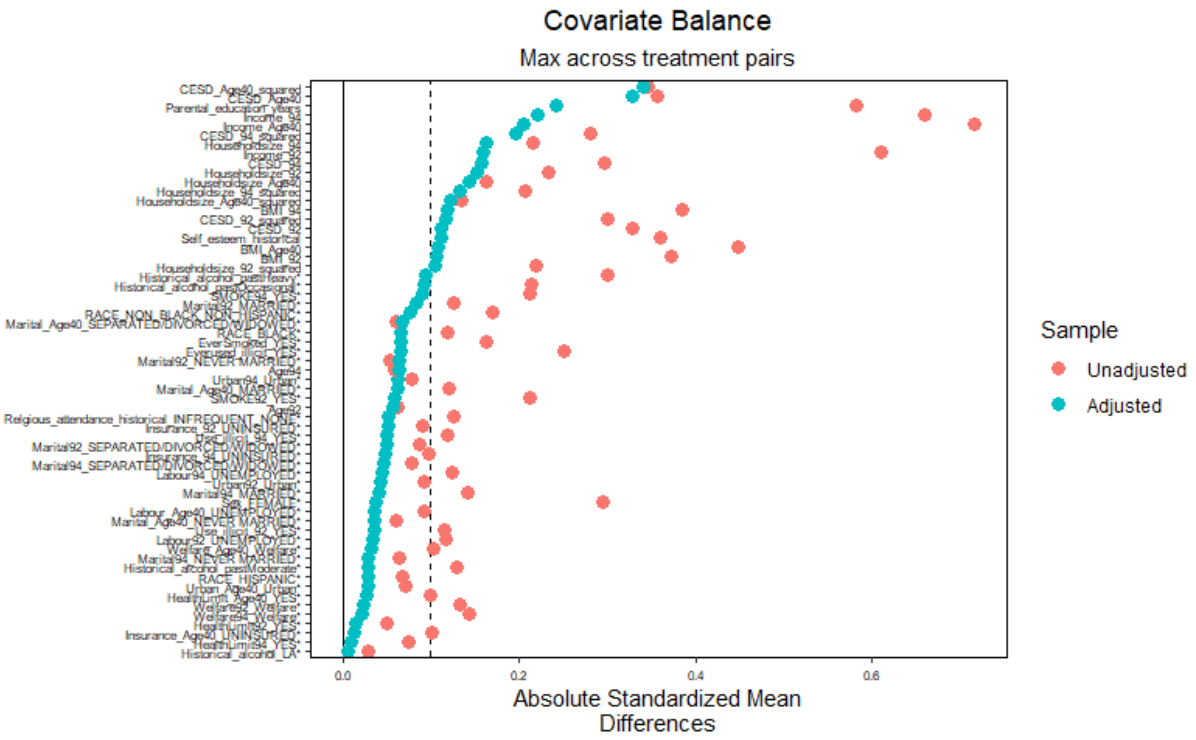
A) 1994



B) 2002



C) 2006



[a] Figures display differences in means for the two exposure groups for which these differences are largest. Variables followed by an asterisk are unstandardised categorical variables. Covariate names (and their categories) are abbreviated for brevity: 'illicit' stands for illicit drugs; 'HealthLimit' stands for self-reported health limitations; 'Historical_alcohol_LA' stands for lifetime abstainer at baseline.

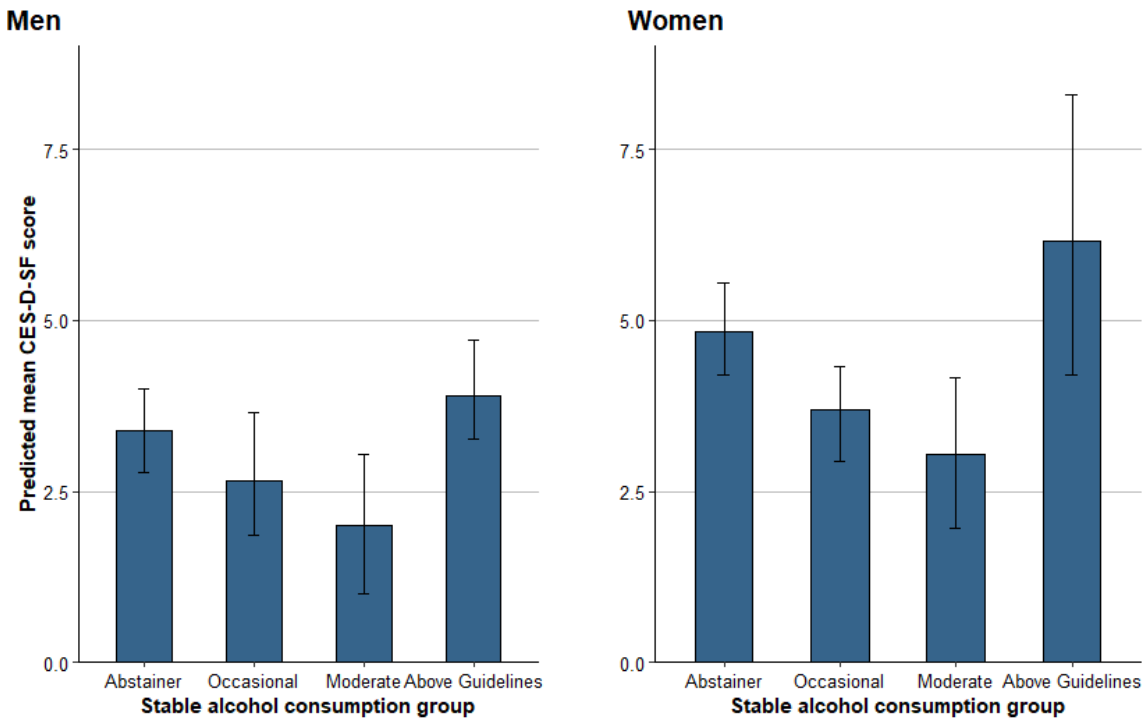
Appendix 3.4. Output from contrasts for analyses using continuous CES-D-SF scores as outcome

Contrasts	**Estimates** (b)	**Std. Error**	***P***	**Bootstrapped CI**
Consistent occasional vs Consistent abstinence	-0.84	0.26	0.001	**-1.63, -0.11**
Consistent moderate vs Consistent abstinence	-1.08	0.30	<0.001	**-1.88, -0.20**
Consistent above-guidelines vs Consistent abstinence	0.34	0.25	0.162	-0.62, 1.25

Appendix 3.5. Output from contrasts for analyses using binary probable depression as outcome

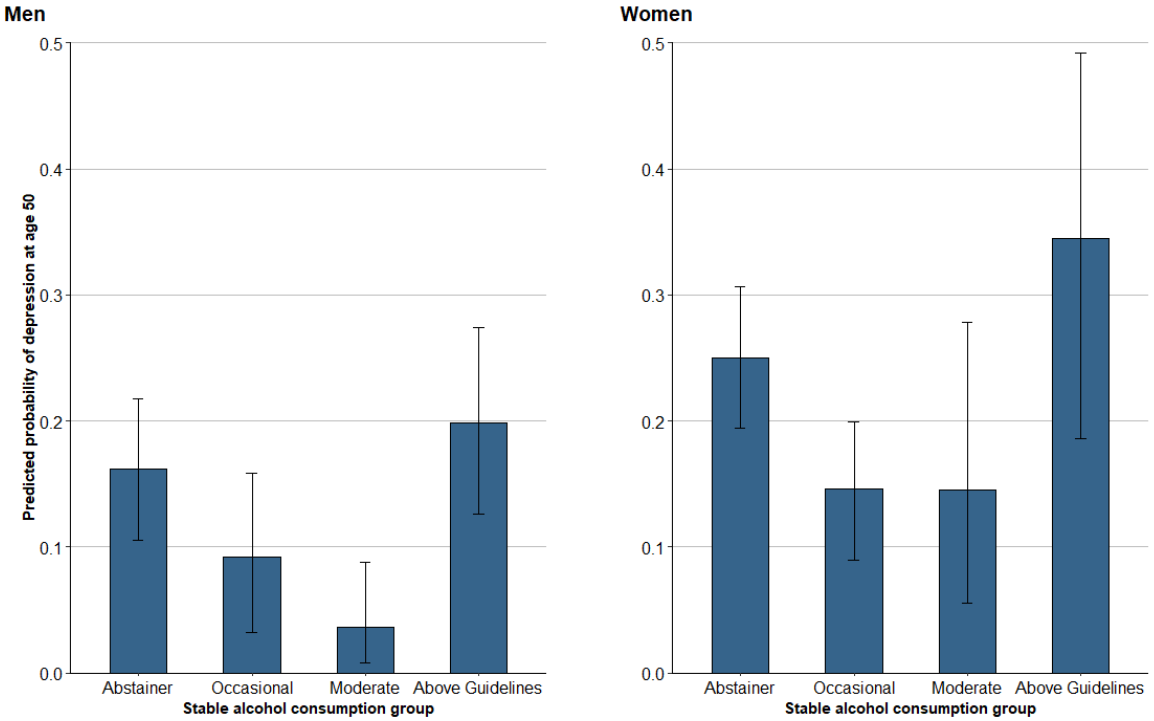
Contrasts	**Estimates** (OR)	***P***	**Bootstrapped CI**
Consistent occasional vs Consistent abstinence	0.58	<0.001	**0.36, 0.88**
Consistent moderate vs Consistent abstinence	0.59	0.005	0.26, 1.13
Consistent above-guidelines vs Consistent abstinence	1.06	0.683	0.66, 1.72

Appendix 3.6. Predicted mean CES-D-SF scores at age 50 for those consistently abstaining, drinking occasionally, moderately or above guidelines over the 1994, 2002 and 2006 measurement occasions, stratified by sex[a]



[a] Whiskers indicate bootstrapped 95% CIs. Groups do not possess a specific number of subjects as they represent hypothetical trajectories.

Appendix 3.7. Predicted probability of probable depression at age 50 for those consistently abstaining, drinking occasionally, moderately or above guidelines over the 1994, 2002 and 2006 measurement occasions, stratified by sex[a]



[a] Whiskers indicate bootstrapped 95% CIs. Groups do not possess a specific number of subjects as they represent hypothetical trajectories.

Supplementary References

Canty, A., & Ripley, B. (2017). boot: Bootstrap R (S-Plus) functions. *R Package Version*, *1*, 3–20.

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, *168*(6), 656–664. https://doi.org/10.1093/aje/kwn164

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application* (Issue 1). Cambridge university press.

Greifer, N. (2017). cobalt: Covariate balance tables and plots. *R Package Version*, *2*(0).

Greifer, N. (2020). WeightIt: weighting for covariate balance in observational studies. *R Package Version 0.9. 0*.

Hothorn, T., Bretz, F., Ag, P., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*(3), 346–363.

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243–263.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52-54. https://doi.org/10.5395/rde.2013.38.1.52

Mathur, M. B., & VanderWeele, T. J. (2019). Sensitivity analysis for unmeasured confounding in meta-analyses. *Journal of the American Statistical Association, 115*(529), 163-172.

Smith, L. H., & VanderWeele, T. J. (2019). Bounding bias due to selection. *Epidemiology (Cambridge, Mass.)*, *30*(4), 509.

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine*, *167*(4), 268–274.

Appendix 4

Supplementary materials for Chapter 5

Appendix 4.1 Supplementary methods

Quality control and assumption checking in generating genetic instruments

After identifying the 93 SNPs significantly associated with drinks per week in the discovery GWAS, the following steps were performed.

First, clumping was performed to identify SNPs in linkage disequilibrium (LD; defined as r^2>.01), and then pruned, removing 14 SNPs. Next, SNPs with a minor allele frequency (MAF) <.01 were identified (none for the instruments based on MAFs in GSCAN data; none for the GRS based on MAFs in UKB data). Next, one SNP without an associated RSID was removed, bringing the set of SNPs down to 84. Then, for the two-sample instruments, palindromic SNPs with a MAF >.42 were removed in order to ensure accurate strand harmonisation between gene-exposure and gene-outcome data (2 SNPs removed). As the inclusion of SNPs associated with confounders could result in violation of instrumental variables assumption 2 (exchangeability) and 3 (exclusion restriction), the remaining 82 SNPs were run through Phenoscanner to identify significant associations with potential confounders of the alcohol–diabetes relationship (smoking, physical activity, and education-related variables; BMI/adiposity and diet were not included as they may be mediators). This resulted in the removal of a further 5 SNPs, and a final set of instruments composed of 77 SNPs. For the one-sample NLMR, after finding no SNPs with a MAF <.01, the remaining 84 SNPs were run through Phenoscanner resulting in the removal of 5 SNPs. None of the remaining SNPs had an information score <.4 (reflecting imputation quality) in UK Biobank, leaving 79 SNPs to comprise the GRS (Appendix 4.2). The F statistic for the GRS was 22.80, which is above the F>10 rule-of-thumb to satisfy the relevance assumption

(although weak instrument bias is unlikely with a GRS in any case; Burgess & Thompson, 2021).

Initially, SNP discovery (identification of variants significantly associated with drinks per week at a genome-wide level; $p < 5 \times 10^{-8}$) had been performed using results in those of European descent from the recent update to GSCAN in 2022 – the largest published GWAS available (N up to 2,428,851 depending on variants).[43] For two-sample MR, gene-exposure summary statistics were also taken from GSCAN 2022. This iteration of GSCAN identified 501 SNPs significantly associated with drinks per week. First, clumping was performed to identify SNPs in linkage disequilibrium (LD; defined as $r^2 > .01$), and then pruned, removing 121 variants. Next, SNPs with a minor allele frequency (MAF) <.01 were identified (none for the instruments based on MAFs in GSCAN data; none for the GRS based on MAFs in UKB data). Then, for the two-sample instruments, palindromic SNPs with a MAF >.42 were removed in order to ensure accurate strand harmonisation between gene-exposure and gene-outcome data (11 SNPs removed). The remaining 369 SNPs were run through Phenoscanner to identify significant associations with the same potential confounders described above. This resulted in the removal of a further 16 SNPs, and a final set of instruments composed of 353 SNPs. For the one-sample NLMR, after finding no SNPs with a MAF <.01, the remaining 380 SNPs were run through Phenoscanner resulting in the removal of 17 SNPs. None of the remaining SNPs had an information score <.4, leaving 363 SNPs to comprise the GRS. However, using these SNPs, MR Egger intercepts for both the HbA1C and T2D analyses were significantly different to zero, indicating that either the average pleiotropic effect was not zero or the pleiotropic effects were correlated with the gene–exposure association, meaning the assumptions required for the IVW estimate do not hold. Significant intercepts

were also found when performing MR Egger within strata used in the UKB non-linear MR analysis. This led to the decision to rely instead on SNPs identified in the older version of GSCAN, as detailed above.

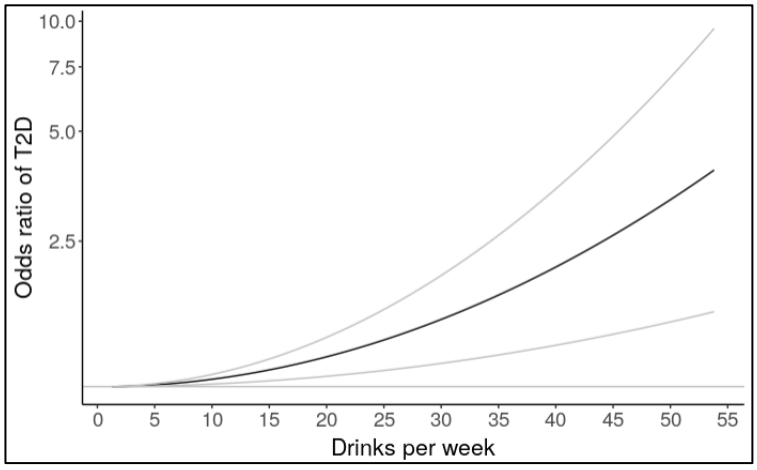
Appendix 4.2. SNPs included in genetic instruments.

Chromosome	SNP	Effect allele	Other allele	Beta	SE	Effect allele frequency
1	rs705687	A	G	-0.010904017	0.001775889	0.215000
1	rs58107686	C	A	-0.009747126	0.001585370	0.672000
1	rs12088813	A	C	-0.009329047	0.001649158	0.733000
1	rs5024204	A	T	0.009702711	0.001628467	0.722000
1	rs10753661	G	A	-0.008638128	0.001569272	0.316000
1	rs28680958	G	A	-0.010996457	0.001769942	0.783000
1	rs823114	G	A	0.008767740	0.001467417	0.447000
2	rs77165542	C	T	-0.026010899	0.003971245	0.965100
2	rs1260326	T	C	0.020889835	0.001488339	0.399000
2	rs13032049	A	G	0.010194550	0.001617979	0.717000
2	rs828867	G	A	0.008757157	0.001463597	0.455000
2	rs11692435	G	A	0.017449894	0.002615850	0.914800
2	rs72859280	G	T	0.022885324	0.003901915	0.963800
2	rs56337305	T	C	-0.009588381	0.001499283	0.617000
3	rs13094887	A	T	-0.010309671	0.001588926	0.699000
3	rs13066454	C	T	-0.008775946	0.001491956	0.602000
3	rs9838144	G	C	-0.009964223	0.001792521	0.791000
3	rs2011092	T	C	-0.008900405	0.001540056	0.661357
3	rs6787172	T	G	-0.008030894	0.001466234	0.446000
4	rs3748034	G	T	-0.011740341	0.002081935	0.857000
4	rs11940694	A	G	0.025949908	0.001485887	0.403000
4	rs4501255	C	G	0.010693415	0.001718941	0.765000
4	rs1229984	T	C	0.150533575	0.003861105	0.037000
4	rs36052336	A	G	-0.018428809	0.003033780	0.938504
4	rs2165670	G	A	0.023080380	0.002364267	0.893663
4	rs79139602	A	T	0.060272310	0.005075844	0.978939
4	rs4699791	G	A	0.018570721	0.002477197	0.904274
4	rs4690727	C	G	0.010816985	0.001619716	0.282000
4	rs12651313	C	G	-0.008642927	0.001467223	0.557000
5	rs4916723	A	C	-0.009952257	0.001478674	0.584000
5	rs12655091	G	A	-0.008312109	0.001460289	0.470000
5	rs55872084	G	T	0.009978744	0.001718941	0.765000
5	rs11739827	G	T	-0.008375229	0.001469112	0.549000
7	rs6460047	T	C	0.011623538	0.001795690	0.792000
7	rs10236149	A	G	-0.013498145	0.002219081	0.877000
7	rs35034355	G	A	-0.008096817	0.001458945	0.479000
7	rs6951574	T	C	0.013222216	0.001462828	0.542000
8	rs13250583	C	T	-0.009717545	0.001780116	0.787000
8	rs1217091	T	C	0.012160827	0.001865386	0.188000
8	rs28601761	C	G	0.009102878	0.001476682	0.580000
9	rs55932213	A	G	0.009488271	0.001654204	0.263613

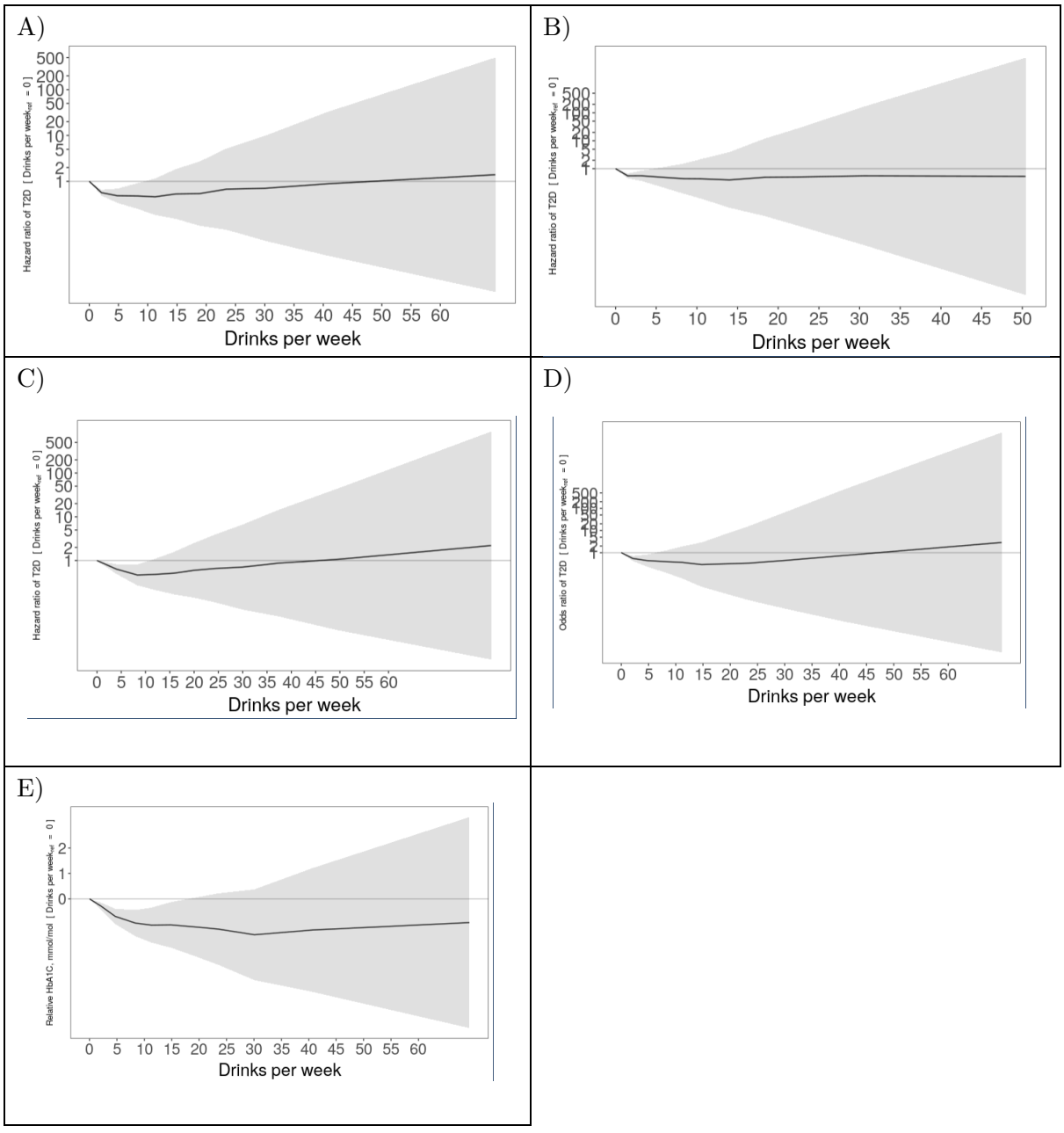
9	rs10978550	T	C	-0.011748331	0.001802112	0.794000
10	rs7074871	G	A	-0.009399850	0.001672158	0.745000
10	rs17665139	C	T	-0.011560138	0.002046764	0.851000
11	rs7950166	C	T	-0.009799202	0.001515663	0.363000
11	rs56030824	G	A	-0.011601888	0.001562978	0.678000
11	rs10750025	C	T	0.010321447	0.001570358	0.314000
11	rs1713676	A	G	-0.007991984	0.001459130	0.477546
11	rs4938230	C	A	0.012810377	0.001998209	0.158000
11	rs682011	T	C	0.008212447	0.001467913	0.441000
11	rs12795042	A	C	-0.008319009	0.001503872	0.377000
12	rs10876188	C	T	-0.007986853	0.001463078	0.543000
12	rs3809162	A	G	0.009060925	0.001489607	0.603000
12	rs10506274	G	T	-0.009037403	0.001458405	0.516000
12	rs4842786	G	A	-0.008797801	0.001478674	0.416000
13	rs500321	A	T	-0.009669361	0.001653426	0.264000
14	rs1123285	C	G	-0.008897365	0.001544160	0.665000
14	rs2180870	T	C	-0.012177537	0.002132803	0.865000
14	rs28929474	C	T	-0.036799624	0.005437642	0.981700
14	rs11625650	G	A	-0.009568091	0.001724050	0.767000
15	rs2472297	C	T	0.010606047	0.001685411	0.751000
15	rs12907323	A	G	0.008496596	0.001481314	0.589000
16	rs2764771	G	A	0.009890596	0.001581736	0.693000
16	rs17177078	C	T	-0.022315490	0.003011759	0.937400
16	rs378421	G	A	-0.011205474	0.001486812	0.596000
16	rs113443718	G	A	-0.010208258	0.001584629	0.695000
16	rs62044525	C	G	-0.012173180	0.001882851	0.816000
16	rs7185555	G	C	-0.011100514	0.002026667	0.847000
16	rs79616692	G	C	0.016302272	0.002350583	0.892000
16	rs1104608	G	C	-0.010974988	0.001475848	0.575000
17	rs4548913	G	A	-0.008359953	0.001511274	0.368000
17	rs3803800	A	G	0.011378876	0.001777082	0.214000
17	rs2854334	A	G	0.009220984	0.001497813	0.385000
17	rs10438820	C	T	0.008971778	0.001593488	0.298000
18	rs9950000	C	T	-0.009117697	0.001490903	0.605000
18	rs4092465	A	G	-0.008291878	0.001513883	0.365000
19	rs281379	G	A	0.013722336	0.001457844	0.492000
20	rs4815364	G	A	0.008582356	0.001498544	0.384000
22	rs9607814	C	A	-0.010181001	0.001858788	0.800000

*Highlighted SNPs were not included in two-sample analyses (palindromic SNPs with a minor allele frequency >.42).

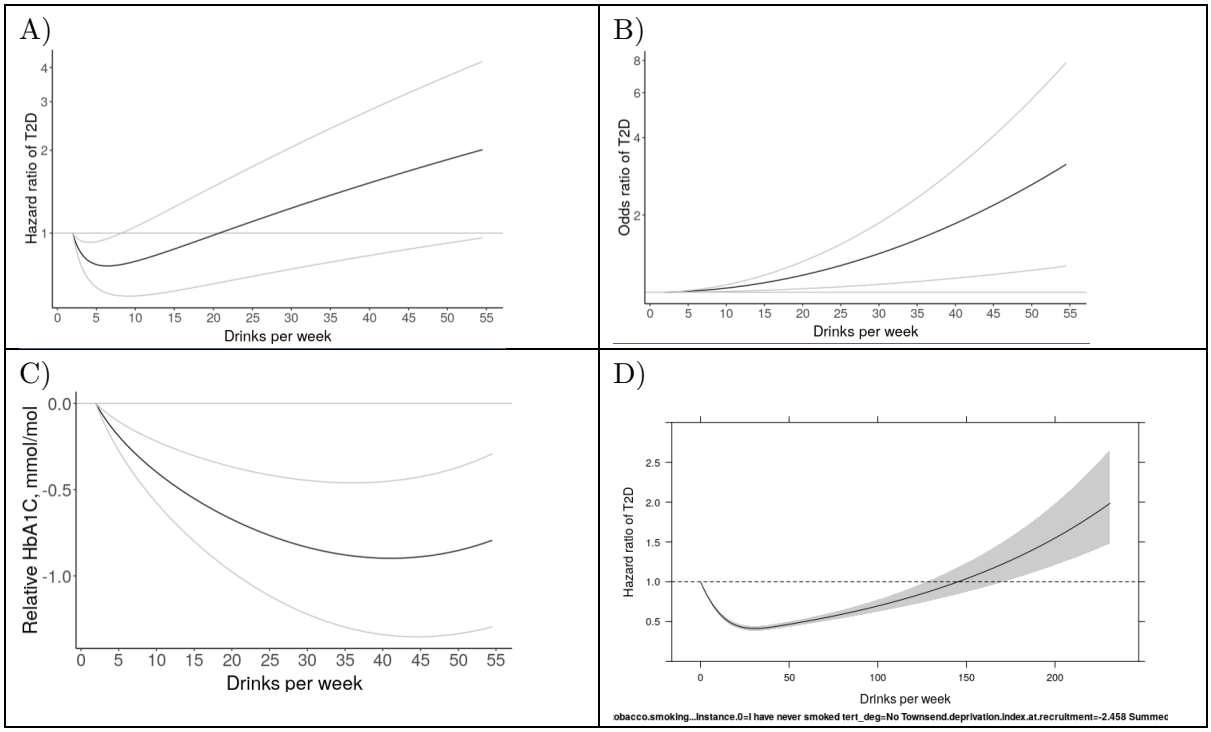
Appendix 4.3. Best-fitting fractional polynomial depicting genetic associations between alcohol consumption and baseline prevalent T2D.



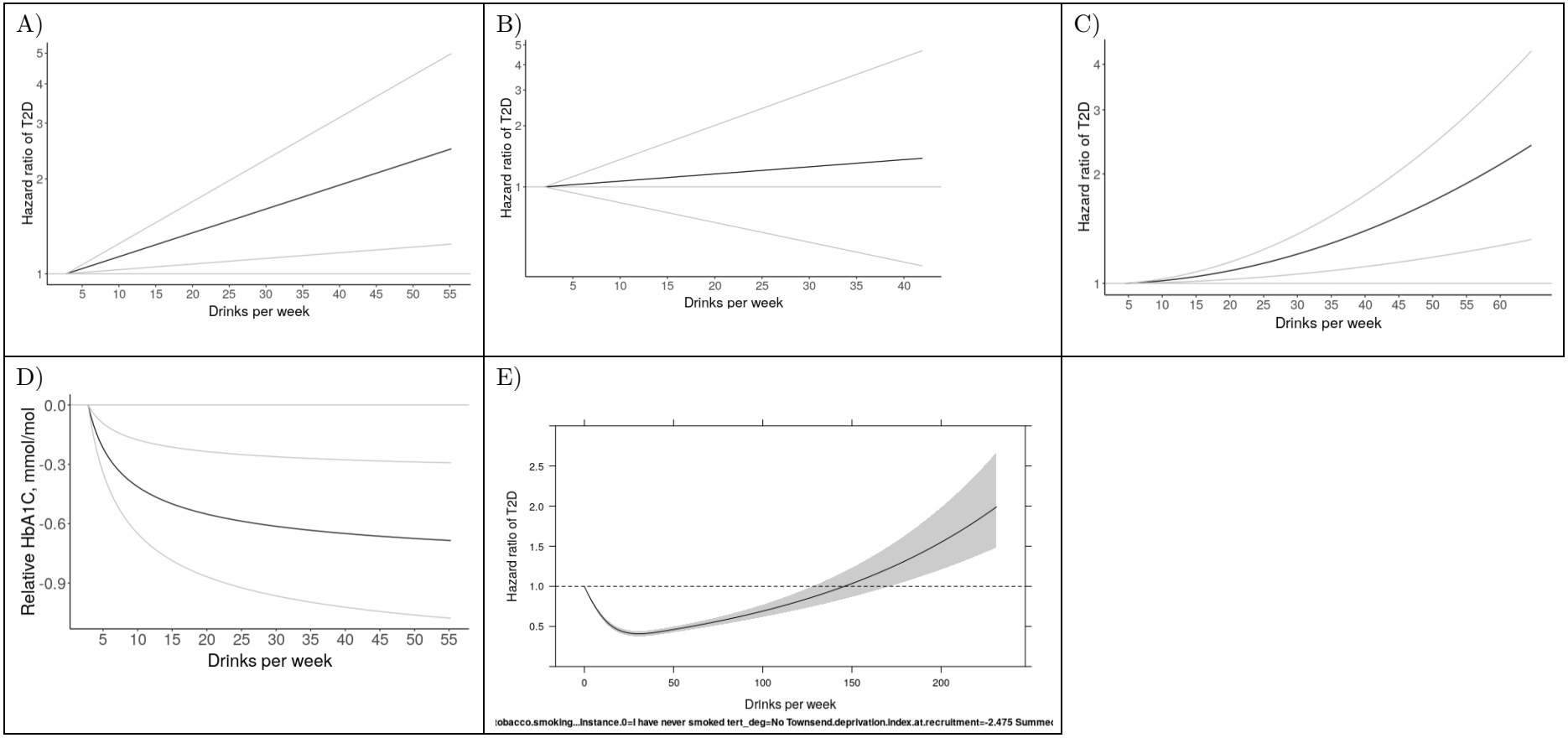
Appendix 4.4. Piecewise linear models for the A) the main survival model; B) the survival model in females; C) the survival model in males; D) the logistic model using baseline prevalent cases; E) the HbA1C model.



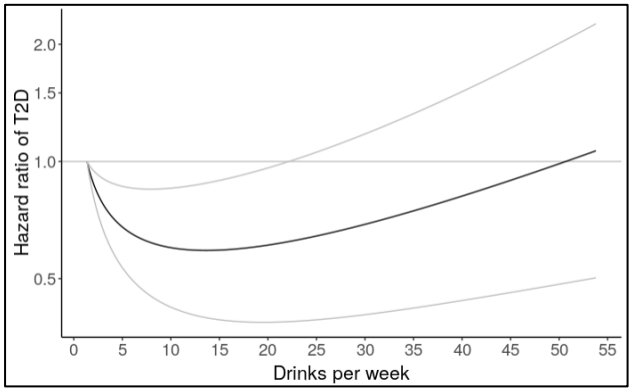
Appendix 4.5. Results for samples excluding former drinkers for A) the main survival model; B) the logistic model using baseline prevalent cases; C) the HbA1C model; D) the observational survival model.



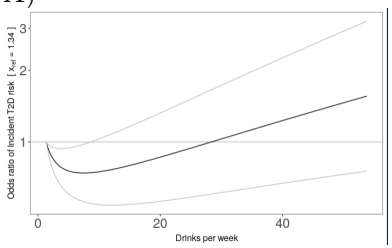
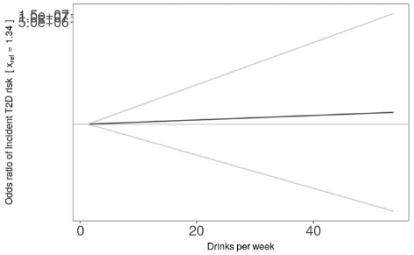
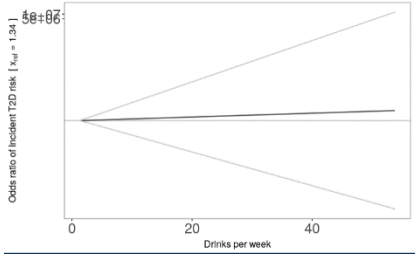
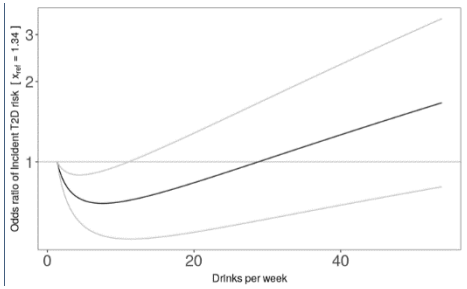
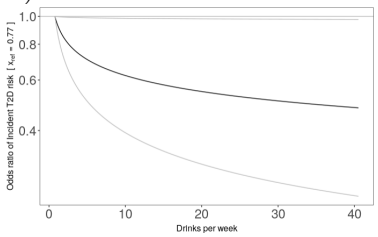
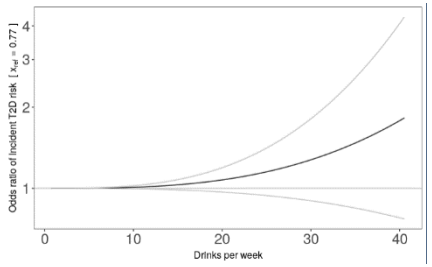
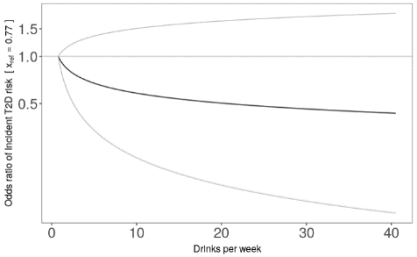
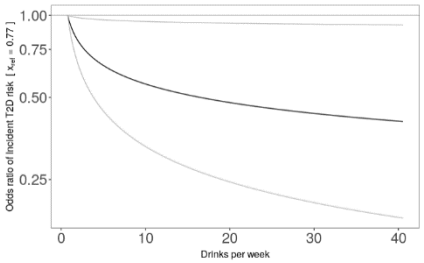
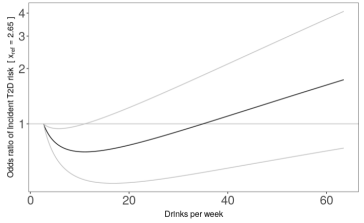
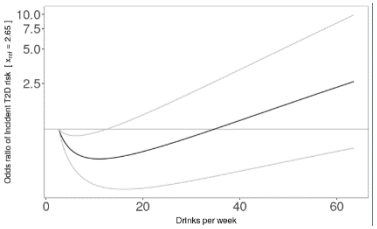
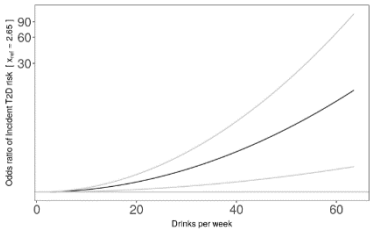
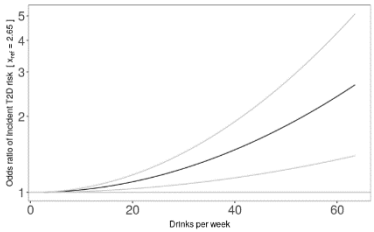
Appendix 4.6. Results for samples excluding former drinkers and never drinkers for A) the main survival model; B) the main survival model in females; C) the main survival model in males; D) the HbA1C model; E) the observational survival model.



Appendix 4.7. Results for the main non-linear MR survival analysis excluding T2D cases identified via primary care data.

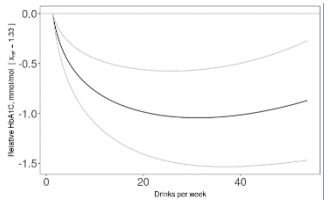
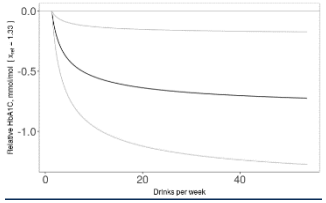
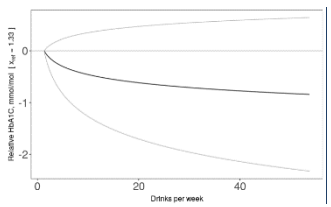
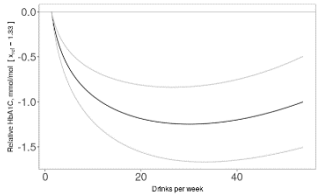
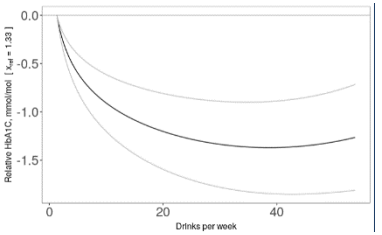


Appendix 4.8. SNP-based non-linear MR results for incident T2D risk in A) the main sample; B) females; C) males.

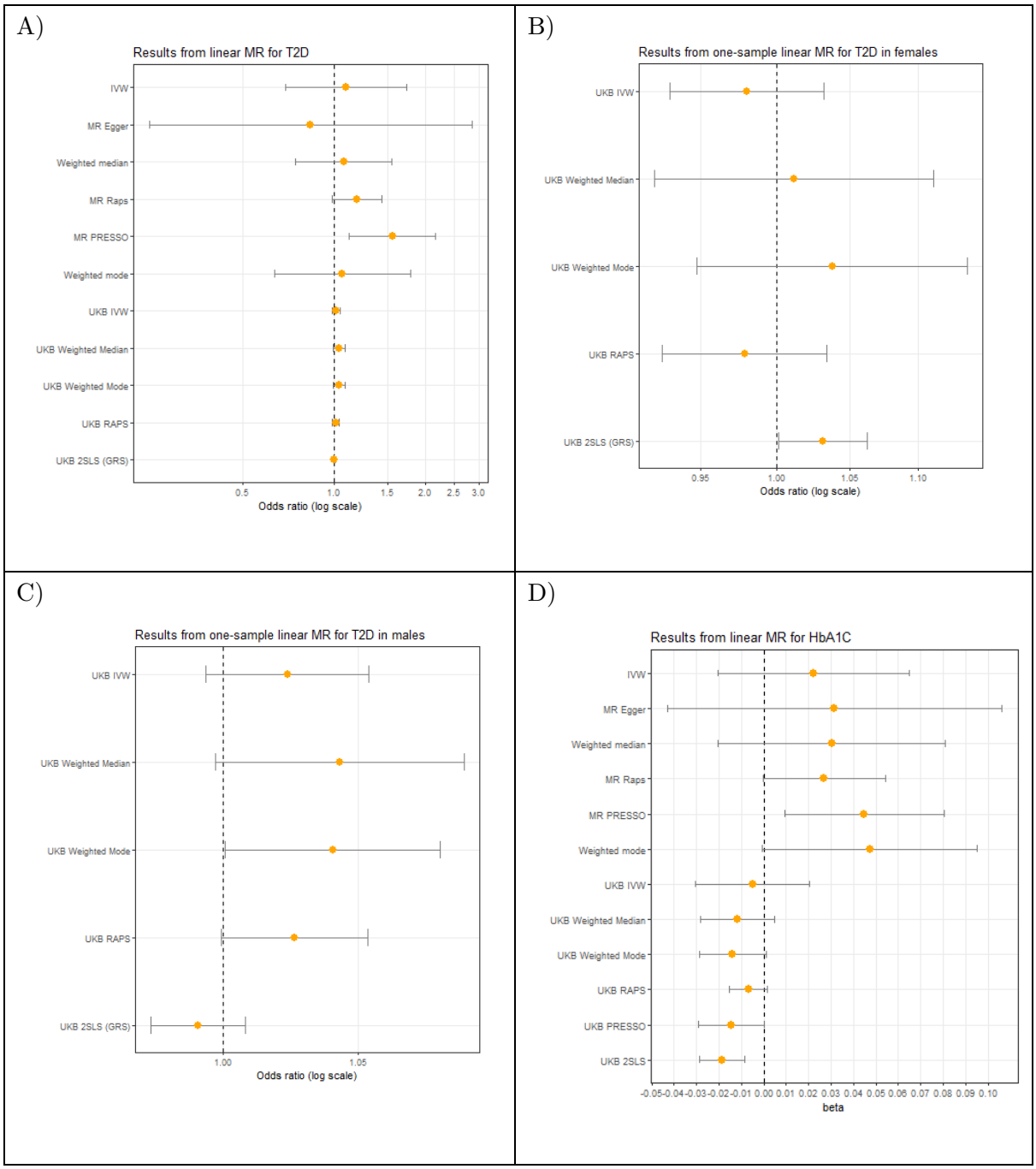
IVW	Weighted median	Weighted mode	MR RAPS
A) 			
B) 			
C) 			

IVW= Inverse-variance weighted. MR PRESSO global tests were not significant in the main sample nor in women, and no significant outliers were detected in men, therefore no outlier-corrected estimates were generated.

Appendix 4.9. SNP-based non-linear MR results for HbA1C in the main sample. IVW= Inverse-variance weighted.

IVW	Weighted median	Weighted mode	MR RAPS	MR PRESSO
				

Appendix 4.10. Results from linear MR for baseline prevalent T2D in A) the main sample; B) females; C) males; and D) for HbA1C in the main sample.



2SLS = two-stage least squares. IVW = inverse-variance weighted.

Appendix 4.11. STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies

Item No.	Section	Checklist item	Relevant text from manuscript
1	**TITLE and ABSTRACT**	Indicate Mendelian randomization (MR) as the study's design in the title and/or the abstract if that is a main purpose of the study	"This chapter reports on the application of one such method, non-linear Mendelian Randomisation (MR), to the analysis of the alcohol–type 2 diabetes relationship."
	INTRODUCTION		
2	**Background**	Explain the scientific background and rationale for the reported study. What is the exposure? Is a potential causal relationship between exposure and outcome plausible? Justify why MR is a helpful method to address the study question	"it is unclear whether these findings reflect true causal relationships, or merely artefacts of confounding and reverse causation" "Only recently have formal methods for non-linear MR – capable of identifying non-linear relationships between an exposure and outcome – been developed and refined (Burgess et al., 2014; Staley & Burgess, 2017). Chapter 2 found no studies implementing these methods to examine the alcohol–T2D relationship (Visontay et al., 2022)"
3	**Objectives**	State specific objectives clearly, including pre-specified causal hypotheses (if any). State that MR is a method that, under specific assumptions, intends to estimate causal effects	"The present study performs the most robust causal analysis of the alcohol–T2D relationship to date. It aimed to apply formal non-linear MR analysis to determine the functional form of this relationship, contrast the relationship across the sexes, comprehensively assess sensitivity findings, and compare results to linear MR and observational evidence."

	METHODS		
4	**Study design and data sources**	Present key elements of the study design early in the article. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:	
		a) Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available.	The UK Biobank is a prospective cohort of over 500,000 participants aged 40-69 at baseline and recruited between 2006 and 2010 (Sudlow et al., 2015). Designed to identify risk factors for disease in middle and old age, it gathered comprehensive data from participants at baseline via questionnaire/nurse interview (probing SES, early life exposures, psychosocial and environmental factors, lifestyle, health status, and cognitive function), anthropometrics/physical testing, biochemical assays, and genotyping. Follow-up of the cohort via health record linkage is ongoing."
		b) Participants: Give the eligibility criteria, and the sources and methods of selection of participants. Report the sample size, and whether any power or sample size calculations were carried out prior to the main analysis	"For inclusion in MR and observational analyses, individuals of European ancestry with complete genetic, alcohol consumption, covariate, and outcome data were eligible. Related individuals (kinship coefficient >0.0884) were randomly removed (Hanscombe et al., 2019). Individuals with either baseline prevalent or incident type 1 diabetes (as ascertained via UK Biobank derived 'first occurrence' variables, or self-reporting insulin use at baseline), were removed for all analyses. This included those with a first diagnosis of T2D and subsequent type 1 diabetes diagnosis, on the assumption that this reflects an initial misdiagnosis. All those with baseline prevalent diabetes of any kind (excluding gestational) for whom T2D was *not* the first

		recorded kind of diabetes were also excluded. Where survival analyses were used (i.e., for the primary non-linear MR and observational analyses), those with baseline prevalent diabetes of any kind (except gestational) were additionally removed." "305,614 individuals were included in Cox non-linear MR, 310,448 for baseline prevalent T2D non-linear MR, 295,712 individuals for baseline HbA1C non-linear MR, and 233,568 individuals for observational Cox analyses."
c)	Describe measurement, quality control and selection of genetic variants	"SNP discovery (identification of variants significantly associated with drinks per week at a genome-wide level; $P < 5 \times 10^{-8}$) was made using results in those of European descent from GSCAN, a meta-analysis of GWAS studies (N=941,280)" "GSCAN identified 93 SNPs significantly associated with drinks per week at a genome-wide level in those of European descent. Several quality control measures were taken to improve confidence that the assumptions of an IV design, described above, would hold, excluding SNPs where necessary (see Appendix 4.1 for further detail). This resulted in 79 SNPs for the individual-level MR and 77 SNPs for the two-sample MR.
d)	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases	"An overall usual drinks per week variable was derived from a range of questions answered by participants at baseline about drinking/non-drinking status, frequency of consumption, and volume of beverage-specific consumption."

				"In 2022, the UK Biobank's derived 'first occurrence' variables were updated to thoroughly incorporate diabetic conditions. Utilising a combination of self-report, primary care, hospital inpatient, and death register data, they provide information for each participant on whether a particular condition was ever recorded, and if so, when first recorded and the source/s used to ascertain presence of that condition. These variables served as the basis for T2D outcome data in all analyses. Glycated haemoglobin (HbA1C), indicative of average blood glucose for the previous ~120 days (Burgess et al., 2021), was measured at baseline."
		e)	Provide details of ethics committee approval and participant informed consent, if relevant	N/A
5	**Assumptions**		Explicitly state the three core IV assumptions for the main analysis (relevance, independence and exclusion restriction) as well assumptions for any additional or sensitivity analysis	"IV designs rely on three assumptions (Burgess & Thompson, 2021): assumption 1 is that the IV is associated with the exposure ('relevance'), assumption 2 is that the IV is *not* associated with confounders of the exposure–outcome relationship ('exchangeability'), and assumption 3 is that the IV affects the outcome *only* via its effect on the exposure ('exclusion restriction')."
6	**Statistical methods: main analysis**		Describe statistical methods and statistics used	
		a)	Describe how quantitative variables were handled in the analyses (i.e., scale, units, model)	"Beverage-specific responses were converted to drinks per week using the following conversions: 1 UK standard drink = 8 grams of alcohol; drinks reported per month were considered as per 4.3 weeks (Topiwala et al., 2022); 1 pint of

				beer/cider=16g (2 drinks); 1 wine = 16.8g (2.1 drinks); 1 fortified wine = 14.8g (1.76 drinks); 1 shot of spirits=8g (1 drink); 1 other alcoholic drink = 12g (1.5 drinks; Biddinger et al., 2022)."
		b)	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected	"To produce GRSs for the one-sample analyses, dosages (number of effect alleles) for each of the 79 SNPs were weighted by the relevant beta coefficients identified in GSCAN (version without UK Biobank participants), and summed."
		c)	Describe the MR estimator (e.g. two-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples	"The ratio method was used to calculate LACEs within strata." "Non-linear MR analyses controlled for age, sex, genotyping array, and ancestry"
		d)	Explain how missing data were addressed	N/A
		e)	If applicable, indicate how multiple testing was addressed	N/A
7	**Assessment of assumptions**		Describe any methods or prior knowledge used to assess the assumptions or justify their validity	"Several quality control measures were taken to improve confidence that the assumptions of an IV design, described above, would hold, excluding SNPs where necessary (see Appendix 4.1 for further detail)."
8	**Sensitivity analyses and additional analyses**		Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations)	"Sensitivity analyses included removing former drinkers, removing former and never drinkers, and removing those with a T2D diagnosis ascertained via primary care records. Additionally, for incident T2D and HbA1c outcomes, a SNP-based approach to LACE calculation was performed using IVW, weighted median, weighted mode, MR PRESSO, and MR RAPS tests, to again account for the possibility of invalid IVs. As there are no existing R

			packages for implementing SNP-based methods in a non-linear MR framework, performing these analyses required that the doubly-ranked strata be manually created (adapting code from Tian et al., 2022), the SNP-based tests run separately within each stratum, and then code from *SUMnlmr* and *NLMR* packages adapted to meta-analyse results across strata."
9	**Software and pre-registration**		
	a)	Name statistical software and package(s), including version and settings used	"All analyses were conducted with R version 4.1.1"
	b)	State whether the study protocol and details were pre-registered (as well as when and where)	N/A
	RESULTS		
10	**Descriptive data**		
	a)	Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow diagram	"305,614 individuals were included in Cox non-linear MR, 310,448 for baseline prevalent T2D non-linear MR, 295,712 individuals for baseline HbA1C non-linear MR, and 233,568 individuals for observational Cox analyses"
	b)	Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (e.g. means, SDs, proportions)	Not reproducible as text.
	c)	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies	N/A
	d)	For two-sample MR: i. Provide justification of the similarity of the genetic variant-exposure associations between the exposure and outcome samples	"This was done to avoid 'winner's curse' bias caused by overlapping samples in the SNP discovery and effect estimates datasets in the

			ii. Provide information on the number of individuals who overlap between the exposure and outcome studies	one-sample analysis (Jiang et al., 2022), and sample overlap in the two-sample analysis."
11	**Main results**			
		a)	Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale	**Appendix 4.2.** SNPs included in genetic instruments.
		b)	Report MR estimates of the relationship between exposure and outcome, and the measures of uncertainty from the MR analysis, on an interpretable scale, such as odds ratio or relative risk per SD difference	N/A for a non-linear relationship.
		c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
		d)	Consider plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure)	Not reproducible as text.
12	**Assessment of assumptions**			
		a)	Report the assessment of the validity of the assumptions	Several quality control measures were taken to improve confidence that the assumptions of an IV design, described above, would hold, excluding SNPs where necessary (see Appendix 4.1 for further detail).
		b)	Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I^2, Q statistic or E-value)	N/A
13	**Sensitivity analyses and additional analyses**			

		a)	Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions	"Most *but not all* of the SNP-based approaches to non-linear MR produced concordant results, with the weighted median and weighted mode methods resulting in positive linear relationships in the overall sample and the protective effect in women disappearing/becoming non-significant (Appendices 4.8A and B)."
		b)	Report results from other sensitivity analyses or additional analyses	"Sensitivity analyses excluding former drinkers did not change the functional form for MR results, although it reduced the number of drinks per week associated with reduced risk of incident T2D and magnified the increased risk for heavy drinking (Appendix 4.5A). When excluding both former and never drinkers – i.e., limiting the sample to current drinkers – protective effects were no longer observed for the overall sample or for women, producing a positive linear relationship and no relationship respectively (Appendices 4.6A and B)."
		c)	Report any assessment of direction of causal relationship (e.g., bidirectional MR)	N/A (incident cases as main outcome)
		d)	When relevant, report and compare with estimates from non-MR analyses	"Survival analysis using observed alcohol consumption and incident T2D cases were consistent with a J-shaped relationship (Figure 5.6A), with the NDE and TMREL corresponding to a very large number of drinks per week."
		e)	Consider additional plots to visualize results (e.g., leave-one-out analyses)	Not reproducible as text.
	DISCUSSION			
14	**Key results**		Summarize key results with reference to study objectives	"Initial findings from non-linear MR indicated a J-shaped relationship between alcohol and T2D

			risk, with a small but significant protective effect up to roughly 8 drinks per week. However, sex-stratified analyses suggested this protective effect may only be present for women. SNP-based sensitivity analyses were not all consistent with the main analyses – in particular, weighted median and mode MR failed to produce significant protective effects in the main sample and for women specifically. Results for HbA1C, an index of average blood glucose levels, were robust across sensitivity tests, revealing a protective effect of alcohol consumption at all levels of genetically-predicted consumption that was larger for women than for men."
15	**Limitations**	Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and any efforts to address them	"Less-than-weekly drinkers in UK Biobank were only asked about usual volume consumed from 2008 onwards" "The other data limitation was that primary care linkage has only been achieved for just under 50% of the cohort" "There are also several limitations stemming from the nature of MR itself." "Finally, despite the precautions taken to remove SNPs with evidence of association with confounders of the alcohol–T2D relationship (see Appendix 4.1), it is not possible to guarantee the assumptions of exchangeability and exclusion restriction."
16	**Interpretation**		
	a)	Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies	"Genetic-based methods for enhancing causal inference do not rule out a possible protective effect of low-level alcohol consumption against

			T2D risk in women, and support a benefit of moderate drinking for average glucose levels in both sexes. Further MR studies in diverse cohorts, and triangulation amongst their findings, are needed for elucidation of these relationships."
		b) Mechanism: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable. Use causal language carefully, clarifying that IV estimates may provide causal effects only under certain assumptions	"has positive effects on biomarkers associated with glucose metabolism, e.g., postprandial glucose response (Brand-Miller et al., 2007) and oxidative stress (Ma et al., 2022), glycated haemoglobin (HbA1C; used also for the diagnosis of diabetes), and fasting insulin (Schrieks et al., 2015). There are also indications from interventional trials that alcohol may promote high-density lipoprotein cholesterol and increase adiponectin, which regulates insulin sensitivity, (Beulens et al., 2008; Brien et al., 2011) – both implicated in T2D."
		c) Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions	"A small beneficial effect of alcohol for T2D in women (if substantiated), as well as the demonstrated reductions in glucose levels for both sexes, is important information for policymakers and clinicians, but must be balanced against the known risks for a host of other conditions"
17	**Generalizability**	Discuss the generalizability of the study results (a) to other populations, (b) across other exposure periods/timings, and (c) across other levels of exposure	"Interestingly, the only previous non-linear MR analysis of alcohol and HbA1C found no evidence of either a linear or non-linear relationship (Peng et al., 2019) – this may be explained by either a genuinely different effect in Asian populations" "There are also several limitations stemming from the nature of MR itself. Firstly, the

			instruments used are associated with lifetime average drinks per week, so the findings offer no insight into the differential impacts of exposure at various ages, nor on the impact of drinking *patterns*."
	OTHER INFORMATION		
18	**Funding**	Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based	N/A
19	**Data and data sharing**	Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article, or report whether the code is publicly accessible and if so, where	N/A
20	**Conflicts of Interest**	All authors should declare all potential conflicts of interest	N/A