Methodological Developments in Constructing Causal Diagrams with Counterfactual Analysis of Adolescent Alcohol Harm

Karl D. Ferguson, BA (Hons)

Submitted in fulfilment of the requirements for the Degree of

Doctor of Philosophy

MRC/CSO Social and Public Health Sciences Unit

Institute of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

April 2020

Abstract

Background and aims: Causal diagrams, or Directed Acyclic Graphs (DAGs), are mathematically formulated networks of nodes (variables) and arrows which rigorously identify adjustment sets for statistical models. They are thus promising tools for improving statistical analysis in health and social sciences. However, a lack of pragmatic yet robust guidance for building DAGs has been identified as problematic for their use in applied research. This thesis aims to contribute an example of such guidance in the form of a novel research method, and to demonstrate this method's utility by applying it to observational data. **Design:** This thesis introduces 'Evidence Synthesis for Constructing Directed Acyclic Graphs' (ESC-DAGs) as a protocol for building DAGs from research evidence. It is demonstrated here in the context of parental influences on adolescent alcohol harm and the resulting DAGs are used to inform analysis of data from the Avon Longitudinal Study of Parents and Children (ALSPAC). **Methods**: ESC-DAGs integrates evidence synthesis principles with classic and modern perspectives on causal inference to produce complex DAGs in a systematic and transparent way. It was here applied to a subset of literature identified from a novel review of systematic reviews, which identified 12 parental influences on adolescent alcohol harm. ESC-DAGs was then further applied to the ALSPAC data to produce a 'data integrated DAG'. The outcome measure was the Alcohol Use Disorders Identification Test (AUDIT) administered to adolescent participants at age 16.5 years. Nine parental influences were measured, alongside 22 intermediates (variables lying on the causal pathway between parental influences and AUDIT score). The DAGs were then used to direct two stages of analysis: 1) weighting and regression techniques were used to estimate Average Causal Effects (ACEs) for each parental influence and intermediate; and 2) causal mediation analysis was used to decompose the effect of maternal drinking on adolescent AUDIT score to estimate Natural Indirect Effects (NIEs) for the intermediates and the other parental influences. **Findings:** Evidence for an ACE was found for each parental influence. Parental drinking, low parental monitoring, and parental permissiveness towards adolescent alcohol use had larger effects that were more robust to sensitivity analysis. Several peer and intrapersonal intermediates had higher effects. There was little evidence of an NIE of maternal drinking through other parental influences. There were substantial NIEs for substance-related behaviours of the adolescent and their peers. **Conclusions:** ESC-DAGs is a promising tool for using DAGs to improve statistical practices. The DAGs produced were transparent and able to direct various forms of data analysis in an immediate sense while differentiating between a comparatively large volume of confounders and other covariates. Future development is possible and should focus on efficiency, replicability, and integration with other methods, such as risk of bias tools. ESC-DAGs may thus prove a valuable platform for discussion in the DAG and wider quantitative research communities. The statistical analyses were performed with methods that were novel to the literature and findings triangulated with the wider evidence base. Mediation analysis provided novel evidence on how parental drinking influences adolescent alcohol harm via intrapersonal and peer pathways.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ACE	Average Causal Effect
ALSPAC	Avon Longitudinal Study of Parents and Children
ANOVA	Analysis of variance
AUD	Alcohol Use Disorder
AUDIT	Alcohol Use Disorder Identification Test
BMI	Body Mass Index
CDE	Controlled Direct Effect
CMO	Chief Medical Officer
CSO	Chief Scientist's Office
DAG	Directed Acyclic Graph
DALYs	Disability-Adjusted Life-Years
EIMOC	Exposure-Induced Mediator-Outcome Confounder
ESC-DAGs	Evidence Synthesis for Constructing Directed Acyclic Graphs
I-DAG	Integrated DAG
IG	Implied Graph
IMD	Index of Multiple Deprivation
IPW	Inverse Probability Weight
MICE	Multiple Imputation of Chained Equations
MRC	Medical Research Council
MRO	Methodological Research Objective
NDE	Natural Direct Effect
NIE	Natural Indirect Effects
POF	Potential Outcomes Framework
SRO	Substantive Research Objective
TCE	Total Causal Effect

Publications and Conference Presentations

The following publications and presentations have resulted from the research described in this thesis:

Publications

Ferguson, K. D., McCann, M., Katikireddi, S. V., Thomson, H., Green, M. J., Smith, D. J. & Lewsey, J. D. 2019. Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. *International Journal of Epidemiology*.

Conference presentations

Ferguson, K. D. 2017. A pilot protocol for graphical evidence synthesis: Parental effects on adolescent alcohol harm. 43rd Annual Alcohol Epidemiology Symposium of the Kettil Bruun Society. The University of Sheffield, Sheffield.

Ferguson, K. D. 2017. Using directed-acyclic graphs for evidence synthesis: Neighbourhood effects on adolescent alcohol harm. 5th UK Causal Inference Meeting. University of Essex, Colchester.

Ferguson, K. D., Lewsey, J., McCann, M. & Smith, D. 2018. How to build the 'right' directed acyclic graph (DAG): a systematic, transparent and accessible method for evidence synthesis. 62nd Annual Scientific Meeting of the Society for Social Medicine. University of Glasgow, Glasgow.

Acknowledgements

It has been my belief that even getting as far as submitting a PhD thesis may not be possible, for mere humans, without support from family, friends, and colleagues. The support I benefited from was rich and reliable, and I am forever grateful.

To Mum and Dad, who let me move back home so that I could go back to school when I finally accepted that my career as a rock star wasn't working out. Not everyone gets that, or the countless other amazing opportunities you've given me, and that you continue to give to me, Sasha, and Dani. I'm keenly aware of the level of tenacious self-sacrifice this has required from you both. It's why I'll never give you grandkids. Thank you also for taking the time to proof-read - Sasha and Dani too.

Mhairi, I feel dumb writing this sitting beside you. You were playing L.A. Noire on your Switch, I was stealthily typing this on my phone. You just ran over a pedestrian. I hope you didn't notice me struggling not to get mushy, or perhaps mistook it for concern over said pedestrian. I'm at an utter loss for how to express the depth of my love, gratitude, and admiration for you. So, I'll leave it at that.

Numerous researchers have shown me not just intellectual guidance during this process, but simple human kindness as well. The image of cut-throat academia did not pertain to my experience as a PhD student at SPHSU. I would like to thank Katy for taking a chance on me; Eric and Frank for encouraging me to be ambitious again; Vittal, Mike and Hilary for setting standards; Jon and Gemma for their hospitality; Mike and Peter for their advice and their time; the PhD students at SPHSU, especially Aidan; Breda for showing me it could be done; Mark, Elaine, Andy, Rory, Debs, and Gerry for the validation and future-proofing; Justine for her trust and guidance; and Laurence for his approachability. Of course, I especially want to thank my excellent supervisors, Jim, Mark, and Danny. Your mentorship and support saved me from more crises of self-doubt than I hope you realise. Additionally, I would like to thank the admin and IT staff at SPHSU for their patience, professionalism, and friendliness - especially Avril, Carol, and Crawford.

And finally, to Granny, who was so worried about the infamous Glasgow climate that she gave me two thermal blankest for when I moved, less than 100 miles, from Belfast. The blankets are horrifically ugly (jagged patterns, primary colours). I've kept one in work and one at home ever since I arrived. Sometimes they get used when it's not that cold. I wish you and Grandad were here to see me do this. But I'm comforted by the knowledge that many of the best parts of who I try to be are a reflection of you. I've tried to represent that in the following pages.

Author’s declaration

I declare that I am the sole author of this thesis, except where the contribution of others has been acknowledged. The work in this thesis has not been submitted in any form for another degree or professional qualification at the University of Glasgow or any other institution.

Karl D. Ferguson
August 2019

Chapter 1 Introduction

1.1 Background

When designing a statistical model in observational health research, a key consideration is which variables should be included to account for confounding bias. This is a challenging task which depends on several factors, including expert knowledge of the relationships under study and an understanding of which types of covariates should and should not be conditioned on given the research question. Take the example of estimating the effect of an exposure (e.g. poverty) on an outcome (e.g. alcohol harm). Common practices which are problematic include conditioning on too few confounders and misidentifying other types of covariates as confounders. In the former case, it is probable that confounding bias will not be accounted for adequately as the model will be similar to the unadjusted version. In the latter, other forms of bias may be induced, for example, by mistakenly controlling for a mediator (Westreich and Greenland, 2013). The end-result in both cases is that it is generally impossible to tell how much of the estimate (e.g. regression coefficient of poverty's effect on alcohol harm) is due to bias. Causal diagrams, hereafter termed as 'DAGs' (Directed Acyclic Graphs), are an increasingly popular method for selecting confounders (Pearl, 2018, Krieger and Davey Smith, 2016). However, while DAGs have great potential to reduce bias in statistical models by virtue of how their formal mechanics systematically identify confounders (Textor et al., 2011, Textor et al., 2016, Greenland et al., 1999), the practical challenge of how to actually build a DAG has not been met, thereby inhibiting this potential (Ferguson et al., 2019, Tennant et al., 2017, Sauer et al., 2013).

The key aim of this research was thus to develop a method which can guide users through the process of building DAGs. This is achieved in this thesis in an applied sense, by using the method to design DAGs pertaining to an important health context, and then to use those DAGs to direct analysis of observational data. Thus, while the analysis and findings on the health context are a crucial

component of this thesis, the primary emphasis was methodological. The method that was developed is called Evidence Synthesis for Constructing Directed Acyclic Graphs (ESC-DAGs), and was published in the *International Journal of Epidemiology* in July 2019 (Ferguson et al., 2019). The chosen health context was parental influences on adolescent alcohol harm. The rest of this chapter is concerned with justifying both the methodological impetus of designing a method for building DAGs, and the choice of health context. It closes by setting out the research objectives and describing how they meet the overall aim of developing and applying ESC-DAGs.

1.1.1 DAGs

DAGs are graphs consisting of nodes and arrows, in which nodes represent variables and arrows represent causal relationships between them. For example, Figure 1-1 below presents a basic DAG, showing an outcome (liver cirrhosis) a distal cause (age) and a more proximal cause (alcohol consumption). See Chapter 2 for a detailed discussion on DAG mechanics.

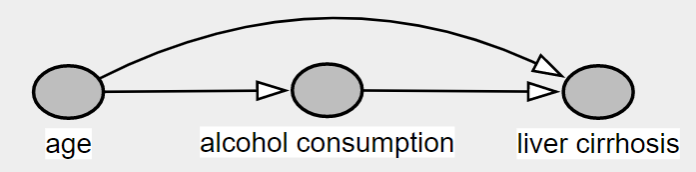


Figure 1-1: A basic DAG

DAGs were originally developed by Pearl in Computer Science (Morgan and Winship, 2015, Pearl, 2009), and were introduced to health and social sciences by Greenland et al in 1999 (Greenland et al., 1999). The authors (which included Pearl) suggested that DAGs are a valuable tool for informing statistical models because they provide a "compact graphical as well as algebraic formulation of assumptions". This value derives from how the "algebraic formulation" provides a systematic and reliable way to identify sources of confounding bias when designing statistical models. For example, the DAG in Figure 1-1 states that age confounds the relationships between alcohol consumption and liver cirrhosis, and so a statistical model estimating the effect of alcohol consumption and liver cirrhosis should adjust or otherwise control for the confounding bias of age. DAGs have gained considerable popularity since 1999. Figure 1-2 below depicts their increasing volume of citations (Web of Science search on 22/07/2019: TS = "causal diagram" OR "directed acyclic graph"). Further, updated editions of key reference texts in health and social sciences increasingly have entire chapters dedicated to DAGs. Such texts include *Modern Epidemiology* (Rothman et al., 2008), *Methods in Social Epidemiology* (Oakes and Kaufman, 2017), and *Structural Equation Modelling* (Kline, 2015). Thus, as per Morgan and Winship, DAGs are recognised as "powerful" tools for

directing statistical analysis, and this recognition drives their popularity in both pedagogical and applied settings (Morgan and Winship, 2015).

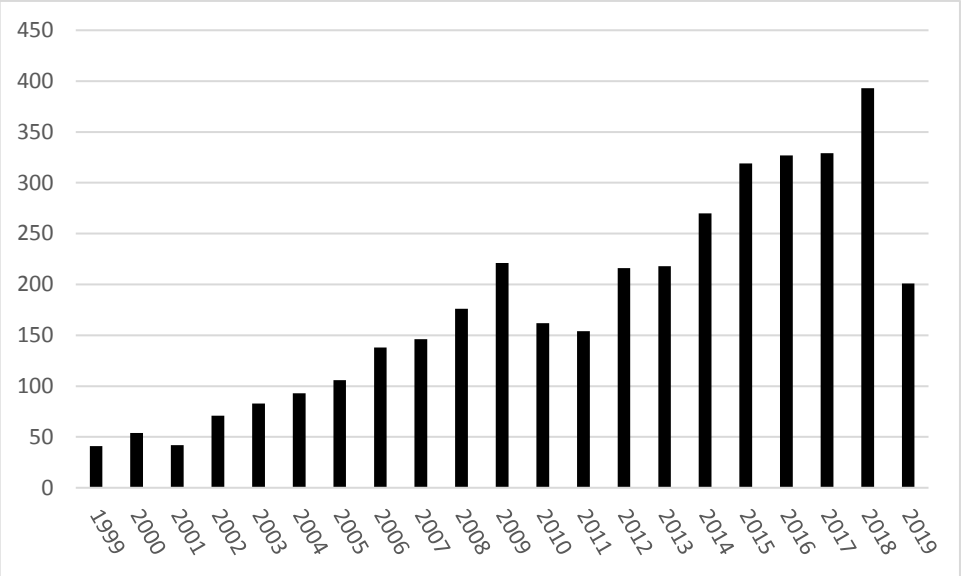


Figure 1-2: Increasing citations of DAGs in Web of Science, 1999-2019

However Greenland et al's seminal paper does not mention how DAGs should be built at all, focusing instead on how DAGs work (Greenland et al., 1999). The literature on DAGs has continued with this trend for 20 years. The limited amount of guidance on building DAGs is restricted to relying on "background knowledge" (Sauer et al., 2013, Robins and Wasserman, 1999), "prior knowledge" (Lederer et al., 2019), or simply "knowledge" (VanderWeele, 2019). Indeed, in studies which have used a DAG, the common practice is to state or imply that the DAG was built by relying on the knowledge of the research team, sometimes with reference to consultation with external experts (Cullen, 2018, Polzer et al., 2012, Ahrens and Schisterman, 2013). Another common approach is to simply list the selected confounders without providing any insight into how or why they were selected (Liljestrom et al., 2018, Zampieri et al., 2019). While pragmatic, these strategies raise questions of reproducibility and reliability. In other words, while there is broad recognition that scientific knowledge must drive how DAGs are built, there has been no progress towards a method for doing so that is systematic or open to close scrutiny.

This has been identified as a problem at least as early as 2013, when Sauer et al called for "a disciplined approach to developing DAGs" (Sauer et al., 2013). More recently, Tennant et al found that a lack of "guidelines for best practice" may indeed be consequential (Tennant et al., 2017). In a critical review on the use of the popular DAG-graphing software 'DAGitty', they found that: DAGs are often over-simplified to an implausible degree; are altered to fit model specification rather than the opposite (for example, not including unmeasured confounders); and are regularly not included in

research papers which state that DAGs have been used. As such, while DAGs are recognised as powerful tools for reducing bias in statistical studies, Tennant et al indicated that a "huge variation in practice" may be inhibiting DAGs from meeting that potential.

The foremost contribution of this thesis is the ESC-DAGs methodology, designed to begin addressing this gap on research methods for building DAGs. It was designed by the author with oversight from PhD supervisors and other academics between 2016 and 2018 and uses various principles from the evidence synthesis and DAG literatures. As well as the publication it was presented at several academic conferences, including the 43rd Kettil Bruun Alcohol Symposium in 2017 (The University of Sheffield, presentation), the UK Causal Inference Meeting in 2017 (The University of Essex, poster), and the Society for Social Medicine Annual Scientific Meeting in 2018 (The University of Glasgow, poster (Ferguson et al., 2018)). It is described in detail in this thesis in Chapter 4.

1.1.1.1 DAGs and causal inference

DAGs are often referred to as a causal inference method. Indeed, an arrow in a DAG does not represent a quantitative association - it is a qualitative and non-parametric causal hypothesis. There are two fundamental reasons why this emphasis on causation makes DAGs useful for causal inference. First, when researcher assumptions are explicated using DAG rules, the resulting diagram can be used to help satisfy important assumptions necessary for valid causal inference, especially 'conditional exchangeability' as discussed in Chapter 3. Second, because DAGs encode causal pathways, they also encode indirect effects (i.e. mediators). Mediation is one way in which to further evidence an exposure as a cause, simply by virtue of identifying some aspect of the causal mechanism (again, see Chapter 3). Furthermore, DAGs are commonly used in tandem with statistical methods from the 'Potential Outcomes Framework' (hereafter; the 'POF') which are also commonly referred to as causal inference methods. For example, the POF methods in this thesis estimate 'average causal effects' before decomposing them using 'causal mediation analysis'. However, neither DAGs nor POF methods are causal inference methods in the truest sense. Firstly, DAGs are tools to help build statistical models – they do not objectively identify causal relationships. Secondly, POF statistical methods are subject to the same kinds of strict and often untestable assumptions that underpin most statistical methods. Nevertheless, this thesis maintains an emphasis on causation throughout. This is driven by both a methodological interest and an underlying epistemological stance. Simply put, that stance is that, rather than "retreating into the associational haven" this research aims to "take the causal bull by the horns" (Hernán, 2005). This is not achieved by using specific methods which are a silver bullet for causal inference – no such methods exist for analysis of observational data. Rather, it is achieved by interpreting all findings in the context of formal assumptions for valid causal inference. In other words, the scope for causal inference, or lack thereof, is transparently considered throughout

this thesis. Indeed, this is what is meant when DAGs and POF statistical techniques are referred to as 'causal inference methods'.

1.2 Chosen health context

Alcohol harm

Alcohol is an important social determinant of health. It is addictive, psychoactive, and toxic, and is associated with over 200 health outcomes, varying from acute alcohol poisoning to accidental injury to chronic liver scarring (World Health Organisation, 2014). It has also been shown to have a dose-response relationship with several forms of cancer (Bagnardi et al., 2014). This translates into notable contributions to death and disease. Globally in 2016, 2.8 million deaths were estimated to be attributable to alcohol (Griswold et al., 2018). This accounts for 2.2% of age-standardised female deaths and 6.8% of age-standardised male deaths. Alcohol is also estimated to account for 1.6% (female) and 6.0% (male) of disability-adjusted life-years (DALYs). A more recent estimate for Scotland specifically is around 8% for females and males (Tod et al., 2018). Alcohol consumption thus leads to various diseases (e.g. cancer), conditions (e.g. addiction), and events (e.g. risky sexual behaviour, injurious accidents, death etc.) that can collectively be understood as alcohol harm, and that are measured in various ways (e.g. DALYs). In Scotland and the rest of the UK, like in most countries, alcohol is nonetheless commercially available to adults. It plays an important economic role and narratives describing its cultural significance are ubiquitous. Unsurprisingly perhaps, this apparent contradiction between its clear risk to health and wellbeing and the perception of its economic and cultural importance has resulted in alcohol consumption being a well-studied behaviour in health and social sciences, as both a risk factor and an outcome. It also attracts a notable degree of media and policy attention. Minimum Unit Pricing in Scotland (Angus et al., 2016, Holmes et al., 2018, Katikireddi et al., 2014) and the ongoing deliberations on a Welsh equivalent are recent high-profile examples. Another was the review of guidelines for low risk drinking by the UK Chief Medical Officer (CMO) in 2016, which argued that the risk of negative health outcomes due to alcohol consumption increases with any amount of alcohol consumed on a regular basis, and lowered the recommended low risk threshold for men from 21 to 14 units of alcohol per week (Department of Health, 2016, Holmes et al., 2016).

Adolescent alcohol harm

The legal age of sale for alcohol is 18 years and above for people in the UK. This is similar to many other Western European states, with some countries having more prohibitive legislation, such as the United States' restriction to age 21 and older. Despite this, 'underage drinking' (i.e. alcohol

consumption by those aged below the national legal age of sale) is recognised as a common problem globally. For example, the World Health Organisation's global strategy to reduce alcohol harm has a notable focus on underage drinking (World Health Organization, 2010). However, in the UK and many other high-income counties, underage drinking has seen a notable decline in the last decade (Oldham et al., 2018). Nevertheless, it is arguably still highly problematic. For example, there is strong evidence that underage drinking increases the probability of later substance dependence in general and of greater alcohol use over the life course, and that the psychoactive aspects of alcohol may have implications for neurological development during adolescence (Hall et al., 2016). As such, with the prevalence of underage drinking indicating that adolescence is the life stage during which alcohol consumption typically begins (Degenhardt et al., 2016), adolescence is thus an "opportune time" for preventing the formation of harmful drinking habits and of adolescent alcohol harm (Yap et al., 2017).

Parental influences

Official guidelines in the UK, established by the CMO in 2009, are consistent with this evidence. They recommend that, because an "alcohol-free childhood is the healthiest and best option", children should not drink alcohol until at least age 15 (Donaldson, 2009). They further recommend that any alcohol consumption between the age of 15 and 17 be supervised by a parent or carer (hereafter referred to as parents) and that this take place in a supervised environment. Further, the report in general heavily emphasises the importance of parents in determining the alcohol-related behaviours of their children during adolescence, including the role of parents' own drinking behaviours. Of course, given the central role of parents in the socialisation of children, the role of parents in determining adolescent alcohol behaviour is well recognised and has been studied extensively, if not exhaustively (Yap et al., 2017, Mattick et al., 2018).

However, parental influences are a broad conceptual domain, and cover parental behaviours, parenting style (hereafter; parenting), and aspects of the parent-child relationship (see Chapter 5). Thus, in a methodological sense, this makes parental influences an interesting candidate for the exposure(s) in this thesis. This is because building a DAG involves differentiating between confounders and intermediates, rather than simply identifying confounders. Thus, DAGs developed from the ESC-DAGs method hypothesise causal pathways *between* exposures, and so identify possible indirect effects. As alluded to above, indirect effects are an important consideration for causal analyses as they can shed light on the 'black box' of causation between the purported cause and the outcome by focusing on the causal mechanism itself. Thus, applying ESC-DAGs to a 'domain' of exposures which are related, but which are relatively distinct, allows for a more holistic examination which is consistent with causal analysis. This has important implications for identifying and analysing prevailing theories in the relevant literature, such as the common theory that parental

alcohol problems inhibit good parenting practice which in turn increase offspring alcohol harm (Mahedy et al., 2018). However, differentiating between confounders and intermediates for a broad domain of exposures also results in a high volume of intermediates which are not from that domain. For example, the indirect effect of parental discipline on adolescent alcohol harm through the adolescent's mental health. Of course, this immediately broadens the scope of the DAGs and subsequent analysis. As such, a pragmatic conceptual framework used to assist in dealing with this complexity is introduced below. Thus, the breadth of parental influences as a group of exposures presents opportunities for unpacking casual relationships, but also engenders logistical challenges via greater complexity, especially of the DAGs.

Summary

Parental influences on adolescent alcohol harm are argued here as a valid health context for application of ESC-DAGs and subsequent DAG-based analysis. Adolescent alcohol harm is widely recognised as an important health outcome, largely because it plays a role in determining later behaviour and because alcohol use is arguably always harmful. Parental influences are an important domain of exposures for adolescent behaviour in general, and underage drinking is no exception. Further, the breadth of parental influences will allow for analysis of causal mechanisms (in the form of mediation analysis of indirect effects). However, there were further recommending factors. Firstly, relevant data that was of sufficient quality was available in the form of the Avon Longitudinal Study of Parents and Children (ALSPAC). It has self-report data from both parents and the adolescent, and it measures adolescent alcohol harm at age 16.5 years using an Alcohol Use Disorders Identification Test (AUDIT), which is a validated quantitative tool (Saunders et al., 1993). ALSPAC is discussed further in Chapter 7. Another aspect more specific to the DAGs and POF statistical techniques especially, is that there is little-to-no evidence on the parental influences on adolescent alcohol harm using these methods (see Chapter 5). As such, their application to the topic would be worthwhile whether or not findings were in line with the literature (i.e. triangulation) or in disagreement. In summary, adolescent alcohol harm is an important health outcome for which parental influences are important and interesting determinants; there is data available on both; and the methodological contributions of this thesis will make a substantive contribution to the literature.

1.2.1 A pragmatic conceptual framework

The socio-ecological framework was used to systematise the high volume of concepts and variables that were engendered by the broad focus on parental influences. Figure 1-3 below is a simple reconceptualisation of Bronfenbrenner's original socio-ecological framework of adolescent development, first introduced in 1979 (Bronfenbrenner, 1979).

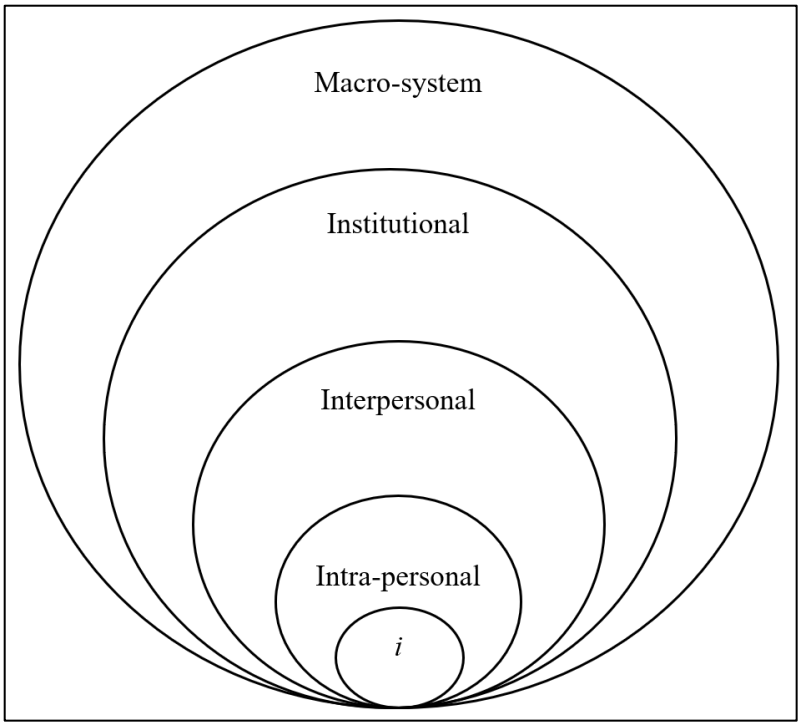


Figure 1-3: Simplified socio-ecological model

It posits that the individual adolescent *i* is influenced by their own intra-personal factors (e.g. personality, mental health, behaviour, attitudes), interpersonal relationships (e.g. with parents, peers, siblings, teachers, etc.), institutional effects (e.g. of school, neighbourhood, work, or community), and the wider macro-system (e.g. policy and macro-economics). It further suggests that each of these 'layers' is influenced by the others, either directly (e.g. institutional on interpersonal) or indirectly (e.g. macro-system on interpersonal via institutional). However, while this model certainly posits an overall theory of adolescent development, its main operation within this thesis is to allow for important concepts pertaining to the adolescent to be differentiated from one another. More specifically, the socio-ecological model as it is applied in this thesis, categorises influences on adolescent alcohol harm as intra-personal, interpersonal or institutional. Parental influences, while arguably taking place in the context of the family institution, are treated as interpersonal influences, although this was mainly to emphasise parent-child interactions over time, rather than for practical reasons (such as limiting the number of variables in any layer of the socio-ecological model, for example). Thus, as alluded to above, this conceptual framework is employed out of pragmatic need rather than as an overarching theoretical stance. Note that the macro-system is not used in this thesis, as there is no way to operationalise it in the ALSPAC data. Further, the causal methods that usually pertain to analysis of the macro-system would have represented a distinct addition to an already methodologically complex thesis (i.e. natural experiments/regression discontinuity).

1.3 Research objectives

This research has a split emphasis between methodological innovation and substantive interest in investigating parental influences of adolescent alcohol harm. The research objectives were divided into halves to reflect this. There were ten research objectives in total. The first five were 'methodological research objectives' and focused on the development and demonstration of ESC-DAGs. The latter five were 'substantive research objectives' and focus on the literature and analysis of parental influences on adolescent alcohol harm. The methodological and substantive research objectives are abbreviated to MROs and SROs throughout. Each is described below with signposting to the most relevant chapters.

1.3.1 Methodological research objectives

The MROs are concerned with defining ESC-DAGs, demonstrating how it produces DAGs, and using these DAGs to direct analysis of observational data.

MRO 1 - Establish a foundation in causal methods for this thesis

A firm foundation in causal inference methods was required in order to perform this research. This was not a small undertaking as causal inference methods are diverse despite existing in a unified framework (Pearl et al., 2016, Daniel et al., 2016). As counterfactual reasoning underpins the POF, the approach taken in this thesis was to first introduce how counterfactual reasoning can be used to infer causation in general and then in research settings. Then DAG mechanics are considered in some detail, often in reference to counterfactual reasoning. Finally, specific statistical techniques for estimating potential outcomes (directed by DAGs and using counterfactual scenarios) are introduced. The selected statistical techniques focus on estimating the 'average causal effect' of an exposure on an outcome and then on estimating the indirect of an exposure on an outcome through mediators. The latter is performed using so-called causal mediation analysis. MRO 1 is met across Chapters 2 and 3.

MRO 2 - Develop a method for building DAGs

As argued above, there is a clear and important need for methods for building DAGs, and the ESC-DAGs methodology of this thesis is one possible solution. This research objective is met in Chapter 4, which describes the ESC-DAGs protocol in detail, first explaining its underlying principles before demonstrating its mechanics, and then explaining its application in this thesis.

MRO 3 - Demonstrate how the method produces DAGs

As ESC-DAGs is a novel and involved methodology, the core of this thesis revolves around a detailed demonstration of how it was applied (to literature and data pertaining to parental influences on adolescent alcohol harm). ESC-DAGs is technically a form of literature review, and so this application is framed as an 'ESC-DAGs review'. It takes place over three chapters. Chapter 5 identifies the important concepts from the literature, Chapter 6 integrates these concepts into a provisional DAG, and Chapter 7 operationalises this DAG in the context of the ALSPAC data. Despite the degree of novelty of this approach, it is very similar to any research project which gives careful treatment to the implications of the pertaining literature and the data used.

MRO 4 - Demonstrate how the method can be used to direct conventional data analysis

An underlying aim of this thesis, which was a key impetus in its early design, was to be a translational study which brings POF methods and DAGs closer to 'mainstream' health and social science. While DAGs are automatically associated with POF methods, they can also be used to improve modelling practices with other statistical techniques. As such, the fourth MRO was to demonstrate that DAGs can be beneficial outside the POF. A further justification was that any findings from 'conventional' analysis can be usefully contrasted to findings from POF analyses. To achieve this objective, Chapter 8 uses the DAGs produced from ESC-DAGs to direct multiple regression models.

MRO 5 - Use the method to direct data analysis under the potential outcomes framework

As DAGs are most commonly used in tandem with a specific (if broad) group of statistical methods, it was necessary to demonstrate that the DAGs from ESC-DAGs could perform the same role. Accordingly, Chapter 8 and Chapter 9 employ POF statistical methods to estimate the effects of various parental influences (and their intermediates identified from ESC-DAGs) on adolescent alcohol harm, and then to analyse indirect effects.

1.3.2 Substantive research objectives

The substantive research objectives are concerned with understanding parental influences on adolescent alcohol harm. Specifically, the contemporary literature is appraised and used to develop DAGs that are then used to analyse parental influence as causes of adolescent alcohol harm.

SRO 1 - Determine the main parental influences on adolescent alcohol harm from the literature

This research objective corresponds to a literature review, with the specific purpose of identifying the parental influences, and thus causal hypotheses, that will shape the DAGs. The method identified for this task was a review of systematic reviews, largely because it strikes the necessary balance between efficiency and comprehensiveness required to support a subsequent ESC-DAGs review. It is the focus of Chapter 5.

SRO 2 - Appraise the evidence on parental influences on adolescent alcohol harm

SRO 2 also pertains to the review of systematic reviews but is more concerned with critical appraisal of the literature. This appraisal is narrative and emphasises methodological gaps in evidence (e.g. mediation analysis); general scope for causal inference; and, as a review of systematic reviews, the methodological quality of the meta-literature. As well as informing the hypotheses in the DAGs, this sets a foundation which can be returned to when interpreting findings. Accordingly, because DAGs are non-parametric, this research objective pertains to their interpretation rather than their design.

SRO 3 – Build DAGs of parental influences on adolescent alcohol harm

This research objective is concerned with the application of ESC-DAGs to the literature on parental influences on adolescent alcohol harm and the ALSPAC data. It is thus highly related to MRO 3 (apply and demonstrate the method) and as such is also met across Chapters 6 and 7.

SRO 4 - Test the causal effects of parental influences and their intermediates on adolescent alcohol harm

This research objective is concerned with testing the degree to which the parental influences identified from ESC-DAGs 'cause' adolescent alcohol harm. The analytical strategy was somewhat novel compared to the pertaining substantive literature, and involved using the DAGs to direct analysis of 'average causal effects'. The method used derives from the propensity score literature, and is a combination of regression analysis and 'inverse probability weighting'. It is explained in detail in Chapter 3, and the results are documented in Chapter 8. The exposures used are the various parental influences and their intermediates as identified from the DAGs.

SRO 5 - Investigate mediators of parental influences on adolescent alcohol harm

One of the strongest benefits of DAGs in general is how useful they are for differentiating between direct and indirect effects (see Chapter 3). Indeed, as noted above, indirect effects play an important role in understanding causation. They also have important implications for

intervention design in Public Health, especially in instances when a less modifiable exposure's effect on an outcome is conveyed by a more modifiable mediator (e.g. the indirect effect of ethnicity on cholesterol via diet). As such, mediation analysis of parental influences was determined as the final research objective for this thesis. Mediational g-computation was used to estimate randomised interventional analogues of the natural indirect effect. This technique is described in Chapter 3, and results are documented in Chapter 9. Chapter 3 also introduces an 'analysis plan' which acts as a bridge between the research objectives as a whole and the statistical models.

1.3.3 Summary of thesis structure

To summarise, Chapters 2 and 3 focus on establishing a foundation in causal inference methods, and do so by examining counterfactual reasoning, DAGs, and POF statistical methods. Chapter 4 then builds on the DAG foundation specifically and introduces ESC-DAGs in some detail, with an emphasis on its application in this thesis. Before demonstrating this application, Chapter 5 outlines the review of systematic reviews that was undertaken to appraise the pertaining literature and identifies important parental influences on adolescent alcohol harm. Chapter 6 then applies the ESC-DAGs protocol to this literature to develop a provisional DAG. Chapter 7 then introduces the ALSPAC data, provides basic descriptive analysis of the key variables, and then proceeds to contextualise the DAG from Chapter 6 in the data. Chapter 8 and 9 then use the DAGs produced from ESC-DAGs and the POF methods described in Chapter 3 to analyse the ALSPAC data. Chapter 8 uses a combination of regression and weighting-based methods to estimate 'average causal effects' of parental influences on adolescent alcohol harm. The intermediates identified from the DAGs are given the same treatment. Chapter 9 then decomposes causal effects across various intermediates using causal mediation analysis. Finally, Chapter 10 is the Discussion for this thesis. It evaluates the ESC-DAGs application, considers further developments for the method, explores the implications of the findings, as well as opportunities for future research.

Chapter 2 Causal inference: Counterfactuals and DAGs

This chapter pertains to MRO 1 (establish a foundation in causal methods for this thesis). This foundation supports the applied POF analyses and novel methods in subsequent chapters, and thus pertains to all other research objectives. The emphasis of this chapter is split between introducing counterfactual reasoning as a basis for POF statistical methods, and the DAGs mechanics which underpin the ESC-DAGs method and the later DAG-based analyses. Accordingly, this chapter features one section on counterfactual reasoning and a second on DAG mechanics.

2.1 Counterfactual reasoning

This section follows the style set out in key texts (Pearl, 2009, Hernán and Robins, 2019, Pearl et al., 2016) by first considering counterfactual reasoning in non-scientific and informal settings before looking at treatment-outcome relationships.

The 2018 film *Avengers: Infinity War* ended when the antagonist, Thanos, defeated the titular superhero team, the Avengers. He used 'the snap' to successfully complete his plan to wipe out half of all life in the universe. In the 2019 sequel, *Avengers: Endgame*, the remaining Avengers struggle to come to terms with their defeat. Early on in the film, Scott asks the question "what if we could go back in time and stop Thanos?". The 'what if' clause is the key aspect of a counterfactual and the act of travelling back in time in Science Fiction to stop the emergence of an undesired set of circumstances in the present is a notably common formulation of a counterfactual. Implicitly; the factual scenario in the past and its observed outcome in the present are contrasted with a

counterfactual scenario in the past and its corresponding *potential outcome* in an alternative present. Scott identifies one 'factual' scenario in which Thanos performs the snap, and one counterfactual scenario in which Thanos does not perform the snap. In the factual snap scenario, the observed outcome is that half of all life is wiped out. In the non-snap counterfactual scenario, the assumed potential outcome is that no one dies.

One implication of this is that counterfactual reasoning is such a common factor in daily life that it can act as the vehicle for a convoluted plot involving numerous alternative realities. Indeed, counterfactual reasoning will be very familiar to most people, especially for anyone who has ever bemoaned a decision and its outcome. For example, consider the counterfactual question 'I am late for an appointment - what if I had cycled instead of taking public transport?'. The counterfactual scenario of cycling is implied here to produce the potential outcome of not being late, compared to public transport having already been observed to result in tardiness. In the POF, it is the contrast between the potential outcomes of counterfactual scenarios that is of interest, although the general logic of counterfactual reasoning can be applied to many Public Health questions. For example, consider John Snow's research and intervention on the Broad Street pump. There was the factual scenario of 'the pump handle was removed and thus could not be used, thereby stopping locals from accessing the choleric water supply' and the counterfactual scenario of 'the handle was *not* removed and thus the pump was continued to be used, thereby allowing locals to access the choleric water supply'. In the observed outcome, the mortality rate dropped off, while in the counterfactual potential outcome, it would be safe to assume that the mortality rate would not have done so due to continued exposure to cholera-infected water (Davey Smith, 2002).

Of course, any observational example invites scrutiny regarding the complex of other circumstances involved. For this reason, experimental research can be a fruitful context for exploring counterfactuals. Take a fictional research setting where researchers are interested in testing the viability of a new type of ethanol that is designed to allow for intoxication without a hangover. They design a controlled trial without randomisation. The sample is the Avengers. The new form of ethanol is given to half of the participants in the form of 8 pints of Pale Ale X at 5% alcohol content, which they must consume within 4 hours. The control group is given 8 pints of a popular commercial alternative, Pale Ale A, also at 5%, and also to be consumed within 4 hours. The next morning each Avenger is asked how hungover they feel on a scale of 1-10. Steve is in the treatment group and drinks 8 pints of Pale Ale X in 4 hours. The next day he is not hungover. However, we cannot infer that the ethanol in Pale Ale X successfully prevents hangovers as we do not know what Steve's potential outcome would have been if he, counterfactually, had consumed 8 pints of Pale Ale A in 4 hours instead of Pale Ale X. Or in other words, we do not what *would have happened if* he had drunk Pale Ale A instead of Pale Ale X. After a seminal paper in 1986 by Holland, this is known as the

'fundamental problem of causal inference'. For the same individual, "it is impossible to observe" all potential outcomes under mutually exclusive treatments. Only the factual treatment and the factual outcome may be observed and thus the effect of treatment on the outcome itself can never be observed in an individual.

What if we were to compare individuals from the treatment and the control arms of the study? Unlike Steve, Scott was in the control group and was hungover the next day. This supports the conclusion that Pale Ale X prevents hangovers – the control lead to one hangover while the treatment lead to one non-hangover. However, we know from a scene the 2011 film *Captain America: The First Avenger*, where Steve attempts to self-medicate with alcohol after a trauma, that his particular superpowers prevent him from being either intoxicated or hungover. Scott, however, has no innate superpowers. With this knowledge, we can assume that Steve's outcome would have been the same regardless of which Pale Ale he consumed and thus he is not a valid comparator for Scott. In other words, as both participants have different confounders, their observed outcomes cannot comment on the other's unobserved (potential) outcome. Thus, simply comparing a pair of individuals who have been given different treatments does not, of course, overcome the fundamental problem of causal inference. However, if two participants are *identical* in every way except for which Pale Ale they consumed, then the observed outcome for one would indeed be equal to the potential outcome for the other. This would allow for counterfactual causal inference by contrasting these potential outcomes. However, the extent to which a comparison between two identical individuals solves the fundamental problem of causal inference is limited - we would not be able to generalise to individuals that would have made poor comparators (e.g. along dimensions of sex, ethnicity, social class, etc.).

The researchers run the study again, this time on a random sample of 1000 Belfast inhabitants between the ages 21 and 25 years old. The sample is then randomised such that 500 participants are in the treatment arm and 500 are in the control arm. The study is now a true Randomised Controlled Trial (RCT). However, similar to the above, we cannot infer what the effect of the treatment is from any individual in either arm, and despite randomisation, we should not generally attempt to infer by comparing an individual from one arm to an individual in the other. However, due to the random sampling and the random treatment assignment, we can now assume that the study arms are comparable *on average*. With that being the case, the researchers can now ask the counterfactual question; 'what would have happened on average to the treated had they been in the control arm instead?' Thus, under various assumptions including adherence to protocols of the RCT, the average outcome in the control arm is equal to the potential outcome for the treatment arm (assumptions for valid causal inference are discussed in some detail in Chapter 3). In other words, an RCT is able to overcome the fundamental problem of causal inference by using randomisation to compare the

average potential outcomes in comparable groups. If the average in the treatment arm is sufficiently low compared to that in the control arm, researchers could infer that Pale Ale X reduces hangovers.

In summary, counterfactual reasoning is common in humans and prevalent in daily life, art, and science. Counterfactuals revolve around 'what if' questions, in which we consider what might have happened to an outcome, if the past had been different. However, counterfactual reasoning cannot overcome the fundamental problem of causal inference as we cannot observe potential outcomes under mutually exclusive treatments or exposures in the same individual. Nonetheless, if we are able to compare the average observed outcome in one group to the average observed outcome in another group, and if both groups are large and only differentiated from one another in terms of their exposure, then, because the observed outcome in either group is equal to the potential outcome for the other group, we can contrast potential outcomes and thus perform counterfactual causal inference. This can be achieved in RCTs and also in observational data (Chapter 3).

Thus far however, this chapter on causal inference has failed to define what a cause actually is. This a vast and dynamic topic in its own right - in his review of the second edition of Morgan and Winship's key text, *Counterfactuals and Causal Inference*, King wrote in 2015 that "more has been learned about causal inference in the last few decades" than in all previous endeavour on the topic (Morgan and Winship, 2015, Pearl et al., 2016). Perhaps surprisingly, this is in the context of millennia of fierce philosophical debate starting at least with Aristotle. However, as this thesis focuses on the POF, it derives a definition of cause from Pearl, who stated that:

> "A variable X is a cause of a variable Y if Y in any way relies on X for its value"
>
> (Pearl et al., 2016)

This definition is simple and effective, but lacks specificity, especially regarding the complex nature of health outcomes such as those investigated in this thesis. Health outcomes commonly depend on 'component causation' (Glass et al., 2013, Rothman and Greenland, 2005a, Kundi, 2006). For example, there may be numerous component causes involved in determining problematic alcohol use, such as mental health, personality, availability of alcohol, cultural importance of alcohol, etc. Further, health outcomes are often 'multi-causal' in that the causal mechanism for an outcome in one setting (e.g. an individual or a country) may involve different causal components compared to another setting. For example, one person may develop liver cirrhosis from problematic alcohol abuse and a high fat diet, while another may develop liver cirrhosis from untreated hepatitis and a high fat diet in the complete absence of alcohol use. Cause and effect in general must also obey temporality – cause must precede effect. Thus, causal components are 'ordered' such that distal causes generate proximal causes which together determine the disease. For example, child abuse may result in poor mental

health which can result in problematic alcohol use. It is thus possible to further refine the working definition of cause. Again, to paraphrase Pearl:

> "…a variable X is a cause of a variable Y if Y in any way relies on X for its value… and… X is a cause of Y if it is a direct cause of Y, or of any cause of Y" (Pearl et al., 2016)

Rather than attempting to establish a comprehensive or authoritative definition of cause, a secondary aim for this chapter is to establish a working definition that is sufficient to facilitate the use and explanation of the causal inference methods used throughout this thesis. As the above definition accounts for multi-causality, component-causation, and temporality, and because it is compatible with counterfactual reasoning and DAGs (as below) it is sufficient for the purposes of this chapter. However, as noted above it is expanded upon again in Chapter 3's discussion on mediation.

2.2 DAGs

As alluded to in Chapter 1, DAGs were developed by Pearl in Computer Science because:

> "in order to deal rigorously with questions of causality, we must have a way of formally setting down our assumptions about the causal story behind a data set" (Pearl et al., 2016)

DAGs 'set down' our assumptions in a systematic and transparent way according to a robust ruleset. They are then subjected to a mathematical algorithm called d-separation which allows researchers to identify an appropriate adjustment strategy for estimating causal effects. This section first describes how the fundamental components of DAGs and the key 'causal structures' they convey represent the causal assumptions of researchers. It then outlines how d-separation considers these assumptions when determining statistical adjustment strategies.

Before continuing it should be noted that, while DAGs are commonly used in POF analysis, they are not technically part of the POF. They are tools, whereas the POF is as much an epistemology with roots in counterfactual philosophy, as it is an epidemiological set of methods (Blakely et al., 2016). In fact, DAGs are merely a special class of a broader group of conceptual models sometimes referred to as 'structural causal models' (Pearl et al., 2016). What makes DAGs unique, and so promising to health and social science, are the practical benefits of d-separation in determining (reproducible) adjustment strategies.

2.2.1 DAGs: Basic components

DAGs represent a cause as an arrow between two nodes. A node may represent a concept or a variable (measured or unmeasured). Arrows are referred to as 'directed edges' and may have only one head – they must be uni-directional. The most basic causal structure a DAG can represent is a cause and effect as depicted in Figure 2-1. This DAG (and all others in this thesis) was constructed using the freely available online tool DAGitty (Textor et al., 2011, Textor et al., 2016). The directed edge starting at X and terminating at Y determines that X causes Y. Or, according to Pearl's definition of cause, X causes Y because Y in some way relies on X for its value – the value of Y differs according to changes in X. Or in a counterfactual sense, assuming a perfectly designed and executed RCT on a treatment group and a control group, if the treatment group had been assigned a value of X and the control group had been assigned its comparator, the value of Y for either group would correspond to the potential outcome for the other, and these potential outcomes are different.

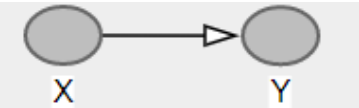


Figure 2-1: Cause and effect

In an observational study, X could be an exposure and Y could be an outcome. For example, the statement that alcohol consumption causes liver cirrhosis would take the following form in a DAG:

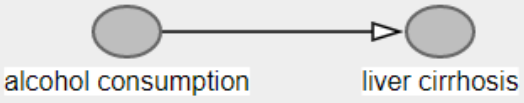


Figure 2-2: Alcohol consumption causes liver cirrhosis

This DAG would be interpreted as follows:

1. The value of the liver cirrhosis variable (e.g. binary) relies on the value of the alcohol consumption variable (e.g. instances of heavy episodic drinking on average month).
2. Because there are no other nodes or directed edges, no other variables which have a direct influence on both alcohol consumption or liver cirrhosis exist.

Of course, DAGs in an applied setting are unlikely to be this simple. A small addition that adds a notable degree of complication is a mediator. For example, in Figure 2-3 alcohol consumption mediates the effect of age on liver cirrhosis.

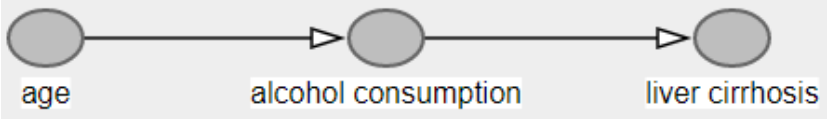


Figure 2-3: Alcohol consumption mediates the effect of age on liver cirrhosis

This diagram encodes twice as many assumptions as the previous:

1. Alcohol consumption is influenced by age.
2. Liver cirrhosis is influenced by alcohol consumption.
3. Liver cirrhosis is not influenced by age other than indirectly through alcohol consumption.
4. No other variables which effect age, alcohol consumption or liver cirrhosis exist.

Note also that, because of the temporality discussed above (i.e. that cause must precede effect), Figure 2-3 describes a person's age at time t1, their alcohol consumption at time t2, and their liver cirrhosis at time t3.

Mediation is a very appealing notion in Public Health, especially regarding exposures that are not readily modifiable such as age, ethnicity, sex, sexuality, parental social class, historical events, etc. In the example of age → alcohol consumption → liver cirrhosis, one would not expect that researchers would consider changing the age of individuals or the age distribution of the population as a practicable intervention for liver cirrhosis. Rather, they might consider alcohol consumption as the more viable intervention site. Similarly, if in any given setting females were at characteristic risk of alcohol harm, researchers would not look towards changing the sex of individuals or the distribution of sex in that population, but rather one of the mediators of sex's effect on alcohol harm (e.g. gendered advertising). Of course, there are likely to be very few instances in applied Public Health research in which the relationship between an exposure and an outcome will be fully captured by a single mediator. Figure 2-4 below is apparently quite simple, but what the directed edge between age and liver cirrhosis suggests is that everything else that happens between t1 and t3, excluding alcohol consumption, can also mediate the effect of age (e.g. hepatitis or non-alcoholic fatty liver disease). This is called the direct effect – everything between the exposure and outcome that does not go through the mediator. Mediation thus recalls the multicausality and component causation concepts introduced above – there is a pool of components involved in the outcome, and the same outcome can be observed in individuals with different sub-sets of this pool of components. Thus, a successful intervention on alcohol consumption would only remove part of the effect of age on liver cirrhosis.

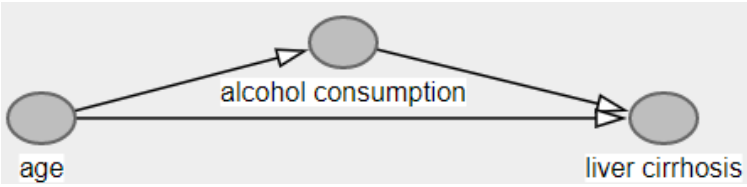


Figure 2-4: Age influences liver cirrhosis via alcohol consumption, and otherwise

Confounding bias is a concern which is perhaps more ubiquitous than questions of mediation. Traditionally, a confounder is thought of as a variable which would bias estimates of the exposure's effect on the outcome because it is related to both the outcome and exposure in some way, while the potential outcomes framework defines a confounder as a mutual cause of an exposure and an outcome (Morgan and Winship, 2007). In a DAG, this equates to a variable with directed edges going to both the exposure and the outcome, as represented in Figure 2-5.

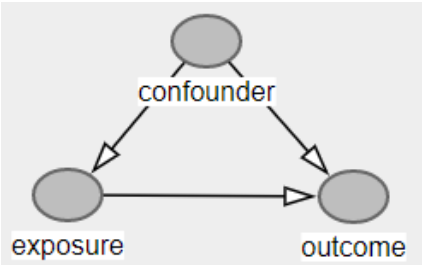


Figure 2-5: A confounder

This DAG encodes the following assumptions:

1. The outcome is influenced by the exposure.
2. The outcome is influenced by the confounder.
3. The exposure is influenced by the confounder.
4. The effect of the confounder on the outcome may also be via the exposure.
5. No other variables that influence the outcome, exposure or confounder exist.

Accordingly, this DAG suggests that the confounder should be adjusted for to estimate the effect of the exposure on the outcome. The DAGitty programme allows researchers to select the exposure of interest and the outcome. It will then search for and identify confounders using a ruleset that is described in some detail below. Figure 2-6 demonstrates this using the same diagram as Figure 2-5 above. The exposure node is yellow with a triangle and the outcome node is blue with a vertical bar. This tells DAGitty that we are interested in estimating the effect corresponding to the directed edge from the exposure to the outcome, which is now shaded green. DAGitty can then identify confounders, which it highlights with red/pink nodes and directed edges. In this case, DAGitty

surmises that the confounder must be controlled or otherwise conditioned to allow for an accurate estimate of the effect of the exposure on the outcome.

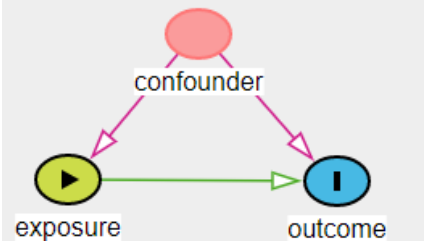


Figure 2-6: A confounder in DAGitty

Figure 2-7 below shows that how we interpret a DAG for the purposes of analysis, and thus how we decide what variables to control for, depends on our research question even under identical assumptions encoded by that DAG.

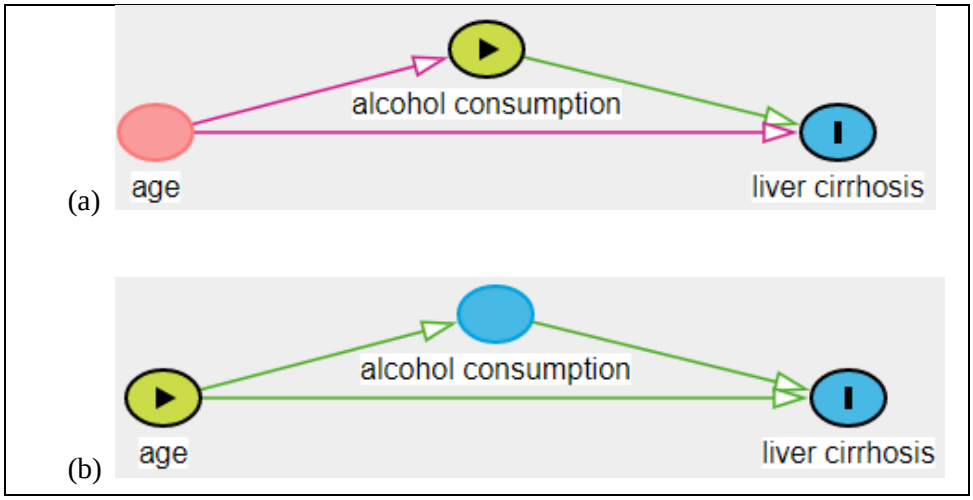


Figure 2-7: Different question, same causal assumptions, different analysis

Figure 2-7(a) indicates that age should be adjusted when estimating the effect of alcohol consumption on liver cirrhosis because it is a confounder. Conversely, Figure 2-7(b) indicates that alcohol consumption should *not* be adjusted when estimating the effect of age of liver cirrhosis because it is a mediator (intuitively, adjusting for alcohol consumption in this example would remove part of the effect of interest, specifically from age to liver cirrhosis via alcohol consumption). Thus, despite how the assumptions in these two diagrams are identical, the resulting statistical models would be very different. This may appear platitudinous, but it is an important idea when applying DAGs, as is made clear in Chapter 8's analysis. It is also helpful to note at this stage that DAGitty allows researchers to set variables to 'adjusted'. In Figure 2-8 below, age has been set to adjusted when alcohol consumption is the exposure, and DAGitty surmises that no further adjustment is required.

Graphically, the red/pink confounder node is now white, and the red/pink directed edges are now black.

Figure 2-8: Adjusted confounder in DAGitty

A third causal structure which is important when estimating causal effects is termed a 'collider'. Where a confounder is a mutual cause of two other concepts or variables, a collider is mutually *caused* by two other concepts or variables. Figure 2-9 presents this in a DAG.

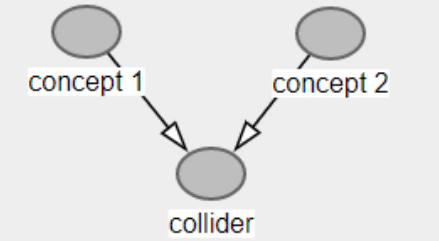


Figure 2-9: A collider

Controlling for a collider creates a spurious association between its mutual causes. DAGs do not make this immediately apparent, but it is a common problem in research, as a collider is effectively DAG terminology for selection bias. A common example used to explain collider or selection bias is selective education (Morgan and Winship, 2007). Consider a university which only admits students with high academic ability or students with high sporting ability. We can thus assume that, as long as this selection process is successful, there will be three 'types' of student at this university: students with high academic ability; students with high sporting ability; and students with both. There should be no students with low academic *and* low sporting ability. In this university there will thus be a negative correlation between academic and sporting ability. To see why consider this simple demonstration in simulated data. As per Table 2-1, there are 100 students in the university, each with a binary indicator of high academic ability and another binary indicator of high sporting ability. Of these, 40 have high academic ability, 40 have high sporting ability, and 20 have both. This produces a moderately strong negative correlation between sporting ability and academic ability of R= -0.667.

Table 2-1: Selection bias at a fictional university

High sporting ability?	High academic ability?		Total
	1	0	
1	20	40	60
0	40	0	40
Total	60	40	100

This correlation, however, is caused by the selection mechanism (i.e. the distribution of both variables in the wider population would be different, for example it would include people with low ability in both academia and sports, and the prevalence of high ability in either category could be assumed to be lower than 60%). Thus, if researchers were interested in estimating the effect of sporting ability on academic ability (or vice versa), conditioning on this selection process would bias the estimate. An example which is more relevant to Public Health, especially for observational social epidemiological studies such as this thesis, selection bias into survey participation. Consider how overall health in the below DAG of depressive symptomology on alcohol consumption acts as a collider (i.e. so-called healthy volunteer bias).

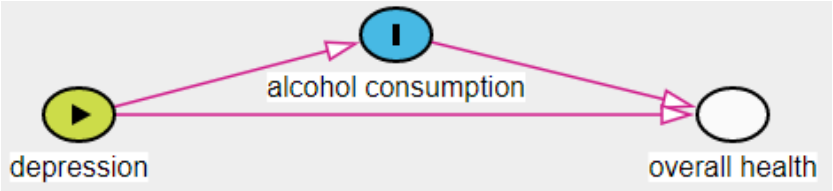


Figure 2-10: Healthy volunteer collider bias

For simplicity, assume that this DAG represents a cross-sectional survey. Because selection is shaped by overall health, and because depression and alcohol consumption influence overall health, any estimate of the effect of depression on alcohol consumption would 'have' collider bias. This is briefly revisited in Chapter 7's discussion of the ALSPAC data. Collider bias is thus a novel articulation of a well-documented problem, but being aware of collider bias, specifically in how colliders should not be conditioned on in statistical models, is crucial for using DAGs to inform analysis.

2.2.2 DAGs: Causal paths

One way to understand DAGs is as a collection of causal paths. A causal path can be 'direct' such as E → O, or 'indirect' such as E → M → O. Crucially, all causal paths must be acyclical – a path started at any variable may never return to that variable (thus directed *acyclic* graphs). This is how

DAGs encode the linearity of time. This collection of direct and indirect acyclic causal paths can be used to identify the causal structures of DAGs described above (i.e. confounders, mediators, and colliders) and their corresponding biases (i.e. confounding bias, 'overcontrol' for mediators, and collider/selection bias). DAGs achieve this using the d-separation algorithm, sometimes known as the backdoor criterion, which can be understood as a method for reliably identifying these forms of bias given the causal paths between variables in any DAG. Or in other words, d-separation determines the 'adjustment set' by identifying variables that cause bias and should be adjusted for, and variables which would cause bias if they were adjusted.

There is only one type of variable that *must* be adjusted for – a variable that 'opens a backdoor path'. A backdoor path starts from the outcome and ends at the exposure, and can therefore be confusing because it requires the researcher to ignore the direction of edges in the DAG. Take Figure 2-11 below. Using DAGitty, the E node has been set to the exposure, and the O node has been set to the outcome. DAGitty automatically performs the d-separation algorithm and has determined that C must be adjusted to estimate the effect of E on O. From the above description of a confounder, we know that is because C is a mutual cause of the exposure E and the outcome O. In DAG terminology, however, we adjust for C because it opens a backdoor path from O to E. Adjusting for the confounder 'blocks' it as a variable, and thus closes the open backdoor path. That path, ignoring the direction of directed edges, is O ← C → E → O (note there are other versions via M as well). This exercise would have to be performed for every possible backdoor path in the DAG.

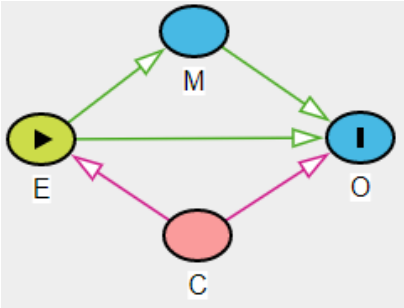


Figure 2-11: Simple DAG for demonstrating an open backdoor path

There are three ways to open a backdoor path – failing to adjust for a confounder, adjusting for a mediator, and adjusting for a collider. Conversely, adjusting for a confounder, and not adjusting for mediators or colliders closes a backdoor path. For example, a closed backdoor path in the above Figure is O ← M ← E → O, because it contains no unadjusted confounders and the mediator is not adjusted. To understand why confounders, mediators and colliders can open backdoor paths, consider each through the lens of bias:

1. Failing to adjust for a confounder opens a backdoor path because the relationship between the exposure and outcome is biased by the confounder's relationship to both.

2. Adjusting for a mediator opens a backdoor path because it systematically removes part of the effect of interest.
3. Adjusting for a collider opens a backdoor path because it creates a spurious association between its mutual causes (a detailed example is included below).

Thus, even if C is adjusted for Figure 2-11, and the backdoor path O ← C → E → O is then closed, the exposure-outcome relationship would still be biased if M were also adjusted for. This is because the previously closed backdoor path O ← M ← E → O has now been opened by adjusting for M. Adjusting for confounders and mediators in the same statistical model thus induces bias. This is a common problem sometimes referred to as the Table 2 fallacy (Westreich and Greenland, 2013). In this thesis, the more specific term 'overcontrol for mediators' is used instead. Note that this is not to say that adjusting for anything that is not a confounder automatically biases effect estimates. For example, the common practice to adjust for 'other risk factors' to improve the precision of estimates does not induce bias as a long as that risk factor's effect is independent of the exposure of interest given the outcome (i.e. it influences the outcome, but not the exposure in any way).

Colliders (i.e. variables caused by two other variables) are likely to be very common in more complex DAGs. However, this is not problematic as the spurious associations induced by adjusting for a collider can often be accounted for by further adjustment. Figure 2-12 below introduces what is sometimes referred to as the M-bias graph, which demonstrates this effect. In Figure 2-12(a), we can see that C1, C2, and C3 each confound the relationship of interest, E → O. The confounding bias of both C1 and C3 is mediated by C2, meaning that C2 is mutually caused by C1 and C3 – it is a collider. However, Figure 2-12(b) shows that, by adjusting for C1 and C2, we can get an unbiased estimate of E → O (i.e. the red/pink arrows are now black). To see why consider first that adjusting for C2 'blocks' the path from C1 to the outcome and the path from C3 to the exposure (such that neither can confound E → O because they do not influence both the exposure and outcome).

However, the spurious association that is induced between C1 and C3 by controlling *only* for the collider C2 would still bias E → O. Think of this spurious association as a latent, if artificial, construct (i.e. analogous to a variable). This is sometimes represented in DAGs by a curved bi-directional arrow, as per Figure 2-12(c). This construct, induced by controlling only for C2 in this example, biases E → O because it has opened the backdoor path O ← C3 ↔ C1 → E → O. In Figure 2-12(d), however, we can see that controlling for C3 in addition to C2 closes this backdoor path because the collider bias is now independent of the outcome. Thus, the spurious association induced by adjusting for a collider can be accommodated using d-separation. The M-bias graph also comments on another aspect of DAGs – there can be several adjustment sets which can satisfy the backdoor criterion (i.e. in this example there are three viable adjustment sets: C1 + C2; C2 + C3; or

C1 + C2 + C3). In an applied setting, the adjustment set that is preferred depends on practical issues (e.g. which has the fewest variables, which variables have the fewest missing data, etc.).

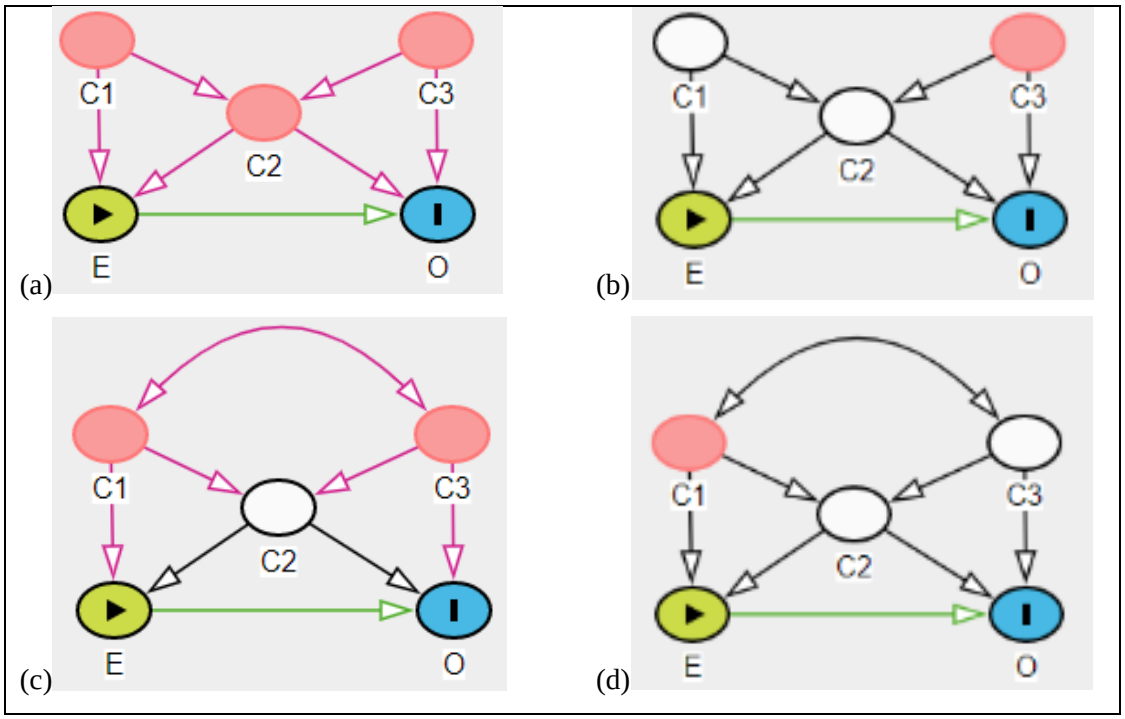


Figure 2-12: The M-bias example of adjusting for a collider

Thus, once researchers have set down their causal assumptions in the form of a DAG, and assuming that there are no cycles in their diagram, d-separation will indicate set of variables that can be adjusted for, thus directly informing adjustment strategies. For example, if researchers wanted to use multiple regression to estimate the effect of an exposure on an outcome, the DAG would indicate which baseline confounders should be added to the model, and which mediators and colliders should not.

2.2.3 DAGs: Limitations

There are clear restrictions in terms of what DAGs are capable of doing. Firstly, they cannot, of course, be used to infer the direction of causation between cross-sectional variables. As alluded to above, they cannot infer the 'truth' of any relationship, nor are they theories of causation – rather they are tools for guiding analysis (Blakely et al., 2016). For example, for cross-sectional variables for which there is no way, using the data, to meet the assumption of temporal ordering, DAGs can be used as transparent and systematic templates for sensitivity analysis, specifically by reversing or deleting the directed edge between pairs of cross-sectional variables. In other words, they can be used to explore how sensitive findings are to different assumed causal structures. However, such DAG-based sensitivity analysis cannot be used to rule out reverse causation between time varying variables

in cross-sectional data (longitudinal data or repeated measures would be required). Of course, measurement of a phenomenon does not equate to its occurrence. For example, even in cross-sectional data, if one or more of the variables in question are assumed to not vary over time per individual (e.g. exposure to historical events, inherent characteristics, etc.) or if the relationships involve a clear causal order (e.g. liver cirrhosis -> death) this should be taken into account. Cross-sectional data does not prohibit researchers from using DAGs to usefully explicate such relationships (e.g. parental ethnicity -> offspring mental health). This DAG-based approach to sensitivity analysis is used in Chapter 8.

A second restriction is that it is unclear how helpful they are for very complicated relationships between many variables (Aalen et al., 2016). However, the DAGs produced in this thesis are good examples of how complex DAGs can still be very useful. Third, DAGs do not explicitly handle either moderation or error terms, as some other structural methods do (for example, structural equation modelling). This becomes more problematic as the number of nodes in a DAG increases, as more moderation effects can be neglected. Further, as noted in Chapter 1 and covered in more detail in Chapter 4, there is a lack of practical guidance for how to build DAGs which may be limiting their use. Finally, high-profile epidemiologists Nancy Krieger and George Davey-Smith have raised concerns that DAGs may be used as a smokescreen to suggest that careful causal thinking has taken place when designing statistical models, when this is not the case (Krieger and Davey Smith, 2016). However, this position has come under heavy criticism for misunderstanding the role of DAGs as simply tools for explicating and interrogating the assumptions underpinning statistical models (Daniel et al., 2016, Pearl, 2018)

2.3 Conclusion

This chapter considered counterfactual reasoning and DAGs in turn. Counterfactual reasoning was described as compatible with common human intuition, in that a factual event and its observed outcome are compared to a counterfactual event and its potential outcome. This was introduced in the context of science fiction and human reasoning, and further explored in a hypothetical RCT. The latter demonstrated how a contrast between average observed outcomes in randomised studies can represent a causal effect, as the observed outcome in one group can equate to the potential outcome for the other. DAG functions were then discussed in some detail, focusing on how DAGs encode causal assumptions, how backdoor paths are opened by failing to adjust for a confounder or adjusting for a mediator or collider, and how d-separation identifies an adjustment strategy for analysis. Chapter 3 continues to expand on this foundation with statistical methods while Chapter 4 expands on

DAGs in general to describe the ESC-DAGs method. The DAG mechanics explained here also underpin the ESC-DAGs review in Chapter 6 and Chapter 7. Chapter 7's discussion on missing data also refers to the above discussion on collider (selection) bias. Finally, Chapters 8 and 9 rely on counterfactual reasoning via POF statistical methods and DAG mechanics. This chapter thus contributes to methodological research objective 1 (MRO 1 - establish a foundation in causal methods) and is crucial for the rest of the thesis.

Chapter 3 Statistical methods

This chapter takes a logical step forward from the previous and describes methods for estimating causal effects from data, thus furthering the foundation in causal methods (MRO 1). It elaborates on how POF statistical methods can be used to compare potential outcomes from different counterfactual scenarios, gives special consideration to assumptions required for valid causal inference, and also refines the working definition of causation. Two types of POF methods are discussed in some detail – the use of inverse probability weights for estimating average causal effects, and causal mediation analysis for decomposing total causal effects. These are the main statistical methods used in the analysis in Chapter 8 and Chapter 9. Methods for handling missing data are also outlined here in brief.

3.1 Data analysis plan

The data analysis plan plays a crucial role in informing each subsequent empirical aspect of this research, including the DAGs, data, and analysis. It is introduced at this stage to act as a helpful reference point going forward. It was developed directly from the research objectives. Thus it performs as a bridge between the overarching goals of this research and their actuation in the empirical chapters. The research objectives that are most concerned with the statistical approach are SRO 4 (test causal effects of parental influences and intermediates on adolescent alcohol harm) and SRO 5 (investigate mediators of parental influences on adolescent alcohol harm). The data analysis plan formalises these goals in terms of the statistical procedures used:

Using the DAGs developed by the novel ESC-DAGs methodology:

1. Estimate regression coefficients for parental influences of adolescent alcohol harm
2. Estimate average causal effects for the parental influences and important mediators
3. Estimate direct and indirect effects for the variables that are determined as causal in step 2

Thus, the end-point of data analysis for this thesis is causal mediation analysis using the variables determined as 'causal' from the previous analyses. However, the methods detailed here are also very relevant to MRO 5 (use the method to direct data analysis under the potential outcomes framework), especially in how the POF operationalises counterfactuals and potential outcomes.

3.2 Counterfactuals and potential outcomes

The POF uses counterfactual thinking to ask 'what if' questions of observational data by comparing the potential outcome for one counterfactual scenario to the potential outcome of one or more alternate counterfactual scenarios (Hernán and Robins, 2019, Morgan and Winship, 2007, Pearl, 2009, Pearl et al., 2016). These counterfactual scenarios are artificially created in data using techniques such as weighting or Monte Carlo simulation. Contrast this with an RCT which uses the observed outcome of one arm of the experiment (i.e. a factual scenario) to make inferences about the potential outcomes of the other. Of course, in observational data, this is not possible because the exposed and unexposed are generally not comparable due to confounding bias and other elements. This is one of the main challenges for causal inference using observational data, and is a focal point of this chapter.

To illustrate, consider hypothetical data comprising a sample of 100 adolescents aged around 16.5 years. We are interested in the effect of the binary exposure 'mother drinks more than 14 units of alcohol per week on average' and the continuous outcome 'number of units of alcohol adolescent drinks per week on average'. The exposure is denoted as X, with 'exposed' = 1 and 'unexposed' = 0, and the continuous outcome is denoted as Y. The POF notation for the potential outcome (adolescent units per week) under the counterfactual exposure (maternal units per week) is Y(X). For each adolescent then, the scenario in which their mother drinks over 14 units per week (possibly counter to the fact) is expressed as Y(1), and the scenario in which their mother does not drink over 14 units per week (possibly counter to the fact) is expressed as Y(0). This example will be returned to throughout this chapter.

As noted in the previous chapter, only one of the potential outcomes Y(1) and Y(0) can ever be observed for each participant, meaning that they can never be contrasted on an individual level. This was referred to as the 'fundamental problem of causal inference' (Holland, 1986). However, the averages for Y(1) and Y(0) can be contrasted. Thus, if it were possible to compare the scenario in which all 100 participant's mothers drink over 14 units per week to the scenario in which no one's mother drinks over 14 units per week for the same 100 participants, then the averages for potential

outcomes Y(1) and Y(0) can be meaningfully contrasted. Therefore, if it were possible to compare these two potential outcomes across all 100 participants, and the various assumptions laid out below are met, then a difference between the averages for Y(1) and Y(0) indicates that X (maternal drinking) has a causal effect on Y (offspring adolescent drinking). Where 'E' refers to the average, a causal effect in the POF is formally stated as:

$$E[Y(1)] \neq E[Y(0)]$$

The difference between these potential outcomes is commonly referred to as the average causal effect (ACE). Calculating average causal effects by contrasting average potential outcomes from mutually exclusive counterfactual scenarios is central to causal inference under the POF.

3.3 Estimating average causal effects

There are two separate but related sets of concerns involved in calculating an ACE. Firstly, there is an underlying group of qualitative assumptions that the POF relies on in order to claim that the contrast between the average potential outcomes is in some way 'causal'. Second, is the practical application of statistical methods to calculate the average potential outcomes that are to be compared.

3.3.1 Assumptions for valid causal inference: Exchangeability

Recall Pearl's pragmatic definition of cause from Chapter 2, that X can be considered a cause of Y if the value of Y depends on the value of X (Pearl et al., 2016). This statement implicitly assumes that there is no uncontrolled confounding bias. The no uncontrolled confounding assumption has been given extensive theoretical treatment by POF scholars and has been rearticulated in several ways, but is perhaps most commonly referred to as the 'exchangeability' assumption. Effectively, in order for the contrast between two outcomes to have a valid causal interpretation, the exposed and unexposed must be comparable except for the exposure – we must be comparing 'like-for-like'. However, exchangeability cannot generally be achieved in observational data, as the exposure we are interested in will be determined by a complex causal structure, meaning that the exposed and unexposed are likely to differ in systematic ways. Thus the observed outcome for the exposed will not equal the potential outcome for the unexposed (and vice versa).

Instead *conditional* exchangeability is achieved in observational data by 'conditioning' on a set of confounders C to make the exposed and the unexposed comparable. Thus conditional exchangeability will be familiar to any researcher who has used regression adjustment or stratification methods in data analysis. Conditional exchangeability is expressed by:

$$[Y(1), Y(0)] \perp X \mid C$$

The technical interpretation of which is that, given C, 'treatment assignment is ignorable'. In other words, what determines the treatment (exposure) value is independent of the potential outcomes, conditional on C. The term 'ignorable treatment assignment' is sometimes used interchangeably with exchangeability (Rosenbaum and Rubin, 1983, Thoemmes and Ong, 2015). Arguably then, as alluded to in Chapter 2, the goal of conditioning on C is to achieve conditional exchangeability by 'emulating randomisation' (Hernán and Robins, 2016).

Effectively, conditional on C, the following three statements communicate that the exposed and unexposed are exchangeable:

- No uncontrolled confounding
- The outcome is independent of exposure assignment
- Exposure assignment is randomised

For further illustration, consider Figure 3-1 below. X is the exposure, Y the outcome, C the confounders, and Z is an ancestor of X. Note, how it encodes exchangeability in two ways: (i) given control for C, there are no other mutual causes of X and Y meaning the X→Y relationship is no longer confounded; and (ii), aside from C (and indeed X) the outcome is independent from causes of X (Z). In other words, because we 'block' the effect of C on X, and because Z has no paths to Y except through X, we can say that what determines Y is independent of what determines X.

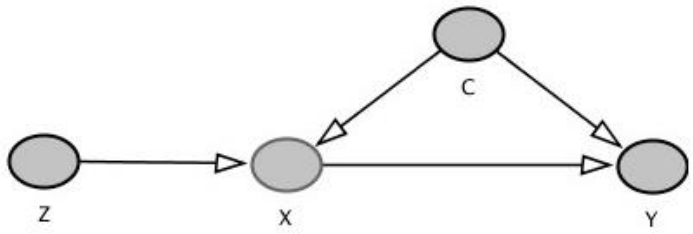


Figure 3-1: DAG showing conditional exchangeability given C

As Chapter 2 discussed in detail, DAGs are powerful tools for identifying causal effects by determining adjustment sets. In other words, one of their primary purposes is to satisfy the conditional exchangeability assumption by selecting C. If conditional exchangeability cannot be assumed then there is no basis for valid causal inference. It is in this regard that DAGs, and the methods proposed in this thesis, find their salience.

3.3.2 Other assumptions for valid causal inference

Other assumptions that are perhaps less essential for understanding and interpreting potential outcomes are consistency, no interference, and positivity. Positivity is simply the concept that there must be both exposed and unexposed participants in each stratum of C. Or, in other words, there must always be a *positive* probability of finding both exposed and unexposed participants in each stratum (Hernán and Taubman, 2008). This is an important concern for building balanced propensity scores, as discussed below.

Consistency means that, for any given value of X, the potential outcome Y(X) is equal to the observed outcome Y. This may appear tautological, but it is important in instances when being 'exposed' may have a systematically different meaning, such that the outcome may not be *consistent* with the exposure value. Hernán and Taubman famously used the example of BMI as an exposure (e.g. for blood pressure) (Hernán and Taubman, 2008). They argued that a low BMI can be achieved in many ways that will lead to different outcome values, including "chopping off an arm". Note however that, while the crux of their argument is that well-defined exposures are necessary for causal inference, if being an amputee is accounted for in C, the distinction between consistency and conditional exchangeability is blurred.

No interference means that the outcome for any one participant is not influenced in any way by the exposure of any other participant. This can be a problem if participants are clustered in settings that are not accounted for in analysis, for example if groups of participants go to the same school or live in the same area. Again, conditioning on C can account for this. Taken together, consistency and no interference are called the 'stable unit treatment value assumption', or SUTVA (Morgan and Winship, 2007).

3.3.3 Calculating and contrasting potential outcomes

This thesis uses a propensity score-based method for achieving conditional exchangeability when calculating ACEs. A propensity score works by calculating the probability (i.e. the propensity) for having the exposure, based on the configuration of confounders (Rubin, 1997, Austin, 2011). It can be expressed as:

$$P = \Pr(X = 1|C)$$

The propensity score P is thus a function of the confounder set C. It is calculated by entering the exposure as the dependent variable in a logit (or probit) regression on C such that P is equal to the fitted probability values from this model. For example, P=0.5 can be interpreted as a 50% chance of being exposed given C. Conditional exchangeability of the exposed and unexposed can be achieved in multiple ways, including, for example, regressing the outcome on the exposure adjusting for P. Thus propensity scores, by modelling on the exposure rather than the outcome, attempt to achieve conditional exchangeability by removing the association between the confounders and the exposure. If there is no association between the confounders and the exposure, then they can no longer be considered as confounders. Or, in DAG terms, propensity scores attempt to remove the edge C→X. Thus, operationalising a propensity score in relation to a DAG involves regressing the exposure on the identified confounders.

Besides adjusting for P, other popular uses of propensity scores include stratifying across different levels of P and matching participants with similar value of P (Austin, 2008, Austin, 2011). The approach used in this thesis is to reweight the data by the inverse of P (called an inverse probability weighting or IPW). IPW is commonly used in epidemiology to create a 'pseudo-population' in which the confounders are unrelated to the exposure (Cole and Hernán, 2008, Thoemmes and Ong, 2015). For each value of P the exposed are reweighted by the inverse of the probability of being exposed, and the unexposed are reweighted by the inverse of the probability of not being exposed. That is, the IPW for the exposed is:

$$\frac{1}{\Pr(X = 1|C)}$$

And for the unexposed:

$$\frac{1}{1 - \Pr(X = 1|C)}$$

This is referred to as an unconfounded pseudo-population because the weighted exposed and unexposed groups are not identical to the population that was actually observed but rather they "could have been sampled from a population in which there was no confounding" (Thoemmes and Ong, 2015). For example, assume that we are interested in the effect of X on Y given univariable C, and that X and C are both binary (Tsai and Goodman, 2013). If the probability (i.e. the propensity score) of being exposed for C=1 is 0.1, then using the above formulae, the IPW for C=1 is 1/0.1 for the exposed and 1/1-0.1 for the unexposed. Thus, for every 9 participants that are unexposed there is one that is exposed, meaning that the one participant who is exposed is given a weight of 10 (i.e. 1/0.1), and the 9 participants who are unexposed are each given a weight of 1.11 (i.e. 1/0.9). As such, there would now effectively be 10 exposed participants and 10 unexposed participants. Using the IPW, an ACE can then be calculated by a weighted regression of the outcome on the exposure, which is "conceptually identical" to running an unweighted regression in the unconfounded pseudo-population (Thoemmes and Ong, 2015). Because the weights themselves are estimated from the propensity score, it is common to use bootstrap standard errors to calculate confidence intervals for the ACE (Cole and Hernán, 2008, Austin, 2011).

IPW has two main advantages over other propensity score-based methods. Firstly, unlike matching (another popular approach), IPW is not restricted by the need to find pairs of participants with identical or very similar propensity scores. Thus IPW can use all of a sample while matching often cannot (Austin, 2011, King and Nielsen, 2016) . Secondly, rather than being used *instead* of regression adjustment, IPW can be used in conjunction with it, in a so-called 'doubly robust' approach. For example, ACEs are calculated in this thesis by first generating the unconfounded pseudo-population and then adjusting for P in the weighted regression models of Y on X. Note that the regression models may feature variables additional to C if the DAG identifies risk factors for Y that are not associated with X. This is considered a doubly robust approach as it has been shown that only the exposure model (i.e. the propensity score) *or* the outcome model (i.e. the adjusted regression on the reweighted data) needs to be correctly specified in order for the output to "remain consistent" with unbiased estimates (Bang and Robins, 2005).

One limitation of weighting-based approaches such as IPW is that they "do not work very well" with continuous X variables, because "they are not very stable" (VanderWeele, 2015). Thus, the analyses in this thesis exclusively used binary exposures and mediators for causal analyses. A further justification was engendered by the data analysis plan introduced at the start of this chapter - the broad focus on parental influences and non-parental influences necessitated a high number of ACEs which was more easily operationalised using binary exposures and mediators. For example, this meant that the interpretation was always in the same form of comparing the exposed to the unexposed. Finally,

as expanded on later in Chapter 7, many variables, while continuous, had heavily skewed distributions (meaning that they would not meet the requirement of normality when they are regressed on covariates, such as when acting as the mediator).

In order for propensity score-based methods in general to achieve conditional exchangeability, each confounder must be balanced at each value of P between the exposed and the unexposed – this is how such methods 'remove' the association between the confounders and the exposure. It is also where the positivity assumption is important, as if there is not a positive probability of finding both exposed and unexposed individuals at each level of P for a confounder, then that covariate will be unbalanced and thus still confound the relationship. Relatedly, special care must also be given to very high and very low values of P, as the higher the propensity for exposure, the lower the probability that participants will be unexposed and vice versa. This can be handled in different ways. Firstly, IPWs can be calculated in a 'stabilised' way, in that the numerator of 1 from the above formulae is replaced by the proportion receiving the observed exposure value (Robins et al., 2000, Thoemmes and Ong, 2015). Secondly, larger weights can be treated as outliers and recoded (e.g. if they lie outside the 5^{th} and 95^{th} percentiles). This is referred to as truncation or trimming (Cole and Hernán, 2008). Propensity score balance can be assessed directly in most statistical software. For example, this thesis uses Stata's in-built suite of propensity score-based methods called `teffects`. It is a flexible approach that can utilise numerous applications of propensity scores, trim large weights, and explicitly identify unbalanced confounders. In cases where the propensity score is unbalanced, then C must be re-specified either by removing unbalanced variables altogether, replacing them with similar measures, or including higher order terms (e.g. interactions) (Caliendo and Kopeinig, 2008). Balance was checked using two popular approaches: a statistical test developed by Imai and Ratkovic (Imai and Ratkovic, 2014) which assesses the overall balance of P across the exposure; and then by inspecting the variance ratios of individual confounders according to guidelines described by Austin (Austin, 2009). The conditions put in place were restrictive - both the statistical test and the individual variables had to indicate balance. Propensity scores were re-specified by removing unbalanced covariates, but this was not common.

Propensity score-based methods are useful additions to the statistical toolkit of social scientists and epidemiologists. Two of their characteristics are of particular note: they can handle high dimensionality, and a misspecified propensity score (i.e. unbalanced) can be rectified. The variation of the full confounder set is expressed by the variable P. As such, problems with overfitting and dimensionality that are common in multiple regression analysis for example, are negated. In fact accurate propensity scores can be calculated with many variables in C (Brookhart et al., 2006). Secondly, unbalanced propensity scores are easily identified and corrected for whereas identifying misspecification in multiple regression models is more difficult. For example, goodness-of-fit

measures such as R^2 do not test model specification nor do they give any indication of how successfully the model has eliminated systematic differences between the exposed and unexposed (Austin, 2011). Relatedly, propensity scores can directly assess whether there are comparable individuals between the exposed and unexposed at any given level of C. Whereas it is very difficult in a multiple regression to assess whether the adjusted effect is calculated via "extrapolation" or from observed exposed and unexposed participants at that level of C (Thoemmes and Ong, 2015). However, it should be noted that despite these advantages, studies comparing the 'performance' of multivariate regression methods and propensity-score based methods have found there to be "no single winner or loser" (Elze et al., 2017, Biondi-Zoccai et al., 2011). Indeed, the coefficient from a multiple regression can be interpreted as analogous to the ACE, given the same causal assumptions as above and correct model specification. This is in line with the main thrust of this thesis that, regardless of which statistical methods are used, a DAG-based approach to variable selection is a valuable tool for quantitative analysts. For example, a DAG would be equally helpful for identifying which variables must be adjusted for in a multiple regression as for which should be in a propensity score.

The propensity score literature is pertinent to both the statistical techniques used in this thesis and the methodological developments. Chapter 1 and Chapter 2 argued that there is only vague advice on how to select variables for analysis (for DAGs specifically). The propensity score literature is a good example of this for other methods. Highly cited papers which have some degree of focus on variable selection (e.g. Brookhart et al. (2006), Austin (2011)) tend to limit their discussions to the technical concern of which 'types' of covariates – those only related to the outcome, only related to the exposure, or both - are the optimal basis for propensity scores. As such, the problematic assumption that the relationships between the exposure, outcome and covariates are already known is prevalent. One of the most commonly cited papers in the propensity score literature is by Peter Austin in 2011 (google scholar citations as of August 2019 >3200). Despite being one of the key didactic articles on propensity scores, it is emblematic of this problem; while there is a detailed discussion on the 'types' of covariates that should be included, the only mention of how to select variables based on evidence or theory is that "the published literature may provide some guidance". Further, the fixation on the technical question of which 'type' of covariate is ideal can be somewhat redundant for many research questions using observational data, as the overwhelming majority of variables that influence an exposure will also influence the outcome.

3.4 Estimating direct and indirect effects: Mediation analysis

Fundamentally a mediation analysis seeks to estimate how much (if any) of the effect of a distal cause on an outcome is conveyed by a more proximal cause. In other words, it seeks to divide a total effect into at least two others: the indirect effect (the part of the exposure's effect on the outcome that is mediated via a purported mediator); and the direct effect (the part of the exposure's effect that is *not* mediated via the mediator). As a statistical approach, mediation analysis neatly aligns with the overall objectives of this thesis. Firstly, because DAGs lend themselves well to effect decomposition, mediation analysis is ideal for exploring the applicability of the ESC-DAGs methodology. A second and third reason were both noted above: mediation analysis is a useful way to explore the causal relationship between an exposure and an outcome; and mediation is a concept with intrinsic value to Public Health (e.g. when mediators are more modifiable than exposures).

There are three forms of mediation analysis that are common in health and social sciences. Sewell Wright's introduction of path analysis in 1921 is recognised as the first formalisation of mediation (Wright, 1960, Mitchell, 2001). Path analysis is most commonly seen in 21st century health and social sciences in the form of Structural Equation Modelling (SEM) (Kline, 2015), but it also informs DAGs (Greenland et al., 1999). In 1986, Baron & Kenny set out their related and very highly cited product of coefficients approach (google scholar citations July 2019 N=>85,000) (Baron and Kenny, 1986). Causal meditation analysis (CMA) is the third form and has been a major focus of POF scholars, particularly over the last decade (Vansteelandt and Daniel, 2017, Vanderweele et al., 2014, De Stavola et al., 2015, VanderWeele, 2015, Daniel et al., 2011). Arguably this prolonged interest can be at least partly attributed to how decomposing an exposure's effect into direct and indirect effects improves the overall evidence base for that exposure (Hafeman and Schwartz, 2009).

The main difference between CMA and SEM and the Baron & Kenny method is how CMA uses 'nested counterfactuals' to decompose total effects by calculating marginal mean differences between potential outcomes, rather than relying on path coefficients (VanderWeele, 2015). The CMA method used in this thesis is called 'mediational g-computation' (Wang and Arah, 2015, Lin et al., 2017). Where non-linearities, interaction terms, and intermediate confounding are highly problematic for SEM, the Baron & Kenny method, and other forms of CMA (De Stavola et al., 2015), the g-computation procedure can handle each (Daniel et al., 2011, Vansteelandt and Daniel, 2017). As such it is used here mainly because it is the most sophisticated POF approach to mediation analysis and least prone to bias due to non-linearities, interaction effects, and intermediate confounders. The following considers several key concepts which are important for understanding effect decomposition in general and for interpreting results from mediational g-computation.

3.4.1 Defining mediation

Pearl's definition of a cause from Chapter 2 (that a variable X can be considered a cause of a variable Y if the value of Y depends on the value of X (Pearl et al., 2016)) can be logically extended to mediation. A variable M is a mediator of the effect of X on Y if M depends on X for its value, and Y depends on M for its value. Figure 3-2 below encodes this definition in a DAG. It states that M is dependent on X for its value, and that Y is at least partially dependent on M. Thus M mediates X. If the directed edge X → M did not exist, then M would not be a mediator but rather an exposure on a different causal pathway. If the edge was reversed to M → X, then X would mediate M. Note that, in a technical sense, X confounds the relationship between M and Y, as it is a mutual cause of both.

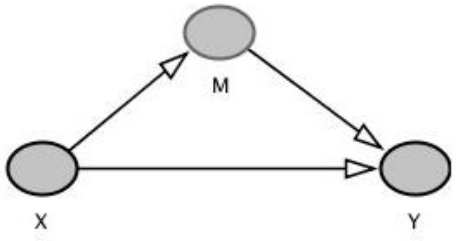


Figure 3-2: Basic DAG showing mediation

The 'total effect' of an exposure on an outcome concerns all paths originating from an exposure and terminating at an outcome. In the case of Figure 3-2 the total effect thus concerns paths X → Y and X → M → Y. The total effect corresponds to the average causal effect described above. The indirect effect can be understood as all paths which start at X, go through M and terminate at Y. Thus the indirect effect in Figure 3-2 is the path X → M → Y. The direct effect is all paths starting at X that terminate at Y, but do not pass through M, in other words; the unmediated effect. The direct effect in Figure 3-2 would be represented by X → Y. However, applied DAGs will invariably be more complex. Figure 3-3 below is still a simple DAG, but adds the variable 'L'. L is another mediator of the effect X on Y that also affects M – it can be understood as a prior or upstream mediator relative to M. According to these definitions of direct and indirect effect, even this simple addition complicates the decomposition of the total effect.

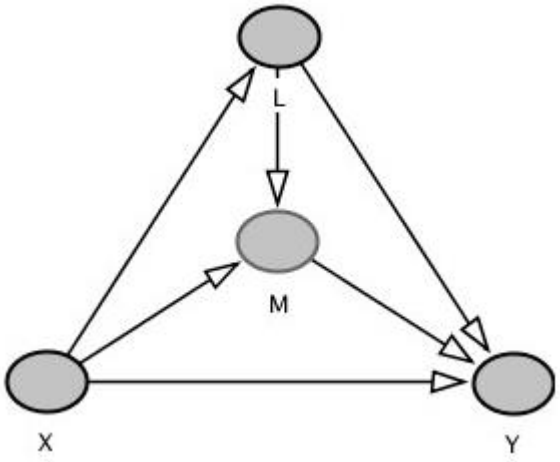


Figure 3-3: DAG showing mediation and confounding

Firstly, the total effect now decomposes across four paths rather than two:

- $X \rightarrow Y$
- $X \rightarrow M \rightarrow Y$
- $X \rightarrow L \rightarrow Y$
- $X \rightarrow L \rightarrow M \rightarrow Y$

Secondly, as the direct effect includes all unmediated paths, it now includes the $X \rightarrow L \rightarrow Y$ pathway. The crucial difference here is that the direct effect is concerned with more than just the directed edge between X and Y. Furthermore, the indirect effect also involves L, as it must include the pathway $X \rightarrow L \rightarrow M \rightarrow Y$. This is a "common" causal structure in applied mediation analysis (Moreno-Betancur and Carlin, 2018). It is referred to as an 'exposure-induced mediator-outcome confounder' (hereafter; an 'EIMOC'). Because an EIMOC is caused by the exposure (i.e. the 'exposure-induced' relationship $X \rightarrow L$) and also confounds the relationship between the mediator of interest and the outcome, it is effectively an earlier mediator of the exposure. This is the intermediate confounding problem noted above - EIMOCs are inherently problematic to mediation analysis because they create a paradox wherein controlling for L blocks part of the direct effect (the path $X \rightarrow L \rightarrow Y$), and not controlling for L allows it to confound the mediator-outcome relationship (and thus bias the indirect effect). The CMA solution is discussed below. First the nested counterfactual approach is explained in detail.

3.4.2 Defining mediation using counterfactuals

Still using Figure 3-3, assume that X, L, and M are all binary and that Y is continuous. There are four effects which can be of interest in causal mediation analysis: the controlled direct effect (CDE); the

natural direct effect (NDE); the natural indirect effect (NIE); and the total causal effect (TCE) (Daniel et al., 2011, VanderWeele, 2015). Each is calculated as the marginal mean difference between two potential outcomes. Thus, it is useful to return to counterfactual notation. The CDE is the most intuitive. Between the two potential outcomes it changes the value of the exposure while holding the mediator at some control value '*m*' (e.g. 0 for a binary variable). It is defined as:

$$\text{CDE} = E[Y\ (1, m)] - E[Y\ (0, m)]$$

Like the potential outcomes discussed above, the CDE can be interpreted as a contrast between the potential outcomes of two counterfactual scenarios. In the first, all participants are exposed, while in the second none are exposed. In both the mediator is held at *m*. Effectively then the CDE statistically adjusts for the mediator while counterfactually intervening on the exposure (e.g. by using simulation techniques given conditional exchangeability). However because the CDE "is a function of *m*" the marginal mean difference may vary under different values of *m* (Daniel et al., 2011). A consequence of this is that the difference between the total effect and the CDE does not then correspond to a controlled *in*direct effect. Note that a difference between the total effect and the CDE may indicate that there is some indeterminate degree of mediation, although this is far from a robust approach to identifying indirect effects. For this reason, the CDE is generally not used in this thesis.

A nested counterfactual, however, can treat the mediator in such a way as to always estimate the same marginal mean difference regardless of the control value used and thus quantify both a direct and an indirect effect. Nested counterfactuals take the form: E[Y(X, M(x))]. The difference is that the mediator is no longer held at an *a priori* control value for all participants, but instead takes on a different value for each, dependent on the (possibly counterfactual) value of the exposure. Specifically, a nested counterfactual is used to calculate the potential outcome Y when X is set to 0 or 1, and when M is set to the value that it would have taken for each participant when X is set to 0 or 1. Thus the counterfactual value or the mediator is 'nested' within a counterfactual value of the exposure such that the value that the mediator takes is one that "evolves naturally" within the counterfactual scenario - thus the *natural* effect (Daniel et al., 2011).

The natural direct effect is defined as:

$$\text{NDE} = E[Y(1, M(0))] - E[Y(0, M(0))]$$

It compares two potential outcomes. The first, E[Y(1, M(0))], is the potential outcome in which everyone is exposed but the mediator is set to the value it would have taken under no exposure. The second, E[Y(0, M(0))], is the potential outcome in which no one is exposed and the mediator is set to

the value it would have taken under no exposure. Thus, it is very similar in form to the CDE in that the mediator is held at a value while the exposure is intervened upon. The difference is that the value the mediator takes derives from counterfactual values of the exposure. As a consequence, it is thus allowed to vary between participants rather than being held at the same value for the full sample. Consider the prior data example where the sample was 100 adolescents, the outcome was the number of units consumed per week and the exposure was a binary variable indicating whether the adolescent's mother drinks more than 14 units of alcohol per week. Assume that the mediator was a binary variable indicating whether the adolescent had negative expectations around alcohol consumption (M=0) or positive expectations around alcohol consumption (M=1). Calculating the potential outcome E[Y(1, M(0))] would involve counterfactually exposing the sample to all mothers consuming more than 14 units per week while letting adolescent alcohol expectations take on the value that it would have taken if all of the sample was counterfactually 'controlled' such that none of the mothers consumed more than 14 units of alcohol per week. This would then be contrasted with the potential outcome E[Y(0, M(0))] in which no one's mother drank more than 14 units per week and everyone's mediator took on the value corresponding to the same exposure status.

The corresponding NIE is then defined as:

$$\text{NIE} = E[Y(1, M(1))] - E[Y(1, M(0))]$$

It contrasts a scenario in which everyone is exposed and the mediator is set to the value that it would have taken under exposure E[Y(1, M(1))], to one in which everyone is exposed and the mediator is set to the value that it would have taken under no exposure E[Y(1, M(0))]. Thus the NIE holds the exposure value at one level while varying the mediator.

The TCE is perhaps more intuitive. It is defined as;

$$\text{TCE} = E[Y(1, M(1))] - E[Y(0, M(0))]$$

In the first scenario E[Y(1, M(1))], all are exposed and the mediator takes the value that it would have taken under exposure. In the second scenario E[Y(0, M(0))], none are exposed and the mediator takes the value that it would have under no exposure. Thus, the estimate produced is simply the difference between the exposed and the unexposed. In effect it is equivalent to the difference between Y(1) and Y(0) and as such is analogous to the ACE from above.

This thesis uses the Stata package `gformula` to calculate the TCE, NDE, and NIE (Daniel et al., 2011). `Gformula` is an implementation of the mediational g-computation procedure. The same assumptions are required for valid causal mediation as for calculating the ACE, with some extensions. Where conditional exchangeability can be assumed when calculating an ACE if there is no uncontrolled confounding of the X→Y relationship, mediation analysis additionally requires that there is no uncontrolled confounding of the X→M or M→Y relationships, including M→Y confounding caused by L (i.e. there are no EIMOCs) (Wang and Arah, 2015). Given a set of confounders C, a set of EIMOCs L, and no uncontrolled confounding, mediational g-computation works by using Monte Carlo simulation to simulate each post-exposure variable in sequence (according to a DAG, for example) to produce a potential outcome. The TCE, NIE and NDE are then estimated by comparing these potential outcomes. For example, assume there were two variables in L as per the below DAG in Figure 3-4. Assume also that C, X, L1, L2, and M are binary while Y is continuous.

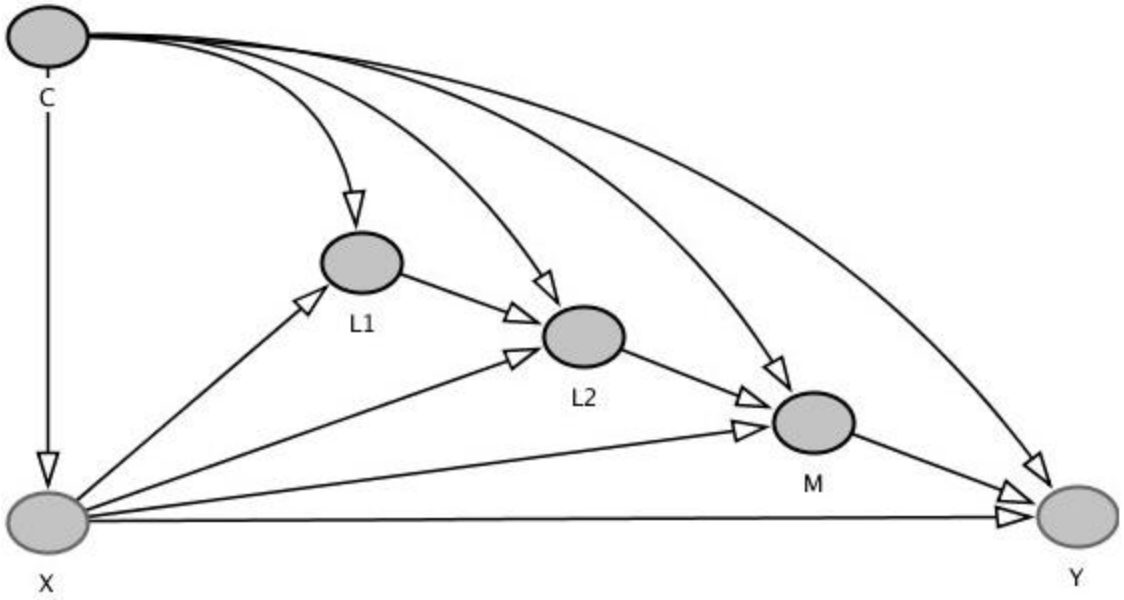


Figure 3-4: Mediation model with two variables in L

The process of 'setting' variables to specific values in order to estimate potential outcomes from nested counterfactuals is achieved by first modelling the relationships in the observed data as per the DAG and then using the parameters from these models to direct Monte Carlo simulation under the interventions of interest. For example, for potential outcome $E[Y(1, M(0))]$, in which everyone is exposed but the mediator takes on the value it would have taken if no one was exposed, L1* is simulated on its conditional distribution given C and X when X is replaced with 0. L2 is then simulated on the conditional distribution given X, C, and L1*. M* is then simulated given X, C, L1* and L2*. Finally, the potential outcome Y* is simulated given X, C, L1*, L2*, and M*. The process

for estimating the other potential outcome necessary for the NIE, $E[Y(1, M(1))]$, is identical except that X is replaced with 1, not 0. The NIE is then the difference between the two simulated potential outcomes (Daniel et al., 2011, VanderWeele, 2015, Thoemmes and Ong, 2015, Wang and Arah, 2015). Mediational g-computation thus only ever conditions on past variables. This is what Daniel et al argue differentiates it from approaches which cannot handle EIMOCs, such as the Baron & Kenny approach (Daniel et al., 2011). The other strength of meditational computation is that exposure-mediator interactions can also be simulated. However, this was not an important focus for this thesis as very few exposure-mediator interactions were detected (covered in detail in Chapter 9). Similar to the IPW estimators described above `gformula` uses bootstrapping to calculate confidence intervals for the TCE, NDE, and NIE.

`Gformula` has been used in multiple settings, including quantifying epigenetic mediators of the effect of smoking on lung cancer incidence (Fasanelli et al., 2015), estimating the direct effect of intrauterine smoking on offspring mental health (Menezes et al., 2013), investigating whether "biogeographical ancestry" has an indirect effect on prevalence of type 2 diabetes via ethnicity (Piccolo et al., 2014), investigating whether number of live births mediates the effect of birthweight on later diagnoses of endometriosis (Gao et al., 2019), and more. It is used in Chapter 9 of this thesis to explore mediators of the effect of parental influences on adolescent alcohol harm, including other parental influences (e.g. parenting and the parent-child relationship), peer effects, and more.

3.5 Key debates in the potential outcomes framework

Until 2017, an important controversy surrounding mediation analysis using nested counterfactuals was that potential outcome E[Y(1, M(0))] is not observable. Outside of simulated data, there can never be a scenario in which a participant simultaneously has the exposure, but in which their mediator takes the value that it would have taken if they had not had the exposure. Thus this potential outcome is impossible in much the same way as the 'fundamental problem of causal inference' discussed above. In the context of natural effects, this impossibility is sometimes referred to as the cross-world independence assumption (Vanderweele et al., 2014). Naimi et al attracted attention with a paper published in the *International Journal of Epidemiology* in 2014, in which they argued that, because of the impossibility of the cross-world independence assumption, the NDE and NIE are not policy relevant (Naimi et al., 2014).

However, in 2014 Vanderweele et al introduced the concept of 'randomised interventional analogues' of the NDE and NIE (Vanderweele et al., 2014). These can be interpreted in much the same way,

except that the value for the nested counterfactual mediator is not held at a unique value for each participant, but rather is held at a random draw from the distribution of the same counterfactual variable. These effects are sometimes denoted as the NDE_R and NIE_R, and they are the output of `gformula`. Notably, in their 2017 paper, Vansteelandt & Daniel explicitly addressed Naimi et al's critique and demonstrated that the NDE_R and NIE_R are 'informative about… particular interventions' (Vansteelandt and Daniel, 2017) and others have shown that they correspond to feasible RCTs (Moreno-Betancur and Carlin, 2018).

Indeed, the POF in general has been a source of intense debate in epidemiology in recent decades. One of the strongest representations of such is an ongoing special issue of the *International Journal of Epidemiology* on causal inference which started in 2016, and in which many of the key authors of the POF (VanderWeele, 2016, Robins and Weissman, 2016, Blakely et al., 2016, Daniel et al., 2016, Pearl, 2018) have conversed with other high-profile epidemiologists (Krieger and Davey Smith, 2016, Vandenbroucke et al., 2016). The most recent response was in December 2018.

The first critique is perhaps the most inarguable: that the POF approach to causal inference should at most be only one part of the epidemiologist's toolkit. However, it is unlikely that POF scholars have ever advocated for a POF methodological hegemony. For example, in 2014 Miguel Hernán, authored an invited commentary in the *American Journal of Epidemiology* on the role of agent-based models for causal inference, in which he concluded that the POF could act as heuristic "intermediate step" between conventional epidemiology and computer science methods for interested epidemiologists and social scientists (Hernán, 2014). A second critique is that the POF relies on too many untestable assumptions. The response has primarily been that the assumptions made in POF analyses are actually present in more conventional methods, and that the POF does not so much rely on more untestable assumptions than usual, but rather treats them in a more explicit manner. The granularity of the conditional exchangeability assumption, for example, supports this response. The critiques of DAGs discussed in Chapter 2 were also from the same special issue of the *International Journal of Epidemiology* (Krieger and Davey Smith, 2016).

A particularly notable and polarising debate was alluded to above regarding the consistency assumption. Hernán and others have strongly argued that exposures should be well-defined. A logical progression of this position is that *interventions* must be well defined. Accordingly, if interventions must be well-defined, can causal effects be calculated for exposures which are non-modifiable, such as a race or sex? As the later empirical chapters in this thesis make clear, this is not a large concern for this research, purely because ethnicity is poorly measured in the data and sex is not associated with the outcome.

3.6 Handling missing data

Multiple imputation was used to address missing data and attrition bias (discussed more in Chapter 7). The objective of any application of multiple imputation is to use the observed data to estimate "plausible values" to replace missing data (White et al., 2011, Royston and White, 2011, Royston, 2004). A random draw of the posterior predictive distribution of the missing data, conditional on the observed data, is used for this purpose. Take the example used earlier where average units of alcohol consumption in 100 adolescents was the outcome, and the exposure was whether or not mothers drank over 14 units of alcohol in the average week. Assume both were fully observed (N=100). Multiple imputation could be used if a third variable with missing data was added, for example if we wanted to adjust for maternal age in years at the time of the adolescent's birth, and the variable was only partially observed (N=80). The multiple imputation approach would first involve using (linear) regression to model maternal age on maternal drinking and adolescent drinking for the 80 participants with observed data on maternal age. A 'single imputation' approach could terminate at this stage by using the regression parameters to predict each of the 20 missing values. However, the imputed data would have no error terms (i.e. uncertainty from the imputation process cannot be quantified). Instead, multiple imputation, uses resampling to make repeated random draws from the predictive posterior distribution to replace the missing data (hence *multiple* imputation). Additionally, a single imputed dataset is not generally regarded as sufficient to capture the error, and so the process is repeated *m* times to produce *m* imputed datasets. Thus, multiple imputation would produce *m* datasets in which the 20 missing values for maternal age vary across *m*.

A very common multiple imputation approach used in practice is Multiple Imputation of Chained Equations, or MICE (White et al., 2011). Its popularity is for several reasons, but one of the most notable is that it can reliably impute missing data for several incomplete variables while quantifying uncertainty in the same way as above. It uses multiple imputation on each variable with missing data in sequence and in iteration. In our example, we also want to adjust for a binary measure of maternal education (has or does not have a university degree). It also has missing data on 20 cases, although only 15 of these cases are also missing on maternal age. Thus both variables have 5 cases which are missing exclusively on them. As a result, 5 of the 80 cases that are observed for maternal smoking will be missing data on maternal education, and vice versa. MICE would then perform multiple imputation on the observed cases for maternal smoking and maternal education sequentially. All 20 missing cases would have imputed values for maternal smoking. A total of 5 of these would be imputed without any input from maternal education. Maternal education would then be imputed, but the 5 cases that were missing exclusively on smoking would now be replaced by their imputation, thus all 80 of the observed cases that are used to predict the imputed value for education would have input

from smoking. The process is then repeated on smoking to use the imputed values for education (hence 'chained equations'). This cycle is iterated multiple times per *m*. Chaining the imputation allows the observed and the unobserved cases to contribute to the imputed values. This is a second reason for the popularity of MICE as it accounts for missing data with more complex patterns (referred to as non-monotone in the literature). An additional reason for the popularity of MICE is that each variable is imputed using its own model. For example, the normally distributed maternal age variable would use linear regression to estimate its plausible values and while the binary maternal education variable would use logit regression.

Multiple imputation is a vast topic and a complicated technique in its own right. Some of the more pertinent technical issues are considered below. See Appendices A, B, and C for a detailed description of the specific imputation process used in this thesis, diagnostics on the imputation models, and descriptives of the imputed data.

When is it appropriate to impute?

There is no real rule of thumb for when MICE should be used over complete case analysis (White et al., 2011). However, for this thesis, MICE was selected over complete case analysis for three reasons. First, complete case analysis was excluded because of how the high volume of variables used in many of the models resulted in them having vastly reduced power or not being identified. Second, the MICE approach is appropriate for the ALSPAC data. MICE is primarily used on data that is 'missing at random' (MAR), meaning that the probability of data being missing and the value of the missing data are dependent on the observed data. Thus, as most surveys including ALSPAC have data that is missing but that can be predicted by the observed data, they can be assumed to be MAR. Indeed, ALSPAC data is frequently analysed after MICE has been performed (Mars et al., 2019, Khouja et al., 2019), including in studies focusing on adolescent alcohol harm (Mahedy et al., 2018, Lassi et al., 2019, Kendler et al., 2018). Third, the degree of missing data in ALSPAC would hamper the external validity of findings from complete case analysis (see Chapter 7).

What variables should be imputed?

Any variable that is to feature in the analysis should be imputed, including the exposure(s) and the outcome. Including the outcome can appear counterintuitive at first, especially in the context of a thesis which is interested in causal inference, rather than prediction. Nonetheless an association with the outcome and the exposures (and indeed confounders) can and should be exploited to predict missing values in the explanatory variables. Thus, including the outcome is entirely in keeping with the aim to replace missing data with plausible values. For the same reason, variables that predict missingness, but that may not be intended to be included in the analytical model, have also been shown to improve imputation (Royston, 2004). Interaction terms and other non-linear variables can

also be included. Historically, the literature is divided on whether interactions should be calculated prior to imputation (referred to as the 'just another variable' technique or JAV) or calculated during the imputation (passive imputation). The approach taken here was to use JAV, as the most concurrent literature suggests (Tilling et al., 2016).

What observations should be imputed?

In longitudinal data, it may be unrealistic to expect that imputations from participants who have not contributed data for a long time may be able to improve the plausibility of the imputed values more than they contribute to uncertainty. For example, an ALPSAC participant who dropped out in the first year offers much less predictive power when imputing data on an outcome 16 years later, thus widening the range of the plausible values (uncertainty). The approach taken in this thesis was to restrict the sample prior to imputation to those participants who had some observed data from the age of 8 years.

How are different variables modelled?

As alluded to above, a wide range of common regression methods can be employed, including linear, logit, ordinal-logit, multinomial logit, Poisson, and more. This flexibility is another reason for MICE's popularity. However, because imputation employs randomness, non-normal continuous measures should not be predicted using linear regression, as a skewed variable will start to normalise under this routine. MICE can instead use an alternative called predictive mean matching (PMM). This works by first modelling the continuous skewed variable on covariates, and then restricting the predictive posterior distribution to a pre-defined range of 'nearest neighbours', from which the imputed value is randomly drawn in each iteration. This has the additional advantage of ensuring the imputation obeys bounded data (maternal age at birth of offspring would not leave age ranges of female fertility that were observed in the sample, for example).

Estimation

When there are *m* datasets, there will be *m* estimates for the models of interest. In the above example, if *m*=20 then there would be 20 different estimates of the effect of maternal drinking on adolescent drinking, adjusting for maternal age and education. These estimates can then be 'pooled' to produce a single coefficient, a confidence interval, and a measure of the Monte Carlo simulation error. Rubin's rules are used for this purpose. They are based on asymptotic theory in a Bayesian framework. Explanation is beyond the scope of this thesis, but it should be noted that each regression and IPW model in Chapter 8 that used the imputed data combined its estimates across *m* using Rubin's Rules as encoded in the STATA `teffects` package. A total of 50 datasets were imputed for the analyses in this thesis (i.e. *m*=50). This number was selected as it was large enough to exploit the law of the big numbers, while still being manageable. However, it was not possible to pool the estimates for

`gformula` in this way, due to its nature as a user-written Stata routine. Instead, the average is provided for each estimate including the confidence intervals, and the number of models for which the confidence intervals crossed the null is also provided as a measure of the variability across the imputations.

Chapter 4 Evidence Synthesis for Constructing Directed Acyclic Graphs (ESC-DAGs)

This chapter directly pertains to MRO 2: develop a method for building DAGs. As discussed in Chapter 1, as far as the author is aware there is currently no guidance available on how to build DAGs in a systematic way. The main proposition of this chapter is that an evidence synthesis protocol that combines DAG theory with both modern and classic causal inference offers an attractive solution. It was on this basis that the ESC-DAGs protocol was developed. The main body of this chapter is dedicated to describing ESC-DAGs in greater detail than was possible in the publication (Ferguson et al., 2019), and to its specific application in this thesis. This is achieved using didactic demonstration of how the protocol would be applied to adolescent alcohol harm using real and hypothetical examples from the literature. The limitations and future directions for the approach, as well as evaluation of how it performed in this thesis, are covered in Chapter 10.

4.1 Contemporary advice for building DAGs

As noted in Chapter 1, the seminal paper which introduced DAGs to the epidemiology community (and social science in general) does not mention how DAGs should be built at all, focusing instead on how DAGs work (Greenland et al., 1999). Since then, advice has been limited to relying on 'background knowledge'. Thus, there has been no progress on the practicalities of building DAGs since their introduction in 1999. As such, while Chapter 1 and Chapter 2 argued that DAGs can be powerful tools for reducing bias in statistical studies, the lack of advice on how to go about building them is problematic.

4.2 Principles for developing the method

The key aim for this chapter was to develop a method for building DAGs from scientific literature. Evidence synthesis is fundamentally concerned with very similar issues and thus is worth brief discussion. In short, evidence synthesis may be defined as the collation and systematic integration of evidence from multiple sources with the goal of comprehensively characterising the scientific understanding of a particular topic (Hanley et al., 2016, Murad et al., 2016). Of the numerous evidence synthesis methodologies that exist, (Sutton et al., 2009) systematic reviews and meta-analysis are perhaps the most common (Higgins and Green, 2011). Clearly then, evidence synthesis protocols share powerful commonalities with the aim to systematise scientific knowledge for building DAGs. However, after appraising the literature, it became clear that evidence synthesis was only able to inform ESC-DAGs very indirectly. There were two mains reasons for this. First, systematic reviews and meta-analyses tend to only be interested in the relationship between very few independent variables (often only one) and a single dependent variable. In other words, they collate evidence on directed edges between only a handful of concepts, while building a DAG is as concerned with how these concepts relate to the wider causal structure. Secondly, evidence synthesis methods are overwhelmingly concerned with characterising the evidence narratively or numerically, rather than graphically. Thus, existing evidence synthesis protocols are not well suited to the task of building DAGs. However, their more general emphasis on systematic, transparent and reproducible protocols was carried forward when designing ESC-DAGs.

As there was no explicit guidance to refer to, and as the related guidance in the evidence synthesis literature was not well suited to the task of appraising numerous directed edges, ESC-DAGs was mainly developed in reference to several underlying principles of graph theory repeated throughout the literature (Morgan and Winship, 2007, Pearl et al., 2016, Greenland et al., 1999, Textor et al., 2011, Pearl, 2009, VanderWeele, 2015). They are discussed below. However, these principles can prove restrictive in an applied setting, especially when research projects are interested in complex social phenomena typical of observational research (such as parental influences on adolescent alcohol harm). Arguably, the restrictiveness of these principles is a key reason why no methods for building DAGs currently exist. Nonetheless, each fundamentally helped shape ESC-DAGs.

1. **DAGs should be built independently of data** because causal phenomena exist independently of the ability of social scientists to measure them. There are at least two characteristics of data that should be ignored when building a DAG - the variables themselves and the timing between their measurements. It must first be assumed that all nodes in the DAG can be replaced with a variable that adequately captures the concept in question. Thus, when a DAG is built and then brought to bear on data, unmeasured confounding may be systematically conceptualised by

comparing the DAG to the data. Second, the timing of measurement between the variables should also be ignored. Thus, it is the relative timing between nodes on a conceptual level that is important – it must be feasible for the purported cause to precede the effect. Given these assumptions, building a DAG independently of data means that it could be used to inform analysis across multiple different data sources, across different research teams, or even to inform data collection itself.

However, focused analysis of a single longitudinal dataset is common in academic, governmental and other research, and as such building a bespoke DAG without consideration of that dataset introduces a high degree of redundancy to the process. Specifically, many directed edges may have only one plausible direction, but be assessed for both (e.g. if variable V2 is only measured 36 months after variable V1 then the directed edge V2 → V1, even if plausible on a conceptual level, is technically implausible in terms of cause and effect).

2. **DAGs should be 'saturated'** such that every possible relationship between the nodes in the DAG has been assessed, and relatedly, it is a much stronger assertion to omit or delete a directed edge than it is to include the same edge.

   This has at least two implications. Firstly, assessing every possible relationship results in a high volume of assessments. To see why, consider how each time a new node is introduced to a DAG the volume of new assessments required would equal the total number of nodes already present in the DAG. Thus, if there were already 49 nodes, then 49 relationships would need to be assessed if a 50th node was added. Similarly, if a 51st node was added, then another 50 assessments would be required. Thus, 99 relationships would have to be assessed simply to introduce two new nodes. If each relationship is assessed in bi-directionally instead of just uni-directionally, the number of assessments doubles. Figure 4-1 below demonstrates graphically how adding a node results in a non-linear increase in the number of required assessments.

   Secondly, if the only directed edges that are not included are those that the researchers are sure do not exist (itself an uncertain concept), then the number of the directed edges, like the number of assessments required to select them, could be very high. This could be problematic in numerous ways, for example when conducting statistical analysis using methods that are susceptible to overfitting. A DAG with 50 nodes might suggest that 48 variables should be adjusted for to analyse the relationship between the remaining pair. It also threatens one of the key strengths of DAGs – their ability to use visualisation to reduce the cognitive load required to conceptualise models during complex data analysis. Having over a thousand directed edges in a DAG may be counterproductive in this regard.

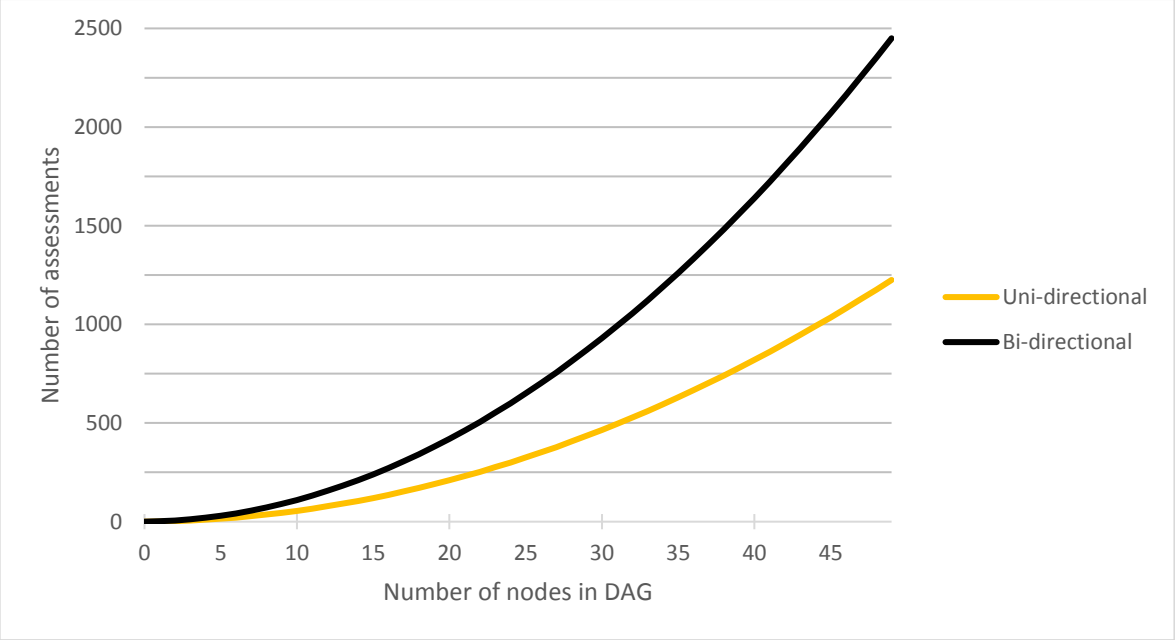


Figure 4-1: Increasing number of directed edges per new node added

3. **DAGs should be built using methods which emphasise the transparency of decision-making** taken in designing the DAG, as DAGs are intended to explicate the causal assumptions of the researcher(s). If a DAG has 50 nodes, then there would be a maximum of 1,225 directed edges (1+2+3… 20+21+22… 48+49+50… etc.). However, a database of 1,225 decisions, while a useful resource, may be cumbersome compared to other options such as relying on the DAGs themselves to communicate the decision-making taken during the DAG-building process. This is discussed below.

4. **DAGs should be informed by the empirical literature** where possible. Again, this seems eminently sensible. However, it could equate to some form of evidence review for every directed edge in a complex DAG. This is not a practical expectation in the context of most research. The approach taken in ESC-DAGs is to focus on the evidence of the focal relationships of the research question, both in terms of the how an exposure of interest effects an outcome, and of the wider confounding structure employed in the literature when investigating this effect.

Rather than a procedural description of how these principles were used to develop ESC-DAGs, the method is demonstrated first before revisiting these topics at the end of this chapter. However, there are two secondary aims that were not related to DAGs or evidence synthesis per se that should be noted. First, explicit consideration was given to how to make the method appealing to the DAG community. If not for application, then at least for discussion. A strategy with two parts was used. On one hand, the targeted journal for the publication had an established record of high-profile

publications in causal inference methods. This was to try and ensure that the method would reach the right audience. On the other hand, ESC-DAGs was designed to be 'modular'. In other words, the protocol operates across several discrete processes, each of which pertains to distinct tasks. As such, the process pertaining to each task may be 'swapped out' for an alternative if users have a preference. This flexibility could also see adaptation of ESC-DAGs or innovation of competitor methods.

The second aim was that ESC-DAGs should be accessible to quantitative researchers who are not familiar with DAGs. This was informed by how this thesis was envisaged as a 'translational' project which might contribute to bringing DAGs and counterfactual causal inference closer to mainstream health and social science. For example, both the publication and this thesis take a didactic stance in several instances. Further, both also make extensive use and recommendation of software such as DAGitty. One reason for doing so is that software automates the arduous task of manually applying the backdoor criterion. In other words, it allows users to build informative DAGs in an applied setting without needing to understand the 'nuts and bolts' of the process. In this way it is analogous to how any regression command in statistical software can be conducted quickly by an applied researcher who can interpret the results correctly without needing an in-depth knowledge of regression minutia. Thus ESC-DAGs may also be applied without the need to have a full understanding of DAGs or previous experience of applying them.

4.3 The ESC-DAGs method

The essential process of ESC-DAGs takes a set of studies from the literature on a particular topic, converts the findings of each of these studies into an individual DAG, and then synthesises these DAGs into one. This process is effectively a form of graphical evidence synthesis. Evidence is thus used to indirectly access the theories that underpin empirical research. The full protocol is presented in Table 4-1 below.

ESC-DAGs Summary

The first step is to use graph theory to convert the narrative and tabulated conclusions of each study identified by a literature search into what is termed here as an 'Implied Graph' or IG. The IG is then saturated such that all variables are connected. These two steps are collectively referred to as 'Mapping' within the ESC-DAGs protocol. The next stage is to evaluate the assertions of the IG using a set of mechanisms described below, and then to correct the IG if appropriate (e.g. by deleting or reversing directed edges). This stage is referred to as 'Translation'. The output of the Translation process for each study is a DAG. The next step is referred to as 'Synthesis'. It firstly involves

extracting each directed edge from each DAG into a 'directed edge index'. These directed edges and their corresponding nodes are then added to a blank diagram in a stepwise fashion until the directed edge index has been exhausted. Once this is done the diagram is saturated and all new directed edges are assessed using the Translation processes. A concurrent stage to Mapping, Translation, and Synthesis is 'Recombination', which involves combining multiple nodes that are deemed to be conceptually similar into one, thus streamlining the process. Once the Synthesis processes are completed, the DAG for the research question is finished. Such DAGs are referred to here as 'Integrated DAGs' or I-DAGs, as they are the synthesis of DAGs derived from the empirical literature. It is unlikely that the output of any application of ESC-DAGs would be a single authoritative I-DAG. Rather, a complex I-DAG, in which there are varying degrees of certainty around different relationships, is more probable. Such diagrams can then be used to direct primary data analysis and as the basis for sensitivity analysis.

4.3.1 Preliminaries

Several preliminary steps and considerations are worth discussing before moving on to demonstrate the method.

What kind of research questions?

ESC-DAGs can be applied to a variety of research questions. For example, research questions pertaining to a single exposure-outcome relationship; a domain of exposures and one outcome; multiple exposures and multiple outcomes; etc. ESC-DAGs is applied in this thesis to parental influences as a domain of exposures and adolescent alcohol harm as the outcome of interest. This was found to be sufficient for identifying mediating structures, although it should be noted that the mediators emerged from this process rather than being *a priori* mediators of interest. However, more complex questions, such as those with a high number of both exposures and outcomes could require an impracticably high workload to produce DAGs using ESC-DAGs. It should also be noted that ESC-DAGs was generally developed with 'causal questions' in mind (does X cause Y, and to what extent?).

What kind of study designs?

While, ESC-DAGs was built with observational studies using longitudinal data in mind, there is potential for the process to be applied within clinical or cross-sectional settings. However, longitudinal data overcomes the key limitation of using cross-sectional data for causal inference – its inability to represent temporal ordering. As alluded in Chapter 2 and expanded on below, temporal ordering is crucial to causal ordering. As such, while ESC-DAGs might be used to analyse cross-

sectional data, longitudinal data should be given priority, especially in cases where relationships between time-varying variables are crucial to the research question (and thus the risk of reverse causation), and in which the DAG is being used to explicate causal (rather than associational) theory.

What kind of DAGs can ESC-DAGs produce?

ESC-DAGs is designed to produce highly complex DAGs with dozens of concepts. There are at least two uses for such a DAG. Firstly, researchers may be interested in generating a complex DAG to act mainly as a conceptual framework to guide multiple forms of research. For example, to help identify potential selection effects for data collection, to act as an over-arching framework for several related projects, or to assess the same relationships in different data sources. Secondly, researchers may be interested in developing a bespoke DAG for a predefined research question, with the intention of using it on a single specific data source. This is arguably the most common reason for wanting to produce a DAG. ESC-DAGs can facilitate each requirement.

Literature review

ESC-DAGs is not a method for developing search strategies. However, the ESC-DAGs protocols should be applied to literature that has been identified using some form of systematic search strategy, for example by using the PICO/PECO guidelines (Population, Intervention/Exposure, Comparison, Outcome (Methley et al., 2014)). Assuming that the ideal I-DAG would be as comprehensive as possible, the ideal review would take the form of a review of systematic reviews or even a novel systematic review, depending on how many extant systematic reviews pertain to the research question. A review of systematic reviews was used for the ESC-DAGs application in this thesis, and the reviews were identified using the PECO strategy (see Chapter 5).

The directed edge index

Early versions of ESC-DAGs attempted to use each study's DAG as the unit of synthesis, rather than the directed edges themselves. In other words, the approach superimposed whole DAGs on to one another. This was found to be inefficient. Instead, for each DAG that is produced from the Mapping and Translation stages, the full list of directed edges is extracted to the directed edge index. This has the additional advantage of compiling all directed edges that have already been assessed, thus helping to avoid repeating the same assessments. In this demonstration, and indeed for the application of ESC-DAGs in this thesis, this list was simply compiled in a spreadsheet file. This is mainly a tool for building the DAGs, rather than a tool for explicating decision-making.

The decision log

A list of how all decisions were made can be kept during the process, as was done for the ESC-DAGs application in this thesis. The main reason for doing so was to explicate the reasoning of the researcher with a view to including a decision log as an appendix. Again, this was done using a simple spreadsheet file. However, once this process was completed the decision log was approximately 200-250 pages long, even after every step was taken to compress its physical space. Furthermore, the decision log offers little benefit over the actual DAGs in terms of transparently explicating researcher assumptions. Instead, the DAGs that were produced for each study are included as Appendix D. Nonetheless the decision log is recommended and demonstrated here, but it is primarily used as a tool for structuring decision-making to help develop the DAGs, rather than a means of presenting decision-making to an external audience.

Scope

Much like a systematic review, there is the potential for the ESC-DAGs process to be highly labour intensive and to produce a DAG with scores of nodes and thousands of directed edges, meaning that it may not always be feasible to use the full version of ESC-DAGs. As such, ESC-DAGs was differentiated into a 'rapid' version. This is discussed further below in the section on applying ESC-DAGs in this thesis (page 71).

Augmentation with data-driven techniques

Data-driven techniques, such as stepwise selection, can be used for helping to design analytical models (VanderWeele, 2019). However, these methods are generally not a suitable starting point for building DAGs, due to their propensity to include variables that should not be controlled for (mediators or colliders, for example). They also cannot be used to identify specific unmeasured confounders. Nonetheless they can be a valuable additional tool, although there are no prescriptive recommendations given in this thesis. For example, stepwise assessment of each confounder's associations with the exposure and the outcome could be used to help build a more parsimonious model when overfitting or related issues are a concern. A recently developed variation on this approach that is specific to DAGs is to use partial correlations to test the conditional and unconditional independencies in a DAG to evaluate how consistent a DAG is with the dataset it is intended to represent (Textor et al., 2016). At the risk of oversimplification, this approach tests the assumption that variable X is independent of variable Y given variable(s) C by controlling C and assessing the correlation between X and Y. Again, this could be very useful in reducing the amount of variables to be modelled. However, unmeasured confounders should still be a concern for any study using data-driven methods. The most transparent approach would be to provide the DAGs from

before and after a data-driven technique was applied, as the difference between the graphs will specify confounders that were identified and not measured.

Nomenclature

Henceforth, the term "study" refers to the published articles that are put through the ESC-DAGs process; "researcher(s)" refers to the authors of those studies; and "reviewer(s)" to ESC-DAGs users.

4.4 ESC-DAGs demonstration

This thesis itself is a demonstration of ESC-DAGs. Chapter 6 is an ESC-DAGs review that demonstrates how the DAGs were built from the evidence covered in the review of systematic reviews in Chapter 5. Chapter 7 is concerned with applying the DAGs to the ALSPAC data. Chapters 8 and 9 focus on analysis directed by the DAGs. The demonstration in this chapter is comparatively simple but still details the protocol from the start of the Mapping stage until the end of the Synthesis stage. The demonstration is effectively a hypothetical application of ESC-DAGs. It assumes that the preliminary concerns have already been addressed:

1. The research question is whether and to what extent an exposure (historical parental drinking) causes an outcome (offspring adolescent drinking).
2. The literature review has been completed and a catalogue of individual studies has been identified for processing (only two studies are used, one hypothetical and the other empirical).
3. Templates for a decision log and directed edge index have been prepared.
4. The appropriate scope is a full application of ESC-DAGs to guide analysis of a single data set.

Finally, it is worth reiterating at this point that the purpose of a DAG is essentially to differentiate variables that should be controlled in analysis (i.e. confounders) from those which should not (mediators, colliders, post-outcome measures, etc.).

4.4.1 Mapping

The purpose of the Mapping stage is to produce an IG - a transparent graphical representation of the conclusions of each individual study. This IG can then be used as a template to apply causal thinking to the assumptions made in the study in the subsequent Translation stage. DAG theory is used to map the relationships without any consideration for their conceptual nature. The only thing that should be controlled from a causal inference perspective is a confounder (Pearl, 2009), thus any variable that was adjusted for in the study is mapped as a confounder, regardless of its conceptual relationship to the other variables. The Translation stage then uses explicit causal and theoretical thinking to assess

these relationships and correct them. For example, if the study controlled for a mediator it will be mapped as a confounder in the IG and corrected to a mediator during the Translation stage.

Table 4-1: Summary of ESC-DAGs protocol

Stage	Purpose	Process
Mapping	To apply graph theory to the conclusions of each study. This creates an 'implied graph' (IG) which acts as a transparent structural template for translation into a DAG.	1. Outcome variable(s) of interest set as DAG outcome(s). 2. Exposure variable(s) of interest set as DAG exposure(s). 3. A directed edge is drawn originating from the exposure(s), terminating at the outcome(s). 4. All control variables are entered as unassigned variables. 5. A directed edge is drawn originating from each control to the exposure(s) and outcome(s) 6. Mediators, instrumental variables etc. are mapped as per study's conclusions. 7. The IG is saturated by drawing directed or undirected edges between all confounders (direction does not matter until the Translation stage). The Recombination process can be performed at this stage to help simplify an overly complex IG.
Translation	To apply causal theory to each relationship in the IG. This creates the DAG for the study. Each relationship in the IG is assessed under sequential causal criteria and a counterfactual thought experiment. See Causal Criteria section below for detailed discussion.	The posited relationship and its reverse are both assessed. Edges may be retained as posited, reversed, or as bi-directional. Otherwise they are deleted. All retained edges are entered into the directed edge index. 1. Temporality – does the posited cause precede effect? (If 'yes', proceed to next criterion. If not, assess reverse relationship). 2. Face-validity - is the posited relationship plausible? (If 'yes', proceed to next criterion. If not, assess reverse relationship). 3. Recourse to theory – is the posited relationship supported by theory? (Always proceed to the counterfactual thought experiment). 4. Counterfactual thought experiment – is the posited relationship supported by a systematic thought experiment informed by the POF? (Once completed, always assess the reverse relationship unless already assessed).
Integration 1: Synthesis	To combine the translated DAGs into one by synthesising all indexed directed edges.	1. A new DAG is created to serve as the Integrated DAG (I-DAG). 2. The focal relationship(s) is added to the I-DAG (as per mapping steps 1 to 3). 3. Each indexed directed edge pertaining to the focal relationship (including its corresponding node) is added to the diagram. 4. Each indexed directed edge pertaining to other nodes is added (e.g. between confounders). 5. Conceptually similar nodes should be grouped together in virtual space to aid the Recombination process.
Integration 2: Recombination	To combine nodes for either practical reasons (i.e. to reduce complexity) or substantive reasons (i.e. to establish consistency).	1. Is there theoretical support for combining two variables/nodes? 2. Do the conceptually related nodes have similar inputs and outputs (i.e. do they 'send to' and 'receive from' the same nodes)?

Take a hypothetical study where researchers were interested in the effect of historical parental alcohol use on their offspring's adolescent alcohol use, and controlled for adolescent sex and substance use. Based on regression coefficients and confidence intervals, the researchers concluded that historical parental drinking was associated with adolescent drinking. Following the process outlined in Table 4-1 above, Mapping begins by entering the exposure (historical parental alcohol use) and outcome (adolescent alcohol use) into the diagram and then drawing a directed edge from the exposure to the outcome. The hypothetical study noted that the control variables were statistically significant in adjusted regressions of the outcome on the exposure and so these are entered into the IG as unassigned nodes. Directed edges are then drawn from both adolescent sex and adolescent substance use to the exposure and outcome, assigning both control variables as confounders. The last step in the Mapping process is to saturate the IG. Directed edges should be drawn between all nodes such that each is connected to all others (the direction of these edges is not technically important at this stage as directionality is assessed during the subsequent Translation stage). The Mapping process is thus completed and the IG for the study is shown in Figure 4-2 below. This IG then acts as a transparent structural template for translation into a DAG.

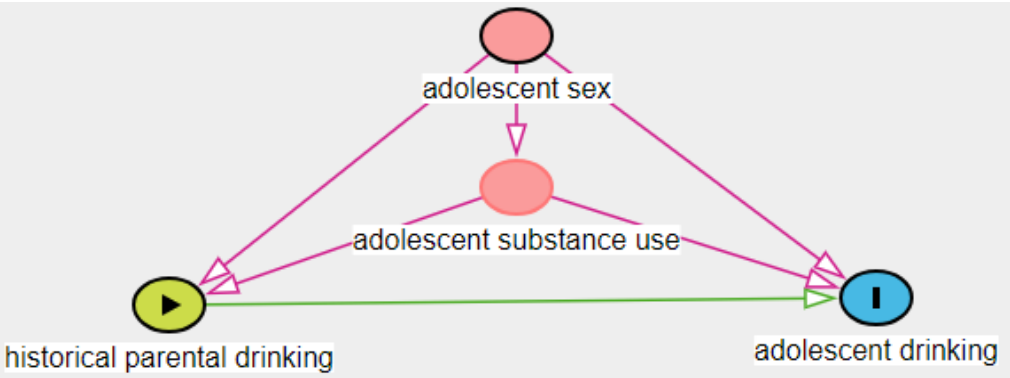


Figure 4-2: Implied Graph (IG) for hypothetical study

4.4.2 Translation

The translation processes were designed to help reviewers efficiently decide on the presence and direction of relationships between nodes in the IG. As such the goal is not to determine the relationship between any two variables in an authoritative sense, but rather to efficiently help design the DAG for that study. This is done in a two-step process for each directed edge. First, three causal criteria are used to determine the plausibility of the directed edge. They are; (1) 'temporal ordering' - that the putative cause *can* precede the effect; (2) 'face-validity' - the relationship is not implausible to the reviewer; and (3) 'recourse to theory' - ideally there should be some theoretical support for the relationship. Each is discussed in more detail below. Next, a counterfactual thought experiment is used to further explicate the assumptions of the reviewer. The causal criteria thus lay the foundation

for the counterfactual thought experiment. The causal criteria are informed by the classic Bradford Hill viewpoints. Table 4-2 below briefly compares them. The counterfactual thought experiment is derived from the POF but is a novel and explicit formulation of a usually implicit cognitive exercise. As such the translation processes can be viewed as a combination of both classic and modern causal thinking.

The process makes two notable assumptions. Firstly, each node represents a variable in the data that is both measured and cross-sectional to the rest of the variables. As noted above, this elevates conceptual and empirical reasoning above simple consideration of the available data. Secondly, bi-directional relationships are possible. This appears to contrast one of the most fundamental aspects of causal graph theory – that no path can return to where it started. This is covered in more detail in the below discussion. The practical implication is that each directed edge should be assessed both in the direction posited by the IG, and its reverse. This is why the saturation process during the Mapping stage is not technically concerned with the direction of the directed edges.

Table 4-2: ESC-DAGs causal criteria and the Bradford Hill viewpoints

Translation causal criteria	**Bradford Hill causal criteria**
Temporality	Temporality
Face-validity	Plausibility (not necessarily biologic)
Recourse to theory	Especially: consistency; coherence and experiment But also: analogy; dose-response; strength; and specificity

The process operates sequentially. In order to assess causal criterion 2 (face-validity), criterion 1 (temporal ordering) must first be established. If the posited directed edge is decided to either not have temporal ordering or face-validity, then it is assumed that it is not a viable directed edge, and the reverse directed edge is assessed from the start of the process. However, this is not the case for the recourse to theory criterion as absence of theory does not necessarily equate to absence of effect (Smith and Pell, 2003). As such, if the posited edge is decided to be temporally ordered and face-valid, the translation process will conduct the counterfactual thought experiment regardless of whether the relationship is deemed to have theoretical support. If the counterfactual thought experiment concludes the posited edge to be plausible, then the posited directed edge is retained and the reverse directed edge must then be assessed. Note that the reverse edge is always assessed after the counterfactual thought experiment on the posited edge is concluded. Once the assessment of the reverse edge is completed, the decision for that directed edge will have been made.

There are four possible outcomes for each directed edge posited in the IG:

1. They can be **retained as posited**, in that the uni-directional relationship is accepted and the reverse relationship is rejected.
2. They can be **reversed** as a uni-directional relationship for which the posited direction was rejected, and the reverse was accepted. This occurs when the study erroneously controlled for a mediator (i.e. the directed edge from the confounder in the IG to the exposure was reversed, thus reassigning the node from a confounder to a mediator).
3. They can be **deleted**. This is only done if it is decided that neither the posited nor reverse edge possess temporal ordering or face-validity, or if both edges fail the counterfactual thought experiment.
4. They can be **accepted as bi-directional**, meaning that both the posited and reverse edges were temporally ordered, face-valid, and passed the counterfactual thought experiment. For research questions in a complex social setting, this will be a common decision. However, because bi-directional relationships cannot be represented in DAGs, bi-directionality should be noted in the directed edge index. This can be an important step for informing later sensitivity analysis.

4.4.2.1 Causal criteria demonstration

Causal criterion 1 – Temporal ordering

Of the Bradford Hill criteria, temporal ordering is the only one that does not require extensive qualification or is yet to be disproven (Glass et al., 2013). For a directed edge to not be temporally ordered, it must be impossible for the purported cause to precede the effect. For example, in the above IG in Figure 4-2, adolescent substance use cannot precede historical parental alcohol use. In this case, the decision would be made to assume that the posited edge is not possible, and so the reverse edge would be assessed from the start of the process. As there is clear temporal ordering from historical parental alcohol use to offspring adolescent substance use, the next step would be to assess this reverse edge for face-validity. Note that, in order to use this criterion, the Translation process always assumes that variables are measured cross-sectionally (i.e. the timing of measurement in the primary study shouldn't be relied upon to aid decision-making). While this may appear to contradict the emphasis placed on longitudinal data above, this is not the case for at least two reasons. First, the temporal ordering criterion aims to establish if either of a pair of concepts *could* occur before each other in a causal sense – it is not to infer causation. Thus, it is the occurrence of the phenomena in question that is paramount. Assuming cross-sectional measurement eliminates the timing of measurement from considerations, allowing researchers to focus instead on the occurrence of the underlying concepts and the nature of their relationship outside of data (this does not contradict the practical impetus of using longitudinal data to rule out reverse causation during the analysis phase). Second, even if it is assumed that all variables are measured cross-sectionally, the temporal

relationship between a time-varying and a non-time-varying concept will not be obfuscated – rather the reverse is true. For example, alcohol consumption in adolescence cannot precede biological sex.

Causal criterion 2 – Face-validity

The face-validity criterion is a means of using reviewer background knowledge to identify implausible relationships, given the temporal ordering established in criterion 1. The reverse edge from historical parental alcohol use to adolescent substance use would be face-valid – it is highly plausible that parental substance use in the form of drinking (whether concurrent or historical) could influence offspring adolescent non-alcoholic substance use. An example of a relationship that is temporally ordered but that is not face-valid in Figure 4-2 would be the reverse directed edge from historical parental alcohol use to adolescent sex (note that this is based on the assumption that sex assignment is effectively random).

Causal criterion 3 – Recourse to theory

Recourse to theory is concerned with why the current directed edge was decided to be face-valid. This step requires the reviewer to explicitly consider their reasoning before moving onto the more formulaic counterfactual thought experiment. This could take the form of identifying a specific theory pertaining to the directed edge. For example, there are numerous candidate theories pertaining to the reverse directed edge from historical parental drinking to adolescent substance use, such as adolescent modelling of observed or described parental behaviours (Ryan et al., 2010), genetic heritability of susceptibility to addictive substances, or genetic heritability of personality types associated with substance use such as sensation seeking (Salvatore et al., 2014). Alternatively, the reviewer could briefly state an informal theory of change. For example, that hearing stories about one's parents' binge drinking might result in a desire to replicate/avoid the behaviour. Thus, the recourse to theory step is not prescriptive and largely open to interpretation. Its value, assuming temporality and face-validity, lies in how explicit consideration of *how* the cause might change the effect prepares the reviewer for the next step. Note that it is perfectly feasible for there to be directed edges in a complex DAG that are completely outside the expert knowledge of the reviewer(s). In such cases a theory of change may be hypothesised or the criterion may even be skipped.

4.4.2.2 Counterfactual thought experiment demonstration

Once the temporal ordering and face-validity of the directed edge have been determined, and some thought has been given to theoretical support, the counterfactual thought experiment is used to further explicate the assumptions of the reviewer(s) in the language of the POF. Recall that counterfactual thinking compares the outcome that would have occurred if all of the sample had been exposed, to the outcome that would have occurred if all of the sample had not been exposed (Pearl et al., 2016,

Morgan and Winship, 2007, Hernán, 2018). The counterfactual thought experiment employs this heuristic in a formulaic and transparent way by comparing the potential outcomes of two 'counterfactual exposures'. This is achieved in reference to how the corresponding variables were measured in the study being translated.

Some aspects of the counterfactual thought experiment are worth explicating before demonstrating the process. Firstly, it makes the exchangeability assumption outlined in Chapter 2. This can be viewed both as an advantage a disadvantage. On one hand, making this assumption frees the reviewer to make a decision without being restricted to data. On the other hand, a thought experiment itself can only rely on reasoning and the reviewer's knowledge, and thus cannot be tested (unless it determines a formal hypothesis). Nonetheless, it is a very useful system to quickly make transparent decisions when building a DAG. Secondly, the counterfactual thought experiment asserts that a non-modifiable exposure (e.g. sex, ethnicity, etc.) can be used for counterfactual causal inference. As noted in Chapter 3, this is a particularly controversial topic in causal inference (Pearl, 2009, Vandenbroucke et al., 2016). However, the stance in this thesis is that, as "causation exists without manipulation", non-modifiable exposures should be treated as causal (Pearl et al., 2016). At least, the lack of the ability to intervene should not preclude a thought experiment. Third, it is recommended that the counterfactual thought experiment assume the exposure is binary for the sake of efficiency. For example, if an exposure in a study was a continuous measure of maternal units of alcohol consumed per week, the counterfactual thought experiment could consider the different potential outcomes when the sample is exposed to all mothers consuming 14 or fewer units compared to all mothers consuming 15 or more units. A fourth consideration is that, while the counterfactual thought experiment is inherently subjective, it nevertheless provides an explicit formalism to log research subjectivity. Finally, the counterfactual thought experiment defines causation in the same way as explained in Chapter 2 – variable X is a cause of variable Y if Y in any way relies on X for its value.

Take the example of the effect of adolescent sex on adolescent alcohol use from the IG. Sex is the exposure node and is a binary variable with values 1 and 0, which we can let equal male and female respectively. The counterfactual exposure would therefore be all participants 'set' to male (or female). The study measures the outcome of adolescent drinking as the number of units consumed last week. Similar to how the POF relies on average effects to overcome the fundamental problem of causal inference (Hernán and Robins, 2019, Holland, 1986), the counterfactual thought experiment then hinges on the question of whether we would expect equivalent average scores between the potential outcomes (number of units consumed per week) for counterfactual exposure 1 (if all of the sample were male) and counterfactual exposure 0 (if all of the sample were female). Based on the temporality, face-validity, recourse to theory, and causal thinking encouraged by the counterfactual thought experiment, the conclusion in this case was that the number of units consumed by the

adolescent last week (the potential outcomes), would differ between the two counterfactual exposures (sex) and so the directed edge from sex to adolescent alcohol use would be retained.

Table 4-3 below examples how the directed edge from historical parental drinking to adolescent drinking could be assessed using a decision log. In this example the Translation process concluded that a uni-directional edge from historical parental drinking to adolescent drinking should be retained in the DAG. To summarise, first the causal criteria assessment found that the posited edge was temporally ordered and face-valid, but that the reverse edge was not temporally ordered and was thus rejected. The recourse to theory step identified parental modelling as a supporting theory and referenced the literature. Then the counterfactual thought experiment concluded that the posited edge was plausible because one would not expect that the potential outcomes (average number of units of alcohol consumed last week by the adolescents) would be the same for counterfactual exposure 1 (exposure to maternal drinking of 15 units or more per week) and counterfactual exposure 2 (exposure to maternal drinking at 14 units or less per week).

Table 4-3: Example decision log – historical parental drinking and adolescent drinking

Information for Posited Edge			
	Puported cause	**Effect**	
Concepts	Historical parental drinking	Adolescent drinking	
Assessed prior? Include reference	No		
Measurement	Number of units per week consumed on average in the last year	Number of units consumed last week	
Causal Criteria			
	Question	**Answer (Y/N)**	
Temporally ordered?	Posited edge	Yes	
	Reverse edge	No	
Face-valid?	Posited edge	Yes	
	Reverse edge	Not applicable	
Theoretical support?	Posited edge		
	Theory:	Parental modelling	
	Reference:	Ryan et al 2010, Sharmin et al 2017	
	Reverse edge		
	Theory:	Not applicable	
	Reference:	Not applicable	
Counterfactual Thought Experiment			
Posited edge counterfactual exposure 1	Maternal alcohol consumption of 15 units or more per week on average over the last year		
Posited edge counterfactual exposure 2	Maternal alcohol consumption of 14 units or less per week on average over the last year		
Hypothesised outcome for posited edge	Counterfactual exposure 1 would be expected to result in a higher average number of units consumed last week by the adolescent		
Reverse edge counterfactual exposure 1	Not applicable		
Reverse edge counterfactual exposure 2	Not applicable		
Hypothesised outcome for reverse edge	Not applicable		
Conclusions			
Outcome of assessment	Retained as posited		
Directionality	Uni-directional effect		
Harmful?	Harmful effect; higher historical parental drinking increases adoelscent drinking		

Conducting this process on every edge in the IG produces the DAG for the study. Some additional aspects of the decision log are worth noting. First, if a directed edge has been assessed before, there is no benefit to assessing it again. Also, the decision log is used here to hypothesise on the nature of the directed edge (i.e. whether or not the relationship is harmful or protective). And lastly, the decision log explicitly requires the directionality of the edge to be recorded (i.e. either uni-directional or bi-directional).

Figure 4-3 below is the DAG that is produced by applying the causal criteria and counterfactual thought experiment to every directed edge in the IG from Figure 4-2. Note how the directed edge from adolescent sex to historical parental alcohol use has been deleted and the directed edge from adolescent substance use to historical parental alcohol use has been reversed. These changes were typical of the ESC-DAGs application to the literature on parental influences on adolescent alcohol harm in that there is over-control for a mediator (adolescent substance use) and over-control for a variable which does not confound the exposure outcome relationship (although this is often done in practice to improve precision of estimates, which is not usually problematic in a causal sense).

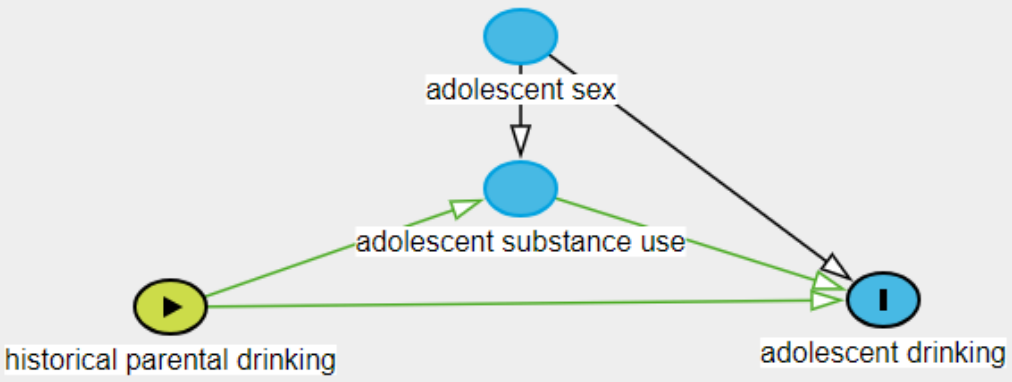


Figure 4-3: DAG for hypothetical study

A final aspect of note on Translation is that the emphasis of the research question is likely to be reflected in the graph, especially regarding relationships that appear to be bi-directional. For example, in Figure 4-3, if the outcome of interest was adolescent substance use rather than adolescent drinking, the DAG is likely to posit drinking as the mediator instead. Dealing with ostensibly bi-directional relationships is a challenge that is discussed throughout this thesis, most notably in Chapter 8.

4.4.3 Synthesis

As alluded to above, the Synthesis stage involves first extracting all directed edges from each DAG into the directed edge index and then entering each edge into a new diagram until the index has been exhausted. This produces the initial I-DAG. However, at this stage, because the exact list of variables used in studies will vary and will possibly be unique in some cases, new relationships between nodes from different DAGs will need to be assessed. Thus, the I-DAG should be saturated and the new edges put through the full translation process. Once this step has been completed, any variables which the reviewers consider to be important but have not been included in the literature may be added. Once again, the relationships between these variables and the rest of the I-DAG should be put through the Translation processes. It should be noted that the order that the directed edges are synthesised in does not matter – the I-DAG will be equivalent.

Synthesis is demonstrated here on the directed edges from the hypothetical study and those from an empirical study by Seljamo et al (Seljamo et al., 2006). They were interested in parental predictors of adolescent alcohol use. They used multiple regression models with early adolescent alcohol initiation, family structure, adolescent sex, and historical parental alcohol use as explanatory variables. The DAG for this study, produced from the Mapping and Translation stages, is presented in Figure 4-4.

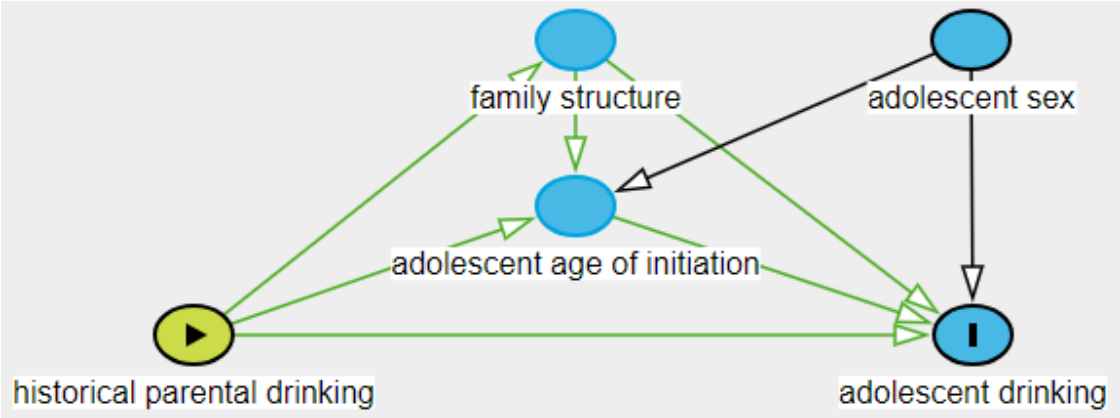


Figure 4-4: DAG for Seljamo et al study

An I-DAG was produced by entering each directed edge from both studies into a new diagram, saturating the diagram, and assessing each new directed edge. For example, when family structure was introduced a new relationship with adolescent substance use was implied. The Translation process concluded that family structure causes adolescent substance use, although the potential for a bi-directional relationship was noted in the directed edge index (on grounds that offspring drug abuse could influence parental conflict). The I-DAG is presented in Figure 4-5.

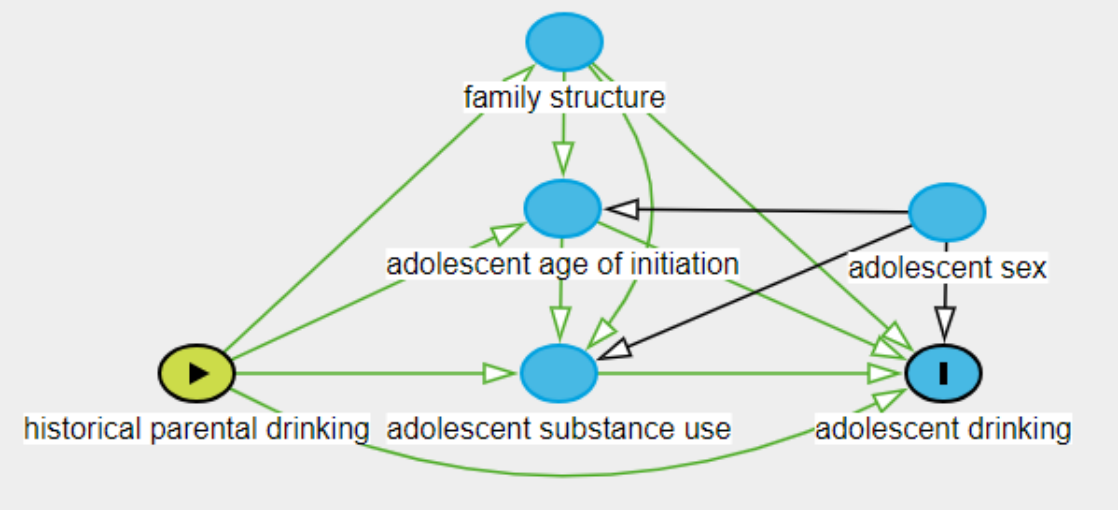


Figure 4-5: I-DAG for hypothetical and Seljamo et al studies

This I-DAG determines that the exposure-outcome relationship of historical parental alcohol use on adolescent alcohol use is unconfounded by any of the suggested covariates. This is typical of the ESC-DAGs application in this thesis, in that over-control for mediators and controlling for other causal pathways (i.e. adolescent sex) is common in the literature on adolescent alcohol harm. The implications of this are that effects that may have been reported implicitly as total effects may instead have been analogous to direct effects. As such, it is possible that the literature may underestimate the importance of parental influences on adolescent alcohol harm. However, several items are worth noting. Firstly, all directed edges between family structure, adolescent substance use, and adolescent drinking were recorded as being potentially bi-directional. Secondly, the directed edges involving adolescent sex, historical parental alcohol use, and age of adolescent alcohol initiation were all uni-directional, mainly because of the historical nature of these variables. Consider, for example, how the I-DAG would have been different if current parental drinking was the exposure instead of historical parental drinking (family structure, age of adolescent alcohol initiation, and adolescent sex might all have been confounders). Third, the only directed edges that were deleted were those from adolescent sex to historical parental alcohol use and family structure. Fourth, any concepts that were not covered in any study may be added to the I-DAG as an 'additional' node. They may then be saturated as regards the rest of the graph, and their directed edges. And finally, this I-DAG is a very simple example for demonstration. In reality, the I-DAGs produced from this process will be exponentially more complex.

4.4.4 Recombination

As the I-DAG grows more complex through further Synthesis, it may be efficient to consider 'recombining' multiple similar nodes. There are at least two indications that recombination may be acceptable for any two nodes. The first is theoretical support. For example, if both nodes are categories of another concept (e.g. parental monitoring and autonomy granting as categories of parenting practice) or if they are used interchangeably in the literature (e.g. alcohol consumption being measured in many different ways such as various quantity-frequency measures, population sales, or clinical questionnaires like CAGE or AUDIT scores). A second indication is if nodes have identical directed edge input and output (i.e. receive from/send to the same nodes). If they feature different directed edges, then there is less support for recombining them.

As an example, if adolescent smoking was added to the above I-DAG, a further six directed edges would also be added (one for each covariate including substance use). However, the directed edge input and output would be identical to adolescent substance use. Further, smoking can be considered a form of substance use and the relationship between substance use and smoking is very difficult to untangle. Thus, on a conceptual level, there may be support for combining these two nodes into one as an efficiency measure that would reduce the complexity of the I-DAG by six directed edges and one node. However, other than suggesting that similarity between nodes conceptually and relative to the rest of the DAG can indicate that nodes *could* be combined, ESC-DAGs stops short of being prescriptive on this front. In this example, there are clear concerns about conflating a drug that is illegal for purchase by adolescents (tobacco) with others that are illegal at any age. Further, this will depend on the research question – it may be inappropriate to combine nodes for one research question which are recombined for another. For example, while recombining illicit substance use and tobacco may be problematic for an alcohol harm outcome, it could potentially be more problematic if the outcome was psychosis (Moore et al., 2007). Any application of ESC-DAGs should keep a record of which nodes, if any, are combined. Again, a spreadsheet document was used during the application of ESC-DAGs in this thesis (specifically an addendum to the directed edge index). Examples included concepts like adolescent school engagement being combined with adolescent truancy, enjoyment of classes, and extra-curricular engagement. The application of recombination is covered in detail in Chapter 6.

The product of the Synthesis stage is the I-DAG for the research question. However, the nodes in the I-DAG at this stage are still only conceptual. As such, ESC-DAGs differentiates between a 'conceptual I-DAG' and a 'data I-DAG'. A conceptual I-DAG is the output of the Synthesis and Recombination stages. It is theoretically applicable to any data source, potentially even by other researchers. It is the template for the data I-DAG. The data I-DAG is produced by replacing as many

nodes as possible from the conceptual I-DAG with appropriate variables from the data. This process is largely comparable to any other quantitative study that selects variables from a dataset to match the concepts of interest. Note that unmeasured nodes should be retained in the data I-DAG.

Once the data I-DAG is produced the ESC-DAGs protocol is concluded. The data I-DAG can then be used to direct data analysis and sensitivity analysis. Note that it is during the transition from the conceptual I-DAG to the data I-DAG that redundancies in the previous processes come to light, especially in terms of building the DAG independently of data. For example, if a DAG had 50 nodes, and the node in question was unmeasured, the 49 relationships that would have been assessed earlier are not valuable for informing statistical models in that data. The exceptions are the directed edges between unmeasured nodes and the exposure(s) and outcome(s). Firstly, determining whether a covariate confounds or mediates the exposure-outcome relationship is always valuable information, even on a purely conceptual level. Secondly, identifying unmeasured confounders can helpfully inform sensitivity analysis of unmeasured confounding.

4.5 Application in this thesis: 'rapid ESC-DAGs'

Evidence synthesis methods like systematic reviews or meta-analysis are rarely applied by a single researcher. Normally a small team of reviewers would be involved, both to increase the degree to which the background knowledge of the researchers covers salient topics, and for efficiency and reliability (Higgins and Green, 2011). However, in the context of this PhD, ESC-DAGs was applied by a single researcher, with oversight from supervisors. Thus, to make sure that the application in this thesis was compatible with the breadth of the research objectives (e.g. using the POF to estimate total and indirect effects of parental influences on adolescent alcohol harm), ESC-DAGs was differentiated into a second version called 'rapid ESC-DAGs'. A summary of it compared to the full version, is presented in Table 4-4 below.

Rapid ESC-DAGs does not differentiate from the full version until the Synthesis stage, meaning that the Mapping and Translation stages still produce the same set of DAGs, and the directed edge index used for Synthesis is identical. However, rather than synthesise all of these directed edges, only those related to either the exposure of the outcome are synthesised. The conceptual nodes are then immediately replaced with corresponding variables from the data source. Then the empirical timing between measures is leveraged to greatly accelerate decision-making. This is achieved by saturating the data I-DAG in a forward direction, such that each variable has a directed edge drawn from it to all other variables that were measured afterwards. These directed edges are then assessed with the

Translation criteria. Finally, any cross-sectional grouping of variables is fully saturated and assessed. The 'rapidity' of this version lies in how it identifies directed edges that can only be uni-directional. In other words, where the full version accepts the redundancy of assessing directed edges that may not even be possible in longitudinal data, rapid ESC-DAGs largely rejects this redundancy to increase efficiency and speed. Box 4-1 below conveys the full rapid ESC-DAGs process in 11 steps.

Several implications are worth noting. First, rapid ESC-DAGs contradicts the principle that DAGs should be built independently from data. This contradiction is covered in the below discussion. Second, rapid ESC-DAGs offers no advantages over the full version if the data are cross-sectional as even in cohort or other longitudinal data, all variables that are measured cross-sectionally to one another must still be saturated and assessed bi-directionally in rapid ESC-DAGs. Further, if ESC-DAGs is being used to generate a DAG to be used across multiple datasets, a single application of rapid ESC-DAGs will not be sufficient and it may be more time consuming than the full version (the saturated conceptual I-DAG in the full version is more immediately applicable to data than the unsaturated conceptual I-DAG in rapid ESC-DAGs). However, if research projects are only interested in analysing a single dataset, or if time pressures preclude the full version, then rapid ESC-DAGs may be preferable. Lastly, redundant assessments may still occur during the Translation stage in rapid ESC-DAGs, in that relationships that were possible in the data sources of the primary studies may not be possible in the data I-DAG.

Table 4-4: Rapid ESC-DAGs vs full ESC-DAGs

Stage or component	Difference	Purpose for change
Mapping	Conducted in identical fashion to the full version, in that saturated IGs are translated into DAGs, and the directed edges are recorded in a directed edge index.	NA
Saturation	All IGs are saturated as per the full version. However, the conceptual I-DAG is not saturated at all and the data I-DAG is only partially saturated: firstly, it is 'forward saturated' such that all earlier variables are posited as causes of all later variables. Secondly, any cross-sectional groups of variables are fully saturated	To avoid redundant assessment of directed edge that are not possible in the data.
Translation	The process itself is the same but it is only applied to the IGs, to the 'additional' concepts in the conceptual I-DAG, and to the saturated cross-sectional groups in the data I-DAG. The full version applies the Translation process to all concepts in the conceptual I-DAG as well.	NA
Synthesis	Only the directed edges that relate to the exposure or the outcome are synthesised from the directed edge index. 'Additional' concepts are added, translated in relation to the exposure and the outcome, and then only further integrated if matched with a variable.	Synthesising the relationships between the non-exposure and non-outcome covariates may induce the same redundancies as above, in that many are unlikely to be possible given the data.
Recombination	Proceeds in the same way as the full version	
IGs	Identical	NA
DAGs	Identical	NA
Directed edge index	Identical	NA

Unmeasured concepts	Only need to be posited and translated in relation to the exposure and the outcome in order to differentiate them as confounders or mediators.	Full saturation regarding an unmeasured concept is redundant for data analysis.
Conceptual I-DAG	Produced only from the relationships from the directed edge index that involve the exposure and the outcome	The conceptual I-DAG's main purpose is to act as a template for conversion into a data I-DAG for any given data source. As such, taking the time to generate a comprehensive list of conceptual relationships introduces redundancies.
Data I-DAG	Initially produced by replacing concepts with variables, and then using the timing of the data to help decide on the direction of as many directed edges as possible	The data I-DAG is produced in a very different way in rapid ESC-DAGs. However, it is highly likely that both versions will produce the same or very similar diagrams given the identical reliance on data.

Box 4-1: Rapid ESC-DAGs in 11 steps

Processes common to ESC-DAGs and rapid ESC-DAGs:

1. Each study is put through the Mapping process, which involves depicting control variables as confounders unless otherwise indicated in the study. This produces an implied graph (IG) for each study.
2. Each IG is then saturated.
3. Each directed edge is assessed using the Translation process (i.e. the causal criteria and counterfactual thought experiment). This converts the IG into the DAG for that study.
4. The directed edges from that DAG are then recorded in a directed edge index until all studies have been translated.
5. Concept Recombination is conducted throughout.

Rapid ESC-DAGs – conceptual I-DAG

6. All directed edges that involve the exposure and the outcome are added to a new DAG. This produces the conceptual integrated DAG (I-DAG) which differentiates confounders from mediators. This is also the step where rapid ESC-DAG departs from the full version, which instead integrates *all* directed edges from the index.
7. Any 'additional' concepts are added at this stage and in relation to the exposure and outcome.

Rapid ESC-DAGs – data I-DAG

8. Conceptual nodes are replaced with variable nodes or are marked as unmeasured.
9. Variable nodes are reorganised in the I-DAG's virtual space such that all cross-sectional measures are grouped together.
10. The I-DAG is forward saturated, meaning that a directed edge is drawn from all earlier variables to all later variables.
11. Each cross-sectional grouping is saturated and all relationships are Translated. Once this is completed for each cross-sectional grouping, the data I-DAG for that data source is complete.

4.6 Discussion

This chapter proposed that an evidence synthesis protocol that combines graph theory with causal thinking could meet the need for guidance on how to build DAGS. ESC-DAGs was demonstrated as a potential tool for doing so. The protocol translates individual studies into DAGs before synthesising them into one. Its systematic and transparent approach was demonstrated with hypothetical and real

examples from the literature on parental influences on adolescent alcohol use. The ESC-DAGs application in this thesis was then discussed, specifically by introducing rapid ESC-DAGs as a more efficient approach than the full version. In reference to this application and the publication in the *International Journal of Epidemiology*, it is the position of the author that methodological research objective 2 (MRO 2 - develop a method for building DAGs) has been met at this stage. This discussion considers the extent to which the full version of ESC-DAGs and rapid ESC-DAGs adhered to the principles of DAG design. The unsaturated conceptual I-DAG for this thesis is presented in Chapter 6 (page 110) and the data I-DAG in Chapter 7 (page 146). Chapter 10 evaluates the strengths and limitations of the application of rapid ESC-DAGs in this thesis, including the issue of whether it conflates temporal order with causal order (page 210). See Appendix E for a brief description of how ESC-DAGs was developed.

4.6.1 ESC-DAGs adherence to DAG principles

The full ESC-DAGs protocol meets each of the DAG principles discussed above:

1. DAGs should be built independently of data

   The DAGs are built under the assumption that each variable could be measured. Thus, concepts that are unmeasured in the data are treated in the same way as concepts that are measured. This ensures that the resulting DAG(s) systematically conceptualises unmeasured confounding. Relatedly, ESC-DAGs also makes no assumptions on the relative timing between nodes other than to use the temporal ordering causal criterion. Note however that the task of producing a data I-DAG from a conceptual I-DAG must, of course, depend on data.

2. DAGs should be built by working backwards from a saturated model

   Saturation is performed three times, and in each instance the full Translation process is applied to any directed edge that has not already been assessed. First, each IG is saturated before its directed edges are put through the Translation processes. The translated directed edges are then extracted into the directed edge index and then entered into an I-DAG, which is then saturated. Additionally, any concepts that the reviewers feel are missing from the literature may be added to the I-DAG at this stage, which should be saturated for a final time to produce the conceptual I-DAG.

3. Deleting directed edges is a much stronger assertion than including them

   The translation process is conservative, it is concerned with asking whether variable X *could* cause variable Y, thus directed edges are only deleted if the reviewers are convinced that they cannot be causal.

4. Transparency

The process uses several tools to aid its transparency. Most notably the final I-DAG is presented alongside the DAGs for the individual (and referenced) studies that were included. The directed edge index may also be included as an appendix to any application of ESC-DAGs. The decision log is a further option.

5. DAGs should be informed by the empirical literature

   The fundamental purpose for using a DAG is to identify the confounding structure for the focal relationship. ESC-DAGs subsequently focuses on the variables that are controlled for in statistical models in the literature pertaining to the research question. However, the extent to which each unique directed edge in the data I-DAG is representative of the empirical literature that the directed edge represents will vary. For example, the relationship between each confounder and the exposure is not derived from research questions focusing on that relationship, but instead from research questions that have included that relationship by factoring the confounder as a control variable. The relationships between the confounders themselves are even more removed from their respective empirical evidence bases, as directed edges between two confounders that are never modelled together in the literature can still be present in the I-DAGs. However, the crucial relationships are those between the focal variables and the covariates, which are given extensive treatment based on how they are used in the literature.

However, while the full version of ESC-DAGs meets these principles, rapid ESC-DAGs conflicts with the principle that DAGs should be built independently of data. The decision to reject this principle was purely pragmatic. Building DAGs independently of data introduces redundancies to an already labour-intensive process, especially if the aim is to produce a DAG to guide analysis of a single longitudinal dataset (as is commonly the case in health and social science and indeed in this thesis). As identified above, this redundancy is in how the relative timing of the measurements will preclude reverse causation for variables representing time varying concepts. The practical implication here is that a directed edge corresponding to two variables (nodes), one measured after the other, can only ever be absent altogether or present and uni-directional from the earlier node to the later. Thus, leveraging the timing of the measures in a longitudinal dataset has the potential to reduce the number of assessments required by 50% without compromising on the comprehensiveness of the final data - DAG(s). Arguably then, the principle to construct DAGs independently of data relies on redundancies and is overly restrictive in an applied setting. Rapid ESC-DAGs meets the rest of the DAG principles in the same way as the full version. Note however that the saturation step depends on the data.

4.6.2 Secondary aims

Each of the two secondary aims introduced above were tackled in two ways. The aim to make ESC-DAGs accessible was addressed by focusing on how the tool was articulated (using worked examples, with minimal reference to the technicalities of graph theory, etc.) and by recommending that algorithmic techniques like those utilised by DAGitty should be used to reduce cognitive load. The aim to make ESC-DAGs appealing to the wider DAG community was addressed by targeting the most appropriate academic journal and by making the approach modular. Discussing this modularity in reference to the above demonstration is helpful before moving on. First, ESC-DAGs is a discrete tool in and of itself – there is no technical need to make changes to it if the goal is to systematically and efficiently develop a DAG. However, it was recognised early in the design process that whatever the end-product of this chapter might be, it would not become the authority on how to build DAGs (indeed, such a thing may never exist). As such, it became a priority to ensure that interested researchers could use ESC-DAGs from start to finish, and that more sceptical researchers could use it in a flexible way, specifically by allowing them to swap out processes they found to be unhelpful for preferred alternatives. For example, the ROBINS-I and upcoming ROBINS-E tools, developed by some key figures in causal inference, focus on assessing risk of bias and the plausibility of the relationships in non-randomised studies of interventions and observational studies of exposures respectively (Bero et al., 2018, Sterne et al., 2016). These tools could feasibly be generalised to assess the viability of directed edges in DAGs, and thus used in place of the Translation processes (although it should be noted that the ESC-DAGs approach is currently much faster). A related concern was that the ESC-DAGs approach could be expanded upon. For example, currently the degree of certainty that reviewers have on any given directed edge is not systematised. A pre-existing risk of bias tool could be integrated for this purpose, for example (see Discussion in Chapter 10).

Chapter 5 Literature review: Parental influences on adolescent alcohol harm

This literature review is concerned with the substantive goal of summarising and critically appraising the evidence on parental influences on adolescent alcohol harm. The relevant research objectives are SRO 1 and 2 (respectively: determine the main parental influences on adolescent alcohol harm from the literature; and what does the evidence suggest about the relationship between these influences and adolescent alcohol harm). As such this literature review also shapes the ESC-DAGs review in the following chapter and thus pertains to SRO 3 (build DAGs of parental influences on adolescent alcohol harm).

5.1 Review of systematic reviews

A review of reviews approach was selected to appraise the evidence for parental influences on adolescent alcohol harm. Unlike a systematic review, the aim of a review of reviews is not to provide a definitive summary of evidence for a topic. Rather, by exploiting pre-existing search, appraisal, and synthesis protocols in the form of original reviews, a review of reviews seeks to arrive at a summary which is as comparable to a systematic review as possible while maximising time and resource efficiency (Noble and Smith, 2018, Smith et al., 2011, Remes et al., 2016, Grant and Booth, 2009). This is achievable because original reviews (especially systematic reviews and meta-analyses) are designed to be comprehensive and critical summaries of evidence on their topics. Indeed, they are often argued to represent the pinnacle of research evidence (Hanley et al., 2016, Murad et al., 2016). A review of reviews is thus an appealing option for this literature review, as the emphasis on

efficiency helps balance the substantive aims of this chapter with the subsequent ESC-DAGs review, in that the search and quality appraisal strategies of pre-existing reviews can be leveraged to identify a wide set of original studies to act as the raw material for the ESC-DAGs processes.

Additionally, as a review of reviews is understood to be especially useful when the research question is broad and there are multiple reviews available (Noble and Smith, 2018, Grant and Booth, 2009), the subject matter of parental influences on adolescent alcohol use is wholly compatible. As indicated by SRO 1 (determine the main parental influences on adolescent alcohol harm from the literature), this thesis takes a broad view of parental influences on adolescent alcohol harm. Further, a scoping search of reviews on parental influences on adolescent alcohol harm indicated that the number of returns was similar to other reviews of reviews (Maniglio, 2010, Marchi et al., 2015, Maniglio, 2009). A review of reviews approach was thus determined as suitable for summarising and appraising the evidence on parental influences on adolescent alcohol harm for this thesis. This chapter covers the search strategy and inclusion/exclusion criteria (Section 5.2), data extraction for the selected systematic review (Section 5.3.1); and data synthesis (Section 5.3 and 5.4). See also Appendix F for a detailed example of how literature searches were performed.

5.2 Search strategy

The approach taken here was to review only systematic reviews, due to their position in the hierarchy of evidence and because their emphasis on systematic search protocols is ideal for the subsequent ESC-DAGs review (and superior to other forms of review). Search terms looked for reviews, systematic reviews, and meta-analysis across three databases (PubMed, Web of Science, and PsychInfo). Appendix F demonstrates the search as performed in PubMed. It is possible to restrict the definition of systematic reviews to only the highest quality examples. For instance, Remes et al (Remes et al., 2016) only included systematic reviews meeting the Cochrane Collaboration definition (Higgins et al., 2011); that used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRSIMA) template; and that met a minimal score on the 'assessment of multiple systematic reviews' (AMSTAR) quality appraisal tool (Shea et al., 2007b, Moher et al., 2009). However, rather than exclude any systematic review, original quality appraisal was conducted on each so as to maximise the number of systematic reviews while still preferring higher quality research for evidence synthesis.

5.2.1 Search terms

The search strategy was further shaped using the Population Exposure Outcome (PEO) guidelines (Richardson et al., 1995) (an adaptation of the PICO guidelines (Methley et al., 2014) recommended by the Cochrane Collaboration (Higgins and Green, 2011)), PRISMA (Moher et al., 2009), and AMSTAR 2 tool (Shea et al., 2007b, Shea et al., 2017). Adolescents acted as the population; parental influences as the exposure; and alcohol harm as the outcome. Given the restriction to systematic reviews, the MeSH vocabulary (Medical Subject Headings) was used to methodically expand on the PEO search terms so as to lend breadth to the returns across databases, as below (Lowe and Barnett, 1994):

- Population: adolescent* OR youth OR student OR teen* OR underage OR minor
- Exposure: parent-child relations OR parental influences OR parenting OR parent*
- Outcome: alcoholism OR underage drinking OR risky drinking OR excessive drinking OR binge drinking OR alcohol use OR alcohol consumption OR drink*

A final restriction was placed on the design of the studies under review by limiting inclusion criteria to systematic reviews of longitudinal studies. The main reason for doing so is the limited capacity of cross-sectional studies to comment on causation, despite their value for indicating associations. Specifically, cross-sectional studies cannot account for reverse causation between time-varying variables. Moreover, more recent reviews of parental influences on adolescent alcohol harm are notably critical of previous reviews for including cross-sectional studies (Visser et al., 2012, Ryan et al., 2010). Thus, given the causal inference emphasis of this thesis, searches were restricted to systematic reviews of longitudinal, prospective, or cohort studies.

5.3 Results

Six systematic reviews were identified, pertaining to 183 unique primary studies. Three databases were searched: Web of Science (25 returns); PubMed (28 returns); and PsychInfo (18 returns). Searches were finalised on the 6th April 2018. Figure 5-1 below is a simplified version of the PRISMA diagram used for systematic reviews and meta-analyses. It summarises how the screening processes identified the six systematic reviews. A total of 42 non-duplicated records were excluded after screening abstracts and full texts because they did not meet the PEO search strategy (e.g. looked at alcohol harm in pre-adolescence; non-parental exposures such as peers; conflated alcohol use and

illicit drug use), or were other forms of reviews (e.g. narrative reviews or 'overviews'). See Appendix F for more detail.

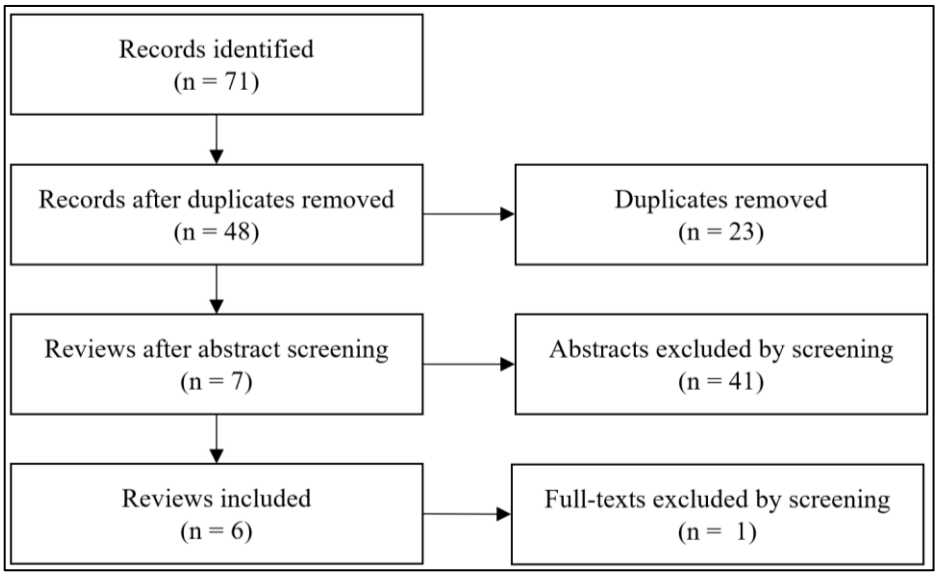


Figure 5-1: Review of systematic reviews flow diagram

5.3.1 Characteristics of included systematic reviews

Data extraction for the six systematic reviews is summarised in Table 5-1 below and is continued narratively throughout this subsection. All were published between 2010 and 2017, with publication dates for included research articles (henceforth; "primary studies") beginning in 1986. Two looked at parental influences in a general sense (Ryan et al., 2010; Yap et al at., 2017). The others looked at more specific elements: the parent-child relationship (Visser et al., 2012); parental drinking (Rossow et al., 2016); parental supply of alcohol (Sharmin et al (a)., 2017); and parental rules about their adolescent's alcohol use (Sharmin et al (b)., 2017). Each reviewed the international literature, but each was restricted to primary studies written in English. Rossow et al's work on parental drinking was a collaboration between Norwegian and UK institutions. Visser et al's review of the parent-child relationship was conducted in the Netherlands. The rest were carried out across several Australian institutions.

Each review detailed its search protocols (including how the research question was devised, the databases and selection criteria used to identify primary studies, as well as the data extraction methods) and used the PRISMA reporting guidelines. Each also augmented its searches through scrutiny of forward and backwards citations. Three systematic reviews featured meta-analysis (Yap et

al's systematic review of general parental influences, and both of Sharmin et al's systematic reviews of parental supply of alcohol and of parental rules about adolescent alcohol use).

Unsurprisingly, the broader the definition of parental influences, the higher the number of primary studies reviewed. The narrower investigations returned between 13 and 28 primary studies, while the broader definitions returned 77 (Ryan et al) and 131 (Yap et al). Of the 183 primary studies, 86 (47%) were unique to the systematic review they were included in. Yap et al contributed 53; Ryan et al contributed 13; Sharmin et al (b) contributed 5, as did Visser et al and Rossow et al; while Sharmin et al (a) contributed 0. The remaining 97 (53%) were included in at least two systematic reviews, 78 of which were included in the Yap et al review. The total number of studies included in meta-analysis after removing duplicates was 53 (1 unique to Sharmin et al (a), 4 unique to Sharmin et al (b), 34 unique to Yap et al, and 14 featured in more than one meta-analysis).

Table 5-1: Characteristics of included reviews

Author & Year	Location	# studies reviewed	# studies in quantitative synthesis	Parental influence investigated	Quality of reporting (AMSTAR 2)
Ryan et al 2010	Australia	77	0	Parenting factors	Low
Visser et al 2012	Netherlands	28	0	Parent-child relationship	Moderate
Rossow et al 2016	Norway/UK	21	0	Parental drinking	Moderate
Sharmin et al (a) 2017	Australia	13	6	Parental supply of alcohol	Moderate
Sharmin et al (b) 2017	Australia	27	13	Parental alcohol rules	Moderate
Yap et al 2017	Australia	131	48	Modifiable parenting factors	High

Numerous different prospective cohorts were utilised, including twin, birth, and parent-child cohorts. US-based studies were the most common, but Western European and Australian counterparts were strongly represented (for example: 89 of the studies that were unique to Yap et al's review were conducted in the US; 3 in Canada; 31 in Europe; 6 in Oceania; and 2 comparative studies of the US and Australia). There was broad variety in outcome measures, including: ever having tried alcohol; age of alcohol initiation; frequency of use; frequency of intoxication; alcohol dependence; alcohol use disorders; etc. There was a similarly broad range of statistical methods used ranging from simple multiple regression approaches (Seljamo et al., 2006) to more advanced multiple mediation analysis (Latendresse et al., 2008). Potential outcomes and DAGs were notably absent from all primary studies, indicating that the analysis conducted in this thesis addresses a gap in the literature.

5.3.1.1 Methods for synthesis

Three of the reviews used meta-analysis (Yap et al and both Sharmin et al). The Sharmin et al reviews compared the odds of 'risky drinking' after their relative exposures (parental supply of alcohol and parental alcohol rules). Yap et al focused on measuring the amount of variation associated with each risk factor for two separate outcomes: the age of alcohol initiation; and later alcohol misuse. Ryan et al used the same pair of outcomes but did not use full-fledged meta-analysis. Instead they employed Stouffer's P to test whether relationships were significant across their primary studies. Rossow et al and Visser et al relied on narrative synthesis. Visser et al used a 'best evidence synthesis approach', while Rossow et al developed 'causal criteria' designed to indicate how much scope primary studies had for causal inference. Despite the lack of quantification in both of these reviews, or perhaps because of it, their assessment of their findings was much more nuanced and less reliant on simple description. This is reflected in the divergent interpretations of the similar literature between them and the meta-analytical reviews as discussed below.

5.3.1.2 Measures of alcohol harm

The numerous ways to operationalise alcohol use and harm present a key challenge for systematic reviews, especially for those with a meta-analytic component. Both Sharmin et al reviews focused on 'risky drinking', and despite stipulating the technical definition of risky drinking, (consuming more than 5 units in one sitting), they nonetheless explicitly acknowledge that they used the term to subsume the numerous other concepts that the primary studies employed, including "alcohol misuse, drunkenness, alcohol dependence, heavy drinking, binge drinking, intoxication, heavy episodic drinking" (Sharmin et al., 2017a, Sharmin et al., 2017b). Clearly these concepts are highly related, if not equivalent. Ryan et al and Yap et al both looked at the same two outcomes individually: early initiation of unsupervised alcohol use, and later misuse. Later misuse was employed as a generic umbrella term, similar to both Sharmin et al reviews' use of 'risky drinking'. Rossow et al used the concept of an 'alcohol-related outcome'. This is a less specific term which conflates the mode of consumption with the consequence of consumption. Visser et al simply let the concept of alcohol use cover all aspects of alcohol use and harm.

5.3.1.3 Categories of parental influences

SRO 1 is specifically concerned with determining the main parental influences on adolescent alcohol harm that are investigated in the literature. The review of reviews found a total of 12 different parental influences. Yap et al and Ryan et al investigated parenting factors in a general sense. Ryan et al developed a typology of parental influences "based on definitions most commonly used in the

literature". The Yap et al review was conducted seven years later, had two of the same co-authors, and used the same typology with some 'slight modification' based on new evidence. The categories used in both reviews were parental alcohol use; parental modelling; favourable attitudes towards alcohol; parental involvement; parent-child relationship quality; parental support; parental communication; parent-child alcohol specific communication; parental provision of alcohol; parental monitoring; parental discipline; parental alcohol rules; and family conflict. As such they subsume the categories of parental influences that are investigated by the other four systematic reviews: parent-child relationship (Visser et al); parental drinking (Rossow et al); parental supply of alcohol (Sharmin et al (a)); and parental alcohol rules (Sharmin et al (b)). Table 5-2 (below) summarises the parental influences investigated across the reviews, as well as the support found for each.

5.3.1.4 Quality appraisal and bias assessment of primary studies

Ryan et al did not conduct any quality or bias assessment. Visser et al used a bias assessment tool designed by Hayden, Cote & Bombardier (Hayden et al., 2006). It assesses studies across six domains: study participation; study attrition; predictor measurement; outcome measurement; confounding measurement; and analysis. They found that nine of 28 studies (32%) were unbiased, ten (36%) had one bias, and the remaining nine studies had two or more biases. They noted attrition rates to be particularly problematic. Rossow et al's causal criteria has a specific emphasis on bias in study design. They determined that sparse use of theory-driven analyses and inadequate control for confounding are common biases and that there was a tendency towards smaller data sets in some studies. They concluded that only four of the 21 studies included in their review (19%) were sufficiently unbiased to possess scope for causal inference.

Sharmin et al (b) used the Newcastle-Ottowa Quality Assessment Scale and found that seven primary studies were at high risk of bias, one study was at moderate risk, and two were at low risk. They noted attrition rates, reliance on self-report for the outcome, and inadequate control for confounding as the more common reasons for studies being at higher risk of bias. Four of the studies they included were trials and were assessed under the Cochrane risk of bias tool. Two were found to be at high risk of bias, while risk in the other two was unclear due to opaque reporting of study protocols. They also conducted publication bias analysis using funnel plots and found no evidence across the studies included in their review. Sharmin et al (a) also used the Newcastle-Ottowa Quality Assessment Scale and found three studies to be at low risk of bias, one to be at moderate risk, and three to be at high risk. Again, they noted the lack of confounder adjustment and reliance on self-reports to be key issues. They used funnel plots once more to assess publication bias. They noted some indication of publication bias but that results were inconclusive due the low number of primary studies.

Yap et al used the same bias assessment tool as Visser et al. They found that 79 studies (60%) did not have any biases, 35 had biases in one domain, and that the 17 remaining studies had two or more biases. Similar to Visser et al, they noted attrition rates as being the most common (81% of studies with at least one bias). Using funnel plots augmented with a trim and fill procedure, they concluded that there may be evidence of publication bias for favourable attitudes to alcohol; parental alcohol use; family conflict; general communication; parent-child relationship quality; parental discipline; and parental involvement.

5.3.1.5 Quality appraisal of included systematic reviews

To ensure that the systematic reviews as a whole were of acceptable quality they were assessed using the freely available and widely used AMSTAR 2 tool (Shea et al., 2017, Shea et al., 2007b, Shea et al., 2007a, Shea et al., 2009, Faggion, 2015). Various elements of systematic review design are covered by AMSTAR 2, such as using search strategies informed by PICO/PEO; the review's approach to bias assessment; whether meta-analysis was used; whether more than one reviewer was involved in screening and data extraction; whether a protocol for the review was registered prior to it being conducted; etc. Quality varied from low (one review, Ryan et al) to high (one review, Yap et al). The remaining four reviews were of 'moderate' quality (Visser et al; Rossow et al; Sharmin et al (a) and Sharmin et al (b)). However, while the tool certainly gives an indication of which reviews employed the most rigorous protocols, more recent reviews featuring meta-analysis automatically scored higher. The Ryan et al 'low' score was not considered problematic given the degree of overlap with the later Yap et al review. In fact, as discussed below, the Ryan et al and Yap et al reviews have highly similar conclusions despite this apparent discrepancy in quality. As such, none of the reviews were excluded from this review of reviews.

Table 5-2: Summary of meta-analyses of parental influences

Parental Influence	Effect	Quantitative synthesis			Qualitative synthesis[4]		
		% variation in initiation[1]	% variation in misuse[2]	OR of risky drinking[3]	Strong	Moderate	Weak
Alcohol use	Harmful	1	2		2	1	
Provision of alcohol	Harmful	4	7	2.0	2		
Favourable alcohol attitudes	Harmful	1	3		2		
Family conflict	Harmful	1				1	
Parent-child relationship quality	Protective	1	1		1		1
Parental involvement	Protective	1	1		1		
Parental support	Protective	1	1		1		
Parental monitoring	Protective	3	5		1		
Parental rules about alcohol	Protective	4		0.64		2	
Parental discipline	Protective	<1				1	
Parent-child communication	Inconclusive						1
Alcohol-specific communication	Inconclusive						

[1]Meta-analysis by Yap et al, 2017 on the percentage of variation on age of first unsupervised alcohol use accounted for by the parental influence.

[2]Meta-analysis by Yap et al, 2017 on the percentage of variation on later alcohol misuse accounted for by the parental influence.

[3]Meta-analysis by Sharmin et al 2017 (a) and Sharmin et al 2017 (b) of the odds ratio of risky drinking.

[4]Of the six systematic reviews, how many reported that the evidence for each parental influence was strong, moderate or weak.

5.3.2 Summary of evidence

SRO 2 is concerned with appraising the evidence for the parental influences identified by SRO 1 above. Table 5-2 (above) summarises the direction of effect, effect size, and the strength of the evidence for each parental influence.

Answering this research objective was more complicated as there are divergences in conclusions between systematic reviews. Firstly, Visser et al concluded there to be "weak" evidence for the parent-child relationship's influence adolescent alcohol harm, while Yap et al (five years later) found the evidence to be 'sound' and 'convergent'. This can be attributed to the relative definitions of the parent-child relationship in the two reviews. Yap et al contended that the 'global' definition of the parent-child relationship used by Visser et al "may have obscured significant associations", and so they deconstructed the concept of the parent-child relationship into constituents. Yap et al also benefited from more recent research.

Secondly, Yap et al and Ryan et al both considered the evidence for parental alcohol use as a harmful risk factor to be strong, while Rossow et al were much more ambivalent. This divergence can be explained by the framing of the reviews. Rossow et al were particularly focused on narrative synthesis emphasising the scope each primary study had for causal inference, while Yap et al were more concerned with meta-analysis. However, Rossow et al did conclude that, while only four of their 21 primary studies met their criteria for causal inference, there was consensus across these studies that parental alcohol use is a harmful risk factor (indeed this consensus was across all studies they reviewed, regardless of their scope for causal inference).

Finally, Ryan et al found that general communication between parents and children was a significant factor, while Yap et al found no evidence for a relationship. Yap et al commented directly on this disagreement, and argued that it was due to how, in their view, research focuses on the quality of communication while neglecting content. In other words, communication may be good but the content may be harmful or protective.

5.3.2.1 Harmful parental influences

Parental provision of alcohol (Yap et al, Ryan et al and Sharmin et al (a)), favourable attitudes towards alcohol (Yap et al), parental alcohol use (Yap et al, Ryan et al and Rossow et al), and family conflict (Yap et al) were all determined as harmful influences on adolescent alcohol harm – i.e. those that parents should aim to reduce. Yap et al found that parental provision was the risk factor with the largest effect, accounting for 7% of variation for adolescent alcohol misuse (favourable attitudes,

parental alcohol use, and family conflict accounted for 2%, 3% and 1% respectively). Sharmin et al's meta-analysis of parental supply found that the odds ratio for risky drinking was around 2 for adolescents whose parents provided alcohol compared to those who did not.

5.3.2.2 Protective parental influences

Parental monitoring (Yap et al and Ryan et al), high quality of parent-child relationship (Yap et al, Ryan et al, and Visser et al), parental support (Yap et al and Ryan et al), parental involvement (Yap et al and Ryan et al), clear rules against adolescent alcohol use (Yap et al, Ryan et al, and Sharmin et al (b)), and parental discipline (Yap et al and Ryan et al) were all found to be protective factors – i.e. those that parents should aim to increase. Yap et al's meta-analysis found that parental monitoring and clear rules against alcohol use had the highest associations with the outcomes and accounted for 3% and 4% of alcohol initiation respectively, and that and parental monitoring also accounted for 5% of later misuse. The other protective factors each accounted for around 1% across both outcomes (except parental discipline which was only investigated on age of initiation). Meta-analysis by Sharmin et al (b) found the odds ratio for risky drinking to be 0.64 for adolescents whose parents set strict alcohol rules compared to those whose parents do not.

5.3.2.3 Parental influences with no relationship

Both Ryan et al and Yap et al found that alcohol-specific communication was not associated with either initiation or later misuse. As noted above Ryan et al and Yap et al diverged on their conclusions on general parent-child communication, although the latter's meta-analysis found no relationship for either outcome.

5.3.2.4 Moderators of parental influences on adolescent alcohol harm

Only Yap et al used any formal assessment of moderation. They found support for follow-up time as a moderator of the effect of parental attitudes to alcohol, family conflict, parental support, parental involvement, and the parent-child relationship. Gender of the adolescent was found to moderate the effect of parental monitoring and the parent-child relationship, and gender of the parent was found to moderate the effect of parental alcohol use, parental attitudes to alcohol, parental support, and parental discipline. This equates to parental influences on adolescent alcohol use being moderated by age and the gender of the parents and their offspring.

5.3.2.5 Mediators of parental influences on adolescent alcohol harm

Across the systematic reviews, some support was found for a small number of mediating mechanisms. Most pertained to mediators of parental drinking and indirect effects were found for parenting (Latendresse et al., 2008), alcohol specific communication (Mares et al., 2011), and alcohol-related memory of the adolescent (Van Der Vorst et al., 2013) while poor adolescent control of inhibition (Pears et al., 2007) and parenting were also found *not* to mediate at all (Van der Zwaluw et al., 2008). Other mediators of parental influences included adolescent self-control as a mediator of alcohol-specific parenting (Koning et al., 2014), adolescent other-sex friendships as a mediator of parental monitoring (Poulin and Denault, 2012), and parental supervision of adolescent drinking as a mediator of parents' favourable attitudes towards alcohol (McMorris et al., 2011). Thus, there is a dearth of literature on mediation of parental influences as a whole. Similar to the above comment on the lack of DAG-based and potential outcomes analysis, causal mediation analysis such as that used in Chapter 9 of this thesis may benefit alcohol studies.

5.4 Discussion

This discussion considers which parental influences should be investigated using the ESC-DAGs processes; the gaps in the literature (emphasising those pertaining to the methodological approach of this thesis); and the scope for causal inference. Causal inference is considered narratively in terms of the consistency in findings across studies, strength of associations, and theory. This is loosely informed by the GRADE approach to the Bradford Hill viewpoints (Schünemann et al., 2011) and Rossow et al's narrative synthesis approach to causal inference (Rossow et al., 2016).

5.4.1 Consistency of findings

Bradford Hill stated that differentiating chance from causal effect can be achieved "only by a repetition of the circumstances and the observations" (Hill, 1965, Susser, 1986). Researchers and policy makers interested in using research evidence to inform further research or intervention design depend on consistent findings and it is on this basis that systematic review methodologies were developed (Higgins and Green, 2011). Consistency across studies and data is of course crucial from a non-causal perspective as well – a pertinent example for a reviews approach would be the GRADE guidelines, which suggest "downgrading the quality of evidence" when studies diverge on their conclusions (Schünemann et al., 2011).

The ‘ideal’ consistent relationship would be an effect of the same direction that is statistically significant across all studies, and that has relatively stable effect size. When trying to establish consistency then, variation in the direction of effect is more consequential than variation in detecting significant results. As such, a body of results in which the direction of effect is homogeneous (even in non-significant results), is more consistent than a body of results in which the direction of effect varies between harmful and protective. Of course, direction of effect and statistical significance should not be considered reductively; a contrary result which is statistically significant should be considered as more consequential than a contrary result that is insignificant. This is of course further complicated by statistical power.

Visser et al reviewed the effect of the parent-child relationship and found that that less than 75% of their studies agreed on the existence and the direction of effect between the parent-child relationship and frequency/quantity of adolescent alcohol use. However, all disagreement was in terms of statistical significance - they found no support for poor quality of parent-child relationship being a protective factor or for high quality of parent-child relationship being a harmful factor. In their review of the effect of parental alcohol rules on adolescent risky drinking, Sharmin et al (b) found that, across 17 analyses in 13 primary studies, only three coefficients suggested a protective relationship, and for only one of these was the lower bound of the confidence interval above 1.0 (i.e. statistically significant). Rossow et al failed to find any primary study that suggested a protective effect of parental drinking on adolescent drinking, and similarly Sharmin et al (a) found no support for parental supply of alcohol being protective.

Across the 12 parental influences they investigated, Ryan et al found mixed results for general parent-child communication, alcohol specific communication, parental discipline, and alcohol specific rules. Parental discipline and alcohol specific rules were equivocal across the primary studies and the authors concluded that there was only weak evidence for general communication being protective on age of alcohol initiation. For parental discipline, one study found that there is a non-linear relationship between the degree of alcohol specific rules and adolescent alcohol outcomes, in that “being overly lax or overly strict” was associated with increased binge-drinking (Guilamo-Ramos et al., 2005, Ryan et al., 2010). This was one of nine studies, three of the others finding no effect, while the other five found a protective effect. Accordingly, parental discipline may be considered as homogeneous in the Ryan et al review. For parental alcohol rules, one primary study was found that suggested strict parental rules were associated with greater heavy drinking in later adolescence. This was one of five studies - only one of the others found a protective effect for later alcohol use. However, when the outcome was age of initiation, results were consistently protective, hence Ryan et al concluded alcohol rules to be protective.

Yap et al did not explicitly report on individual study results, but their meta-analysis was equivocal for parent-child communication and alcohol specific communication, and the broadest confidence intervals over the other 10 parental influences they investigated were for parental discipline and alcohol rules. As such, their findings on general parent-child communication, alcohol specific communication, parental discipline, and alcohol specific rules were largely in line with Ryan et al.

In summary, of the 12 potential parental influences under investigation across all reviews, two were so inconsistent as to be inconclusive (general communication and alcohol-specific communication across two separate systematic reviews). Two had a minority of contrary findings (parent-child relationship and parental alcohol rules). Across the other nine parental influences, no primary studies found contrary results although, for family conflict, as many studies reported no relationship as detected a harmful effect. Therefore, the parental influences with the most consistent findings were parental alcohol use, parental provision of alcohol, parental attitudes to alcohol, and aspects of the parent-child relationships and parenting strategies. As such, the evidence base for parental influences of adolescent alcohol harm is consistent in general, with few instances of contrary results and only a minority of studies reporting no relationship. These effects tended to indicate that problematic parental influences (more drinking, poor parent-child relationship, harsh parenting strategies) had harmful effects, perhaps indicating that overall family dysfunction is an important determinant, although no studies used this construct.

5.4.2 Strength of association

According to Bradford Hill, strength of association is a causal viewpoint in its own right (Hill, 1965). In the most technical sense however, this view has come under criticism because not all component causes of an effect need to have a high strength of association (Rothman and Greenland, 2005b, Rothman and Greenland, 2004). Indeed, across all the systematic reviews, most effect sizes were described as small. However, Yap et al's meta-analysis suggests that, in total, parental influences account for around 18% of variation in age of initiation, and around 20% of variation in later misuse. Considered against the array of alternative factors across the socio-ecological model (genetic, peer, school, neighbourhood/community) this is not an insubstantial effect. The exceptions were parental monitoring, parental supply of alcohol, and to a lesser degree, parental alcohol use – each of which was reported as of moderate effect. These three parental influences were also among the most consistent.

5.4.3 Theory

If explicitly theory-driven analysis is taken as a crucial indicator of causal inference, as suggested by Rossow et al, then the empirical evidence base is weak in this regard. Indeed, the paucity of mediation analysis discussed above indicates as much. While the evidence base appears consistent in terms of the effects produced, it appears to be less successful at explaining these effects. However, the explanation for the lack of theory is as likely to lie in the nature of the peer review process as anything else. For example, in Yap et al's review, they dedicate one paragraph of the entirety of their article to discussing the theoretical role of parents in the socialisation process. The reason why, of course, is that their purpose was publication of a specialist meta-analytical summary, not theoretical debate. A similar argument may be made in defence of the primary studies.

5.4.4 Risk of bias

Additional to low effect sizes and lack of theory in empirical papers, bias in study design was a concern. Except for the Ryan et al review, which did not conduct any formal bias assessment, all systematic reviews reported a moderate level of bias across their primary studies on average. Each considered various dimensions such as sample bias, measurement bias, and model misspecification. Inadequate control for confounders (model misspecification) was ubiquitous. As were problematically high attrition rates (sample bias). Measurement bias mainly pertained to the heterogeneity of outcome measures.

A notable trend across the reviews was ambiguity in how the results of bias assessments of primary studies informed review-level evidence synthesis. For example, none of the reviews that used meta-analysis (Yap et al and both Sharmin et al reviews) applied any kind of weighting mechanism to emphasise higher quality studies, and only Sharmin(a) performed sensitivity analysis investigating the potential impact of bias (they found results were robust to removing weaker studies). Similarly, Visser et al's narrative synthesis carried out an extensive risk of bias assessment, but relegated its results to caveats on their conclusions. Rossow et al, with their explicit emphasis on causal inference, were the exception to this trend. They did however, uncover consensus on the harmful effect of parental alcohol use between the studies they considered to be of no risk of bias and those of high risk.

Yap et al found publication bias across several factors: favourable attitudes to alcohol; parental alcohol use; family conflict; general communication; parent-child relationship quality; parental

discipline; and parental involvement. Publication bias (i.e. publication of positive results to the detriment of negative results) of course reduces confidence in conclusions across the literature.

5.4.5 Strengths & limitations

To the author's knowledge, this review of systematic reviews is the first of its kind which summarises the evidence for parental influences on adolescent alcohol harm. The main strength is how the review of reviews approach identified a large number of primary studies (n = 183) using longitudinal data. The findings at the review level have been summarised and discussed critically, and 12 parental influences have been identified. The explicit causal inference emphasis of this thesis manifested with a strong discussion on the conceptualisation, measurement, and treatment of bias in study design. Most notably, inadequate confounder control was identified as a ubiquitous problem for which there are numerous solutions, including ESC-DAGs. Further, this review of systematic reviews achieved an additional objective of providing an informed entry point to the literature for the subsequent ESC-DAGs review below.

There are limitations to a review of reviews approach in general. Firstly, the emphasis on efficiency means they are less robust and less comprehensive than an original systematic review. Indeed, there are few guidelines available for how to conduct a review of reviews. There is also the concern of multiple levels of heterogeneity in study design, both across primary studies, and across the systematic reviews themselves. However, this was not problematic in this case given the level of consensus across systematic reviews and primary studies.

The current review of systematic reviews has its own limitations. Probably the most notable was the relatively low number of systematic reviews at six, especially in how only three of the six conducted meta-analysis. While there are no strict guidelines on how many reviews should be included for any given research question, the conclusions of this review of reviews are perhaps overly reliant on the Yap et al systematic review and meta-analysis. They reviewed 131 of the primary studies (72% of total), including 53 that were only covered in their review (29% of total). These limitations were not deemed to be critical for two reasons: the AMSTAR 2 tool concluded that the Yap et al systematic review was the highest quality of the six; and the fairly large number of primary studies (183) was deemed sufficient to compensate for the relatively low number of systematic reviews. However, the lowest quality systematic review (Ryan et al) according to the AMSTAR 2 tool had almost identical conclusions to the highest quality study (Yap et al). While conducting an original meta-analysis may have improved the certainty of the findings from this systematic review of reviews, the subsequent

ESC-DAGs review is effectively a novel form of qualitative synthesis, and as the main novel contribution of this thesis, it was preferred over original quantitative synthesis.

There were notable limitations across the systematic reviews also. Firstly, despite each systematic review identifying a lack of control for confounding factors, none devoted anything more than minimal consideration to what the missing confounders might be, and so what they considered 'adequate' control of confounding was mostly unclear. Secondly, while the reviews generally engaged in extensive bias assessment, including publication bias, these assessments were not operationalised fully in any of the evidence synthesis. Relatedly, none of the systematic reviews attempted to explain how such biases would need to behave in order to invalidate the results. For example, Rossow et al noted there to be low scope for causal inference, despite consistent findings. Without such consideration, their conclusions appear overly cautious. Overall, there was a tendency towards a using a plurality of sophisticated evidence synthesis methods that were not then integrated in any meaningful way – indicating that the full value of these disparate methods are not being realised in research.

5.5 Conclusion

This review of systematic reviews found fairly consistent support for harmful and protective parental influences of non-negligible effect sizes across six systematic reviews, indicating that parents have various influences on the alcohol use and harm of their offspring during adolescence. Effect sizes for individual parental influences ranged from modest (Yap et al found that parental discipline accounted for less than 1% of variation in alcohol initiation) to fairly high (Yap et al found parental supply of alcohol accounted for 7% of variation in later alcohol misuse and Sharmin et al (a) found that adolescents whose parents have supplied alcohol to their offspring are twice as likely to engage in later risky drinking). Meta-analysis by Yap et al suggests that parental influences account for around 18% and 20% of variation in adolescent age of initiation and later alcohol misuse respectively. There was a very strong tendency for protective factors to be consistently protective across the literature, and for harmful factors to be consistently harmful across the literature. In summary, there is strong support that parental influences are an important determinant of adolescent alcohol harm. There was a strong indication that DAG-based analysis using counterfactual causal inference can help address real gaps in the literature on parental influences on adolescent alcohol harm. Most notably, the systematic reviews drew attention to a dearth of theory driven analysis and adequate control for confounding. As noted in Chapter 4, ESC-DAGs aims to access theory indirectly via the empirical literature and articulate it in DAG form.

This review of systematic reviews was thus used to answer SRO 1 and SRO 2 of this thesis. SRO 1 was concerned with determining what parental influences on adolescent alcohol harm are commonly investigated in the literature. The review of reviews found 12 parental influences in total, approximately covering the conceptual domains of parental alcohol behaviours, parenting strategies, the parent-child relationship, and parent-child communication. SRO 2 was concerned with appraising the evidence on these influences. It found at least some supporting evidence for all but one of the 12 parental influences (alcohol specific communication). There was strong evidence for seven of the others, including parental drinking; parental supply of alcohol; the parent-child relationship; parental support; parental involvement; parental monitoring; and favourable attitudes towards alcohol. The evidence from the review of reviews can be summarised with the following statements:

- Parental provision of alcohol, parental drinking, favourable parental attitudes towards alcohol, and family conflict are parental influences that have harmful associations with adolescent alcohol harm.
- The parent-child relationship, parental support, parental involvement, parental alcohol rules, and parental discipline are parental influences that have protective associations with adolescent alcohol harm.
- Parental alcohol use, parental supply of alcohol, and parental monitoring were notably consistent and of higher effect size.

Chapter 6 ESC-DAGs review 1: Conceptual I-DAG

The rapid ESC-DAGs review in this thesis is divided across two chapters. This chapter corresponds to the portion of ESC-DAGs that produces the conceptual I-DAG while the subsequent chapter is concerned with replacing these concepts with variables from ALSPAC. Thus, the current chapter covers Mapping the primary studies into IGs, converting these IGs into DAGs using the Translation process, and integrating the key relationships in these DAGs using the Synthesis process to produce the conceptual I-DAG. In other words, causes of adolescent alcohol harm, on a conceptual level, are gathered and integrated into a single DAG. This current chapter is thus mainly concerned with MRO 3 (apply and demonstrate the method) and SRO 3 (based on the evidence and using ESC-DAGs, build DAGs of parental influences on adolescent alcohol harm). It has been structured such that the reporting of results demonstrates the key stages and outputs of rapid ESC-DAGs (MRO 3) and culminates by presenting the conceptual I-DAG to be taken forwards to the data (SRO 3).

6.1 Study selection

A secondary reason for using the review of reviews approach in Chapter 5 was to leverage the most comprehensive search protocols available on the literature on parental influences on adolescent alcohol harm. The systematic reviews thus act as efficient entry points to the literature for the ESC-DAGs review. However, within the wider context of this thesis and in consideration of how detailed the evidence synthesis processes in ESC-DAGs are, the 183 studies identified across the systematic reviews were deemed too many to be workable. Accordingly, in order to reduce this amount, two additional inclusion criteria were implemented for study selection into the ESC-DAGs processes. These inclusion criteria additionally seek to make the participants contributing data to the ESC-DAGs processes more similar to the ALSPAC participants.

Firstly, all studies published before 2005 were excluded (N=65). This increased the likelihood that the studies would focus on participants from the same generation as the ALSPAC cohort, the oldest of which were turning 15 in 2005 (one year before outcome measurement). It also placed an emphasis on more recent research. Secondly, as discussed briefly in Chapter 1, one of the most crucial policies affecting adolescent drinking is legal age of purchase (Room et al., 2005, Wagenaar and Toomey, 2002). Studies in settings with a different legal age of purchase to the UK (e.g. the USA) were excluded (N=70) to try and ensure that the availability of alcohol at a policy level was similar for participants across the primary studies and ALSPAC. Canada was included despite some provinces having a legal age of purchase of 19 (Canadian Centre on Substance Abuse, 2017). Studies in Low- and Middle-Income Countries were also excluded, although this pertained to only one published after 2005 (Hung et al., 2009).

In effect, studies included in this ESC-DAGs review were those published after 2004 and were based in Australia (N=9); Canada (N=3); Finland (N=4); Germany (N=2); the Netherlands (N=18); Norway (N=1); Sweden (N=6); and the UK (N=5). Of the 183 primary studies included in the review of systematic reviews, 48 (26%) met these inclusion criteria. One was later excluded for focusing on alcohol outcomes in children under 10 years old. Forward citations of the systematic reviews were scrutinised in April 2018 and 3 studies (Australia; Sweden; Norway) were added that met inclusion criteria. In total, 50 studies were included in the ESC-DAGs review, summarised in Table 6-1 (below).

6.2 Methods

The application of ESC-DAGs in this thesis was using the rapid version. Box 4-1 is included here again for ease of reference in Box 6-1 below.

6.3 Results

For this or any other application of rapid ESC-DAGs, the processes used to produce the conceptual I-DAG (Mapping, Translation, Synthesis and Recombination), can feasibly be demonstrated in a sequential manner. For example, the Mapping process can present the implied graphs for all the studies, and thus the implied directed edges and concepts; the Translation process can present the DAGs produced from the implied graphs and the decision log for each primary study; Recombination

could cover the process of reducing all variables used in the literature to a workable amount, including the minutia of the graphical justifications; Synthesis could present the entirety of the directed edge index that is extracted from the translated DAGs and the conceptual I-DAG.

Box 6-1: Rapid ESC-DAGs in 11 steps

Processes common to ESC-DAGs and rapid ESC-DAGs:

1. Each study is put through the Mapping process, which involves depicting control variables as confounders unless otherwise indicated in the study. This produces an implied graph (IG) for each study.
2. Each IG is then saturated.
3. Each directed edge is assessed using the Translation process (i.e. the causal criteria and counterfactual thought experiment). This converts the IG into the DAG for that study.
4. The directed edges from that DAG are then recorded in a directed edge index until all studies have been translated to produce a DAG.
5. Concept Recombination is conducted throughout.

Rapid ESC-DAGs – conceptual I-DAG:

6. All directed edges that involve the exposure and the outcome are added to a new DAG. This produces the conceptual integrated DAG (I-DAG) which differentiates confounders from mediators. This is also the step where rapid ESC-DAG departs from the full version, which instead integrates all directed edges from the index.
7. Any 'additional' concepts are added at this stage and in relation to the exposure and outcome.

Rapid ESC-DAGs – data I-DAG:

8. Conceptual nodes are replaced with variable nodes or are marked as unmeasured.
9. Variable nodes are reorganised in the I-DAG's virtual space such that all cross-sectional measures are grouped together.
10. The I-DAG is forward saturated, meaning that a directed edge is drawn from all earlier variables to all later variables.
11. Each cross-sectional grouping is saturated and all relationships are Translated. Once this is completed for each cross-sectional grouping, the data I-DAG for that data source is complete.

Table 6-1: Summary of primary studies included in ESC-DAGs review

Author(s)	Setting	Parental influence(s)
Alati et al. (2005)	Australia	Parental control; physical punishment; maternal drinking
Alati et al. (2008)	Australia	Maternal drinking
Alati et al. (2010)	Australia	Parental control; parental physical punishment; family conflict
Alati et al. (2014)	Australia	Parental drinking; parenting style
Berge et al. (2016)	Sweden	Parenting style; parental drinking
Bodin and Strandberg (2011)	Sweden	Various
Cable and Sacker (2008)	UK	Parental drinking; parental control; adolescent disclosure to parents
Chan et al. (2013)	Australia	Attachment to parents; parental supervision; parental attitudes towards alcohol
Danielsson et al. (2011)	Sweden	Parental supply of alcohol
Degenhardt et al. (2015)	Australia	Parental drinking; parental supply of alcohol
Dick et al. (2007)	Finland	Parental monitoring; shared activities; parental drinking
Dickson et al. (2015)	Sweden	Parental autonomy granting
Engels et al. (2005)	Netherlands	Parenting style; family functioning
Hanewinkel et al. (2008)	Germany	Parental drinking; parenting style
Hemphill et al. (2011)	Australia	Parental attachment
Janssen et al. (2014)	Netherlands	Alcohol-specific rules
Kendler et al. (2013)	UK	Parental drinking
Koning et al. (2012)	Netherlands	Alcohol-specific rules; alcohol-specific communication
Koning et al. (2014)	Netherlands	Alcohol-specific rules
Koutakis et al. (2008)	Sweden	Parental attitudes towards alcohol
Kuntsche et al. (2009)	Netherlands	Parent-child relationship
Latendresse et al. (2008)	Finland	Parental drinking; parent-child relationship; parenting
Latendresse et al. (2010)	Finland	Parenting style
Maggs et al. (2008)	UK	Parent-child relationship
Mares et al. (2011)	Netherlands	Parental drinking; parental alcohol attitudes
Mattick et al. (2018)	Australia	Alcohol-specific rules; parental monitoring; parental drinking; parental supply of alcohol
McCann et al. (2016)	UK	Parental attachment, parental monitoring
McMorris et al. (2011)	Australia	Favourable attitudes; positive family management
Pape and Bye (2017)	Norway	Drinking with parents
Pedersen and von Soest (2013)	Norway	Parental drinking; parental support
Pieters et al. (2012)	Netherlands	Alcohol-specific rules; parental drinking
Poelen et al. (2007)	Netherlands	Parental drinking
Poelen et al. (2009)	Netherlands	Paternal drinking
Poulin and Denault (2012)	Canada	Parental knowledge
Raninen et al. (2018)	Sweden	Parental disapproval of alcohol; alcohol-specific rules
Rioux et al. (2016)	Canada	Parental monitoring; parental coercion
Seljamo et al. (2006)	Finland	Parental drinking
Shortt et al. (2007)	Australia	Parental attachment; parental drinking
Spijkerman et al. (2007)	Netherlands	Parental drinking; parental alcohol norms
Strandberg et al. (2014)	Sweden	Parental control; parental attitudes towards alcohol; parental warmth

Van der Vorst et al. (2006)	Netherlands	Parent-child relationship; Parental control
Van Der Vorst et al. (2007)	Netherlands	Alcohol-specific rules
Van der Vorst et al. (2010)	Netherlands	Parental supervision of drinking
Van Der Vorst et al. (2013)	Canada	Parental drinking
Van der Zwaluw et al. (2008)	Netherlands	Parental drinking
Van der Zwaluw et al. (2014)	Netherlands	Alcohol-specific rules
Van den Eijnden et al. (2011)	Netherlands	Alcohol-specific rules; alcohol-specific communication
Vermeulen-Smit et al. (2012)	Netherlands	Parental drinking; alcohol-specific rules
Visser et al. (2013)	Netherlands	Parenting style; Parent-child warmth and rejection
Weichold et al. (2014)	Germany	Parental drinking; parent-child relationship

An ESC-DAG review, even the rapid version, can thus produce an unwieldy amount of information, and this was the case for the application in this thesis. For example, as noted in Chapter 4, the decision logs alone, if compressed such that they were minimally legible, were around 250 pages long. Thus instead of presenting everything, a triage approach was taken: key information is presented in the main body of the thesis (such as the conceptual I-DAGs and the recombination of the variables into concepts); less critical information is relegated to appendices (such as the translated DAGs); and information that, while integral to the process, was neither a key output nor needed for illustration, was not included in this thesis. For example, the translated DAGs make the corresponding decision logs redundant in terms of information and also in terms of space. The implied graphs are also not included as they are redundant to the translated DAGs, as per design.

As described by Figure 6-1 below, the reporting strategy for the rapid ESC-DAGs application in this thesis is as follows. First, the concepts and variables identified by applying the Mapping processes to the primary studies are reviewed beside their Recombination into a more workable quantity of variables. In other words, the concepts that constitute the relationships in the subsequent I-DAGs are discussed first. The IGs are not discussed in any detail, nor are they included as an appendix. Next the conceptual I-DAG is introduced. This involves a degree of discussion around the Translation processes and reference to the translated DAGs in Appendix D. The decision logs are discussed but not presented in text or as an appendix, other than the example given in Chapter 4 (page 66). The current chapter concludes at this stage, with the subsequent Chapter 7 covering the variables and data I-DAG.

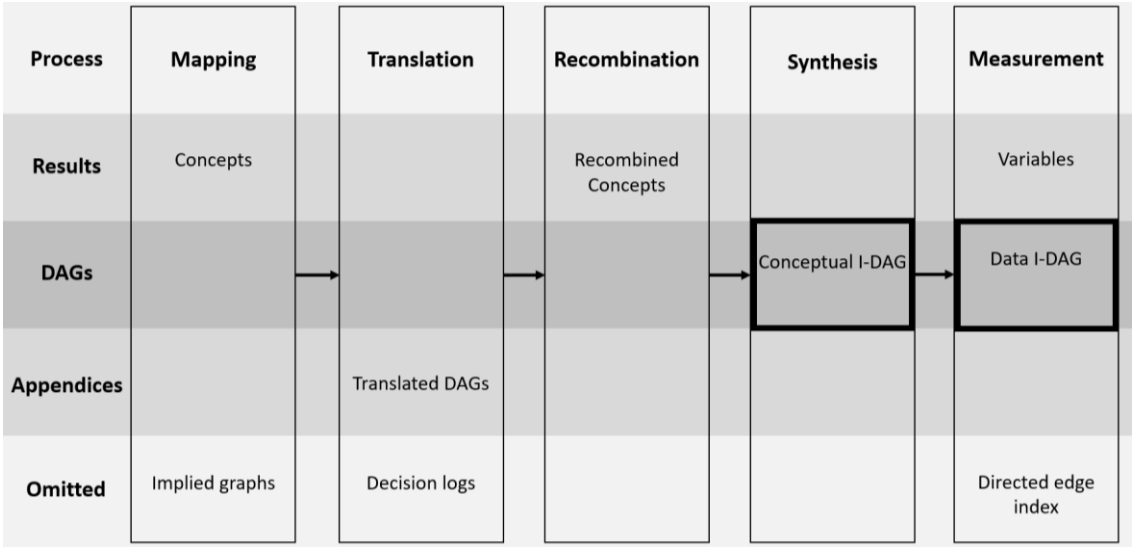


Figure 6-1: Flow diagram for reporting of ESC-DAGs

6.3.1 Concept selection and Recombination

The primary studies had a wide variety of outcome measures for adolescent alcohol harm, such as age of initiation, ever having tried alcohol, frequency-quantity constructs, instances of HED, various composite measures etc. Recombination was used to substitute each study's outcome for generic 'adolescent alcohol harm', but only during the synthesis stage – the original measure from the study was used in the corresponding IGs and DAGs, and only changed to the generic outcome when added to the directed edge index. Some outcome measures that were not recombined into the generic outcome were those relating to attitudes and early alcohol initiation, which are covered in section 6.3.1.2 below on intra-personal factors.

Excluding measures of adolescent alcohol harm, the ESC-DAGs Mapping process found 200 variables across the 50 primary studies. Using the socio-ecological model for structure (included here again for ease of reference in Figure 6-2 below), 61 of these were intra-personal (mental health, personality, behaviour, physical health, attitudes, etc.); 119 were interpersonal (parents, peers, siblings, etc.); and 20 were institutional (school, neighbourhood, etc.). The high number of interpersonal factors reflects the research question's focus on parental influences. Using the ESC-DAGs Recombination approach iteratively across the Mapping, Translation and Synthesis processes, as discussed in Chapter 4, the 200 variables were reduced to 51 concepts. The following contrasts these 51 concepts with the original variables across levels of the socio-ecological model.

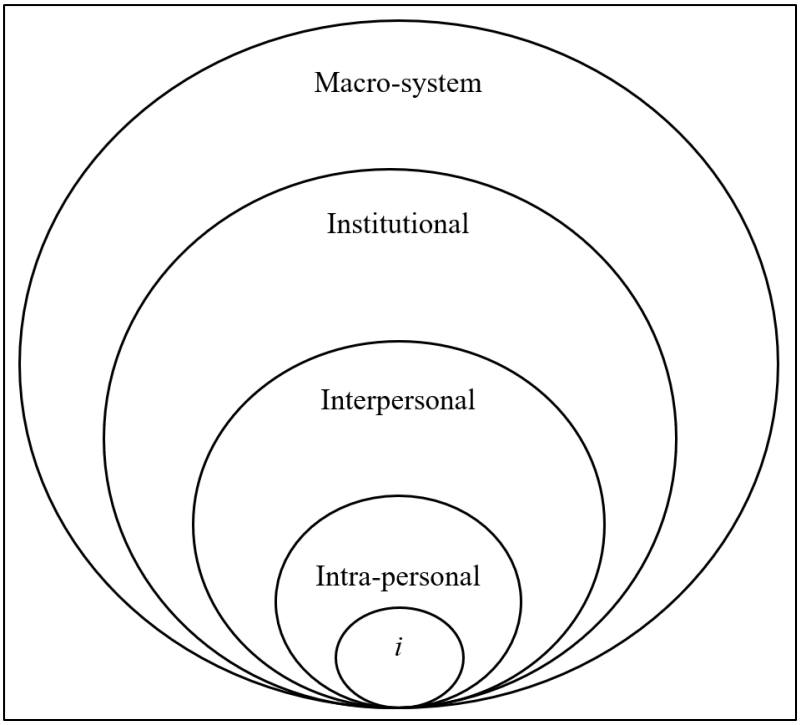


Figure 6-2: Socio-ecological model of adolescent development

6.3.1.1 Parental influences

There were 97 parental influences employed in the 50 studies, compared to the 12 discussed in the review of systematic reviews. However, 75 of these were operationalised in conceptually similar ways to the variables used in the systematic reviews (e.g. parental warmth; parental attachment; parental rejection; parent-child tension). Of the other 22 parental influences, five were related to mental health. The remaining 17 pertained exclusively to socio-demographics. This was entirely in keeping with the systematic reviews. As the primary studies conceptualised socio-demographics as control variables, the ESC-DAGs processes including Translation kept the same distinction, which in effect differentiated the parental influences such that the 'modifiable' factors (including mental health) were assigned as exposures, while socio-demographic factors were assigned as confounders. Thus, mental health was assigned as a 13th category of parental influences.

Table 6-2 (below) summarises the conceptual overlap between the parental influences discussed in the systematic reviews, and those used in the 50 primary studies. A central reason for the variety in measurement across parental influences was the differential focus on maternal and paternal influences. This was most notable for parental alcohol and the parent-child relationship. A final conceptual grouping of parental influence that was not explicitly articulated in the reviews but was covered in five of the primary studies was parental smoking, which was accordingly introduced as the 14th category of parental influences. While other substance use was not measured in any of the

primary studies, attitudes to alcohol, smoking, and substance use were all covered. Thus, the alcohol attitudes category was expanded to include substance attitudes.

There is of course a degree of conceptual overlap between the 14 categories. For example, Visser et al in their systematic review used a global definition of the parent-child relationship that subsumed parental involvement and parental support. This was supported by how both were notably under-developed categories across the 50 primary studies. Alcohol-specific rules was also relatively under-developed. This may be explained by the conceptual overlap between it and parental supply of alcohol as well as with other more general parenting rules. In other words, its ratio of 'uniqueness' to overlap with other concepts is arguably lower than most of the other categories.

None of the parental influences that were investigated in the systematic reviews were missing from the subset of 50 primary studies included in the ESC-DAGs process. As such, the ESC-DAGs review will produce a conceptual I-DAG containing all 12 of the parental influences from the review of systematic reviews, plus the additional categories of parental mental health and parental smoking.

Table 6-2: Summary of recombined parental influences

Recombined concepts	Concepts from ESC-DAGs review
Drinking	Mother's drinking; father's drinking; parental drinking; mother's alcohol problems; father's alcohol problems; parental alcohol problems; parental alcohol-related problems; parental binge drinking; parental problem drinking; historical parental drinking; drinking at home
Smoking	Mother's smoking; father's smoking, parental smoking
Supply of alcohol	Drinking with parents; alcohol availability at home; parental supervision of drinking; drinking at home; parental supply
Substance attitudes	Drug attitudes; attitudes towards anti-social behaviour; alcohol attitudes
Mental health	Maternal mental health; maternal depression; maternal anxiety; parental stress
Family coherence	Parental conflict; family conflict; family problem solving; family stress; positive family relations; family management; family resilience; family problem-solving; family functioning; prosocial family engagement; family connections to community; family connections to school

Parent-child relationship	Maternal attachment; paternal attachment; rejection; warmth; tension; attachment; parental knowledge; parental care; parent-child relationship quality
Parental involvement	Shared activities; family-oriented leisure time; reading to adolescent during childhood; time with family
Parental support	Father's support; desire for adolescent to fit in
Parental monitoring	Autonomy granting; parental control; overprotection
Parental discipline	Parenting style; perceived parenting; physical punishment; authoritative parenting; parenting consistency; parenting of older siblings
Alcohol-specific rules	Permissiveness towards adolescent alcohol use
Parent-child communication	Disclosure to parent; frequency of communication; quality of communication
Alcohol-specific communication	Alcohol-specific communication

6.3.1.2 Intra-personal

There was an extensive array of intra-personal factors, including personality, mental health, physical health, other substance behaviours and more as described in Table 6-3 below. A notable omission was genetic predisposition towards alcohol and other addictive substances, although this may be explained by how family history of alcohol problems, covered below, is a much more convenient measure. Indeed, genetics were beyond the scope of this study and for this reason were not included as an 'additional concept'.

Table 6-3: Summary of recombined intra-personal influences

Recombined concept	Concepts from ESC-DAGs review
Alcohol attitudes	Attitudes to alcohol; alcohol-related memory; alcohol expectations
Early alcohol initiation	Early alcohol initiation
Smoking	Is a smoker; has tried cigarettes; age of smoking initiation
Substance use	Perceived availability; age of substance use initiation; cannabis use; attitudes to substances
Risk personality	Sensation seeking; rebelliousness; impulsivity; extroversion; self-control; emotional stability; stimulus response; working memory capacity

General personality	Aggression; social skills; problem-solving; conscientiousness; agreeableness; 'ableness'; resourcefulness
Beliefs	Moral order; religiosity
General behaviour	Anti-social behaviour; conduct problems; delinquency; sports; gaming
Mental health	Depression; anxiety; coping; negative life events; internalising; externalising; life transitions; somatic complaints
Physical health	BMI; disability; gestational age; birthweight; post-natal ICU; foetal alcohol or tobacco exposure
Economic activity	Employment; disposable income

6.3.1.3 Interpersonal

Asides from parents, interpersonal peer influences were widely investigated, especially in terms of their drinking, substance use, and anti-social behaviour. The general peer network was also represented in terms of numbers and age composition. Sibling effects also received attention. Table 6-4 below summarises the interpersonal concepts from the ESC-DAGs review versus the recombined concepts. Notably, teachers were not included in any of the 50 studies, although this may be explained by the overlap between teachers as interpersonal actors and as representatives of schooling on an institutional level. The extended family was also largely ignored (represented only by family history of alcohol problems). A final notable omission was experience of child abuse, which was surprising given evidence that forms of child abuse (physical, emotional, sexual, and neglect) are related to adolescent alcohol harm (Sartor et al., 2007, Maniglio, 2009) and given the wider emphasis on adverse childhood experiences (Dube et al., 2002, Dube et al., 2006, Scottish Government, 2018).

Table 6-4: Summary of recombined interpersonal influences

Recombined concept	Concepts from ESC-DAGs review
Number of friends	Number of older friends; significant adult attachment; other sex friends
Peer drinking	Peer alcohol attitudes; best friend's drinking; school friends' drinking; peer alcohol norms
Peer smoking	Peer smoking; school friends' smoking; peer smoking attitudes
Peer substance use	Peer substance attitudes; best friend's cannabis use
Peer anti-social behaviour	Anti-social behaviour; negative non-social peers; peer problem behaviour
Sibling effects	Sibling drinking; older siblings; zygosity

6.3.1.4 Institutional

There were three main groupings of institutional factors for adolescents; school; neighbourhood/ community; and family. School factors focused on academic performance and engagement with school. Neighbourhood and community factors were mainly concerned with community engagement, neighbourhood safety and neighbourhood SES. Family institutional factors mainly pertained to aspects of family function, resilience, and conflict. As family effects were largely utilised as exposures in the primary studies they were categorised as exposures here also, and as such were included in the parental influences table above. Table 6-5 below summarises the institutional effects. Work may also be considered as institutional but it was reserved as an intra-personal effect to emphasise how the primary studies focused on the income aspect of engagement with the work force.

Table 6-5: Summary of recombined institutional influences

Recombined concept	Concepts from ESC-DAGs review
School engagement	Leaving plans; drop out; bullying; absence; rewards at school; prosocial rewards; prosocial opportunities; safe school environment
Academic performance	School grades; academic self-confidence; type of school
Neighbourhood SES	Neighbourhood deprivation; neighbourhood safety
Neighbourhood attachment	community disorganisation; prosocial rewards; prosocial opportunities; neighbourhood attachment; school clustering
Family coherence	See parental influences
Urbanity	Urbanity

It was surprising that availability of alcohol at the neighbourhood level was not investigated in any of the 50 studies. Again this is a widely studied exposure with evidence to suggest a harmful effect, although the illegality of underage drinking may explain its absence (Bryden et al., 2012). One study did consider exposure to alcohol in film media (Hanewinkel et al., 2008). Another aspect that was largely neglected was the character of space in the physical neighbourhood (e.g. greenspace, derelict space, etc.), which was only indirectly articulated by neighbourhood SES. This can be an important determinant of drinking context (Jackson et al., 2014).

6.3.1.5 Socio-demographics

Socio-demographics are considered separately here as they constitute the main confounders from the 50 primary studies. The list was fairly typical for studies of adolescents and included age, sex,

ethnicity, family structure, and various measures of socio-economic status (household income, parental education, parental occupation, etc). A socio-demographic confounder that was commonly used but that may be more typical of studies of adolescent alcohol harm than other topics related to adolescents was family history of alcohol problems, which was viewed as likely to increase the adolescent's own chances of experiencing alcohol problems. One concept that was missing across all 50 studies was sexuality. This was surprising given the strong evidence that sexuality has a notable differential influence on adolescent alcohol harm (Marshal et al., 2008).

Table 6-6: Summary of recombined socio-demographics

Recombined concept	All concepts from ESC-DAGs review
Parental socio-economic status	Family income; father's occupation; living arrangements; parental education; parental occupational status; NS-SEC
Parental education	Maternal education;
Family history of alcohol	Family history of alcohol
Maternal age	Age of parents
Family structure	Parental divorce
Ethnicity	Religiosity
Sex	Sex

6.3.1.6 Additional concepts

Concepts that were not covered in the 50 primary studies but that were added to the conceptual I-DAG were experience of abuse or neglect, relationships with teachers, extended family effects, the alcohol environment at a neighbourhood level, the character of space in the neighbourhood, and sexuality. The process of selecting these concepts relied on comparing the prior list of concepts to the background knowledge of the author and the supervisory team. Including the outcome, this took the total number of concepts to 52. As this thesis was interested in generalising to the UK population using the ALSPAC data, policy-level determinants (which are often one of the main constituents of the 'macro-layer' of the socio-ecological model) were not incorporated.

6.3.2 Conceptual I-DAG

The conceptual I-DAG focused on the relationships between the above concepts. It was generated by using the Synthesis processes on all directed edges identified from the Translation processes to involve at least one exposure or outcome node. The conceptual I-DAG is presented in Figure 6-3 below. While the DAGitty software was used extensively when building the IGs, DAGs and the

conceptual I-DAG, a word processor was used at this stage to generate a graph which legibly relates 52 concepts. For consistency, the colour coding of nodes (boxes) and directed edges (arrows) replicates DAGitty output. This section considers key aspects of the Translation process and describes how it applied to some examples from the 50 primary studies, and how the resulting DAGs shaped the conceptual I-DAG. First, however, the conceptual I-DAG is given some interpretation.

The purpose of the conceptual I-DAG is not to guide analysis directly, but to act as a template from which a data I-DAG can be built using any data source. For example, if the aim was to use two data sources, then the conceptual I-DAG would be used as the template for the data I-DAG for both data sources. Given this purpose and how rapid ESC-DAGs does not saturate the conceptual I-DAG, building statistical models based on this graph alone would be problematic. However, while naïve of data, the conceptual I-DAG encodes the following pathways:

1. Baseline confounders → exposures → intermediates → outcome
2. Baseline confounders → exposures → outcome
3. Baseline confounders → intermediates → outcome
4. Baseline confounders → outcome
5. Exposures → intermediates → outcome
6. Exposures → outcome
7. Intermediates → outcome

Nevertheless, these causal relationships do imply some statistical models. For example, when estimating the effect of any of the exposures (e.g. parental discipline), all of the base confounders should be controlled for. Similarly, when estimating the effect of any of the intermediates (e.g. smoking), all of the baseline confounders and exposures should be controlled for. However, because none of the concepts within each grouping are inter-related, the conceptual I-DAG cannot be used to decompose the total effect of the exposures (i.e. there is no way to differentiate EIMOCs from mediators).

6.3.2.1 Conceptual I-DAG: Translation

As noted above the causal pathways encoded in the conceptual I-DAG were generated by synthesising the directed edges pertaining to the exposure-outcome relationship that were determined by the Translation process. Rather than discuss the rationale behind numerous individual directed edges (for example every directed edge from each parental influence to each intermediate in the conceptual I-DAG), the main groups of directed edges are discussed in turn. They are: those from parental influences to the adolescent's intrapersonal effects; those from parental influences to interpersonal influences, especially peer effects; and those from the confounders to the parental influences.

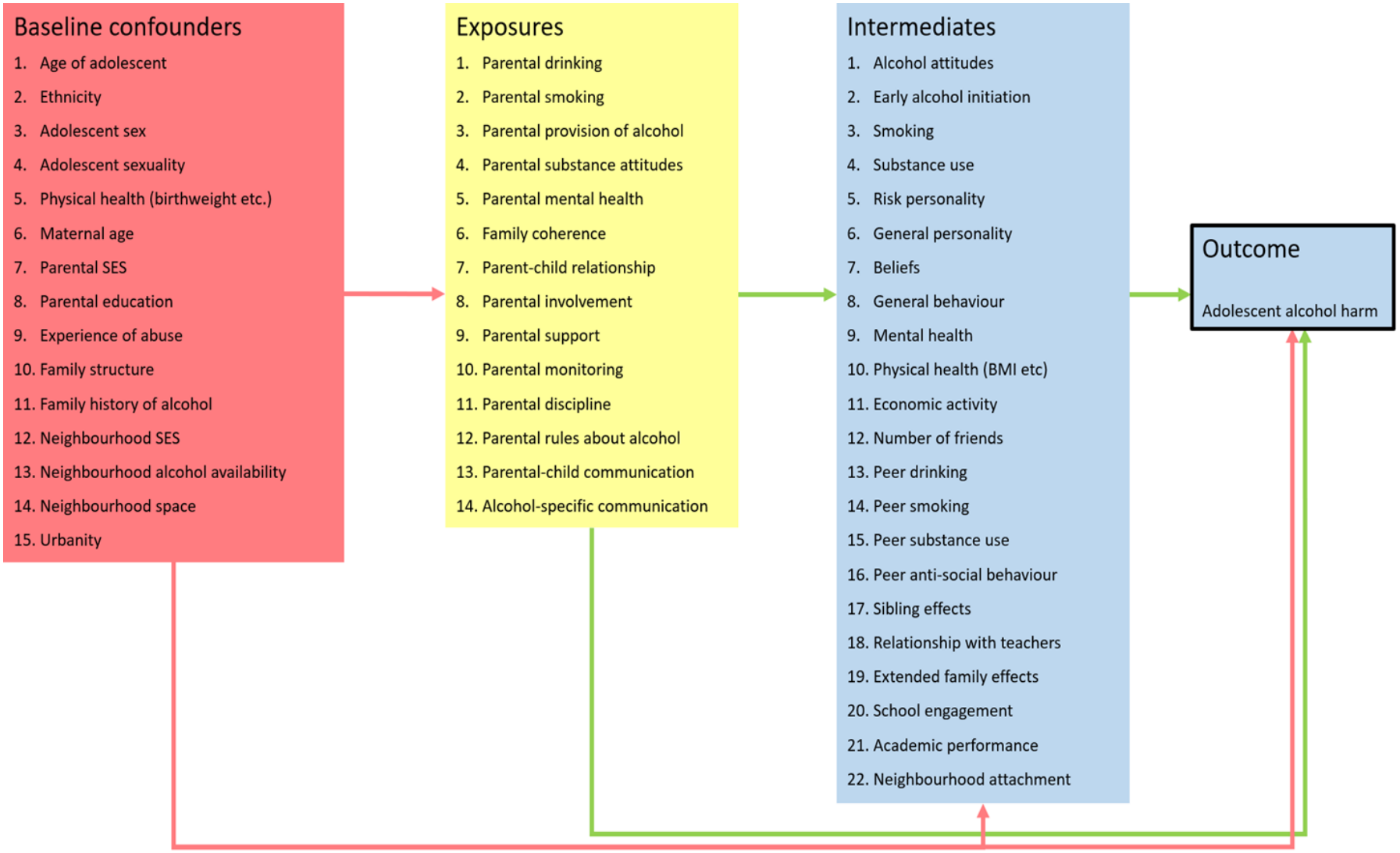


Figure 6-3: Conceptual I-DAG

Parental influences → intra-personal influences

The first pattern is that time-varying intra-personal concepts pertaining to the adolescent (e.g. attitudes, behaviours, personality, mental health) are all downstream of the parental exposures (parental exposures → intra-personal intermediates). Thus, the Translation procedures reflected the strong trend in Public Health for analyses of children to control for their parents, but not vice versa. For example, a study by Poulin and Denault reviewed by Yap et al, used anti-social behaviour as a control variable when estimating the effect of parental monitoring on alcohol use (Poulin and Denault, 2012). The Translation process used the causal criteria and counterfactual thought experiment to conclude that parental monitoring has a causal effect on the adolescent's anti-social behaviour (although it was noted that there could be an aspect of bi-directionality between them). Thus, for both the DAG for that study and the Synthesis step, adolescent anti-social behaviour was posited as an intermediate of the effect of parental monitoring and the other parental influences. Similarly, the Translation process concluded that adolescent anti-social behaviour is a cause of alcohol use (which Poulin and Denault measured using a composite score), again with some degree of bi-directionality. The corresponding DAG is presented below in Figure 6-4 below. It also shows that the Translation processes concluded that the number of other sex friends and adolescent substance use are caused by (and mediate) parental monitoring; that anti-social behaviour causes both; and that number of other sex friends determines adolescent substance use. The effect of anti-social behaviour on substance use was suppressed for illustration. The convention used across this and the following DAGs is that, for any 'column' of variables, ones below are always influenced by ones above. This is revisited in the following chapter in the context of the data I-DAG. All relationships in this DAG were noted to have a degree of bi-directionality.

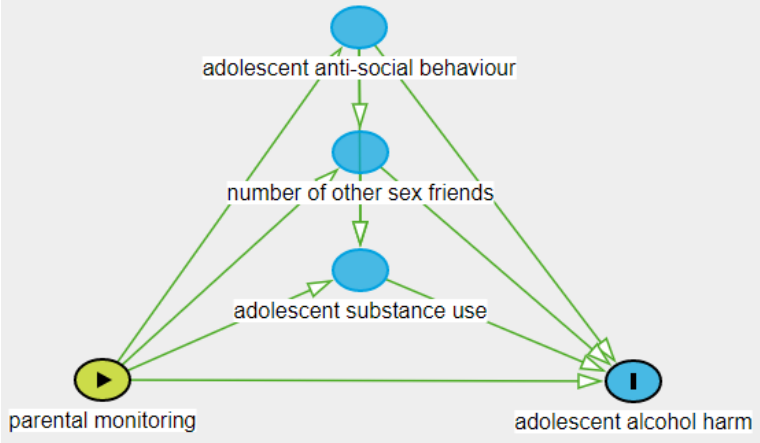


Figure 6-4: DAG for Poulin and Denault (2012)

Parental influences → Peer influences

Secondly, all peer effects in the conceptual I-DAG are downstream of parental influences (parental exposures → peer intermediates). This follows the same underpinning logic as above – adults

influence children. More specifically, adults influence children *more* than children influence adults. Note, however, that this is not to say that statistical control for the adolescent would fail to remove confounding bias when estimating effects of parents. Rather it implies that, when estimating the effect of parenting on adolescent behaviour in a cross-sectional setting, controlling for the adolescent would induce more bias than it would remove (i.e. it would open a backdoor path by controlling for a mediator). This includes the adolescent's relationships with other interpersonal actors such as peers and siblings. In the above example from Poulin and Denault, number of friends was posited as being downstream from parental monitoring. In another study reviewed by Yap et al, Spijkerman et al controlled for peer drinking when estimating the effect of parental drinking (Spijkerman et al., 2007). The translation process concluded that there is a causal effect of parental drinking on peer drinking, but that any reverse effect was likely to be very small or negligible. Thus, the DAG drew a directed edge from parental drinking to peer drinking, which was then added to the conceptual I-DAG via the directed edge index. Figure 6-5 below is the DAG for the Spikjerman et al study. It implies that the effect of parental alcohol norms and parental drinking influence adolescent alcohol harm via peer drinking norms, peer drinking, alcohol perceptions of the adolescent, and early alcohol initiation. It also implies that peer drinking norms influence all other intermediates; that peer drinking influences adolescent alcohol perceptions and earlier initiation; and that perceptions influence initiation. These directed edges were again suppressed for illustration. Once more, all of these directed edges were recorded as candidates for DAG-based sensitivity analysis (i.e. the directed edges could be reversed and still make some sense). Note that, while relatively simple at only 7 nodes, this DAG is already dense with causal hypotheses, possessing 48 causal paths.

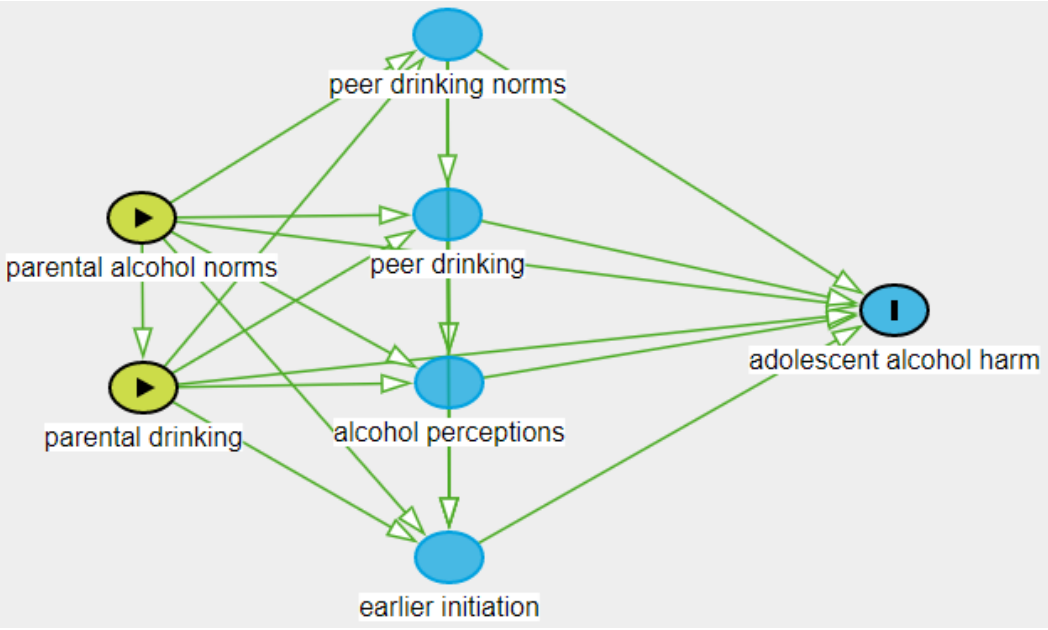


Figure 6-5: DAG for Spijkerman et al (2007)

Socio-demographics and historical events → parental influences

While there is some degree of reverse causation of the parent → adolescent pathway, the baseline confounders were more straightforward. They can be divided into two groups. First are time invariant demographics and events that are historical in regards to parenting of adolescents and adolescent alcohol use (age, ethnicity, sex, sexuality, family history of alcohol, birth-weight). In studies for which any of these were included, the reverse edge was rejected due to it being non-temporal (e.g. age is determined by birth year, ethnicity by parents/grandparents, sex randomly at birth etc). For example, Alati et al conducted a study reviewed by Rossow et al in which they controlled for birth weight and gestational age when estimating the effect of the quantity of alcohol mothers drank at when the adolescent was aged 5 on a quantity-frequency measure of adolescent alcohol use at age 14 (Alati et al., 2008). For birth weight and gestational age, the Translation processes concluded that both would influence later maternal drinking as well as adolescent drinking, but that neither later maternal drinking or adolescent drinking could influence birth-weight or gestational age.

Figure 6-6 below is the DAG for the Alati study. The column of red nodes are the confounders, the column of yellow are the exposures, and causation always flows downwards in these columns. This study and DAG are good examples of different ways in which the Translation processes can play out. First, sometimes the directed edge under assessment will be concerned with an obvious process, for example gestational age influences birthweight. Second, even in studies with a high degree of confounder control, there can be variables that are not influenced by any others and are as such 'exogenous' (e.g. sex). Third, careful consideration of the measurement is required, as per the causal assumption of consistency discussed in Chapter 3. For example, maternal age is not a measure of age but rather a measure of how old the mother was when the child was conceived. As such, perhaps counter-intuitively, maternal age is not exogenous (in much the same way as age of initiation). Relatedly, often the timing of a measurement can change the conceptual meaning of a variable. Maternal characteristics measured during pregnancy are a good example. In the below DAG mental health is treated as a confounder that influences other confounders (birthweight etc.). However, in the context of this thesis, maternal mental health is treated as an exposure. This speaks to the influence the research question imposes on the DAG, mentioned in Chapter 4. Fourth, even the DAG for an individual study can become complex enough to be challenging to interpret. However, most of the complexity in this example was in terms of the confounders – there are only 14 causal paths compared to 48 in the above Spijkerman example with more mediators. Finally, this DAG has examples of when it is not possible to assume that variables are cross-sectional, as recommended in the Translation process. Clearly, there is no benefit to assessing the reverse edge for the relationship between an individual's birthweight and their behaviour during adolescence.

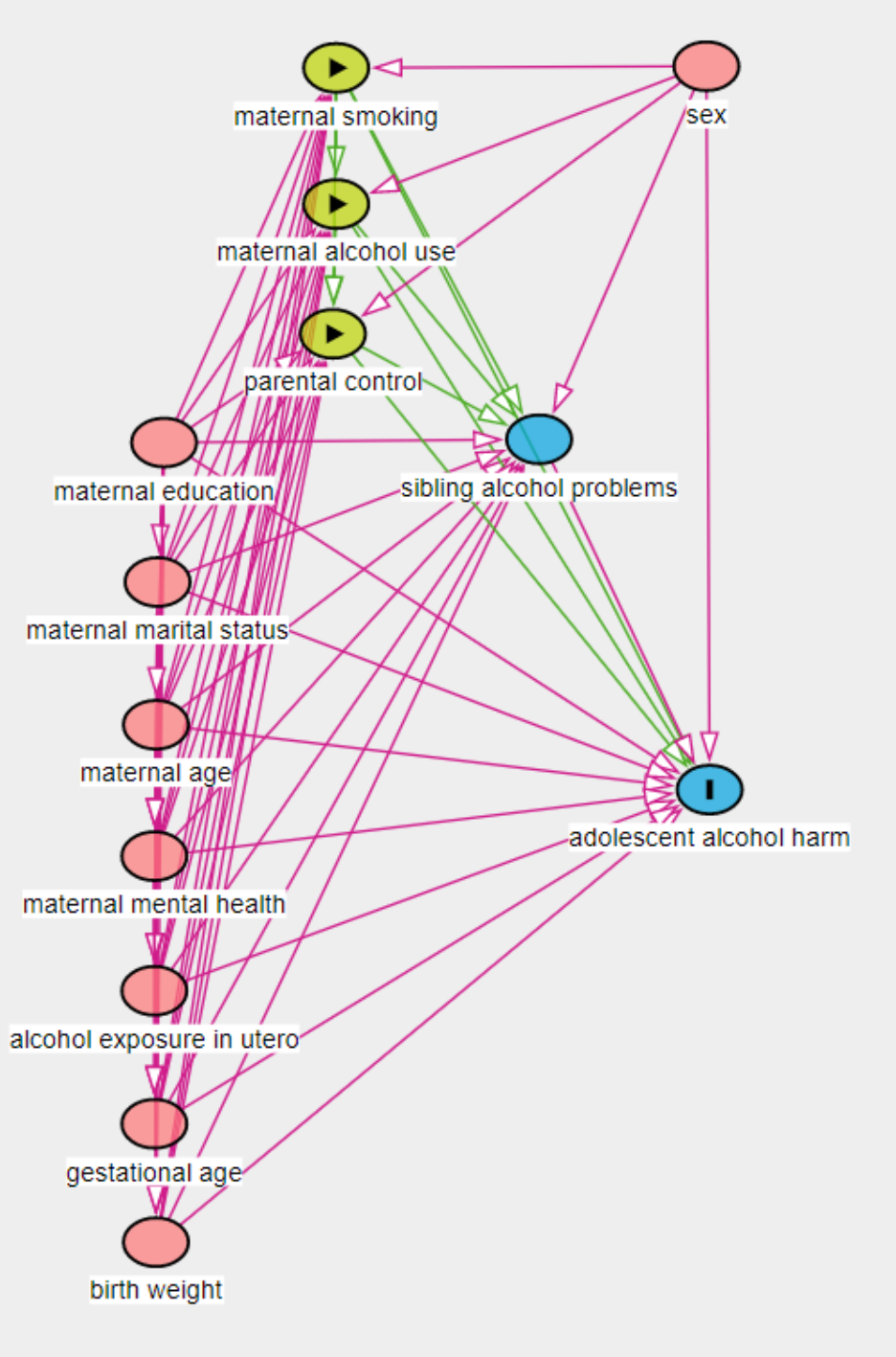


Figure 6-6: DAG for Alati et al (2008)

Time dependent confounders → Parental influences

The second set of confounders were time varying aspects that occur during the adolescent's childhood and influence the adolescent and the parent (ante-natal and childhood health, parental SES, parental education, experience of child abuse, family structure, neighbourhood SES, neighbourhood alcohol availability, neighbourhood space, urbanity). In the above DAG, this would include maternal education and marital status, for example. Several other studies controlled for the effect of parental

SES (Mattick et al., 2018, Alati et al., 2010, Van den Eijnden et al., 2011). For example, Kendler et al used the ALSPAC data to estimate the effect of parental alcohol consumption in units per week on adolescent alcohol use disorders at age 18 using the Alcohol Use Disorder Identification Test (AUDIT) while controlling for parental occupation (Kendler et al., 2013). The DAG is included as Figure 6-7 below. The Translation process concluded that parental occupation is likely to influence parental drinking and adolescent AUDIT score. However, it did note that alcohol dependence is likely to influence occupation status, and so a possible bi-directional relationship was recorded. In other words, the first set of confounders were generally assumed to have little-to-no bi-directionality with the exposures, while there was some degree of bi-directionality in this second group as regards the exposures, but not with the intermediates or outcome. Note that the intra-personal and institutional influences on the adolescent are shared between the baseline confounders and the intermediates, as summarised in Table 6-7 below.

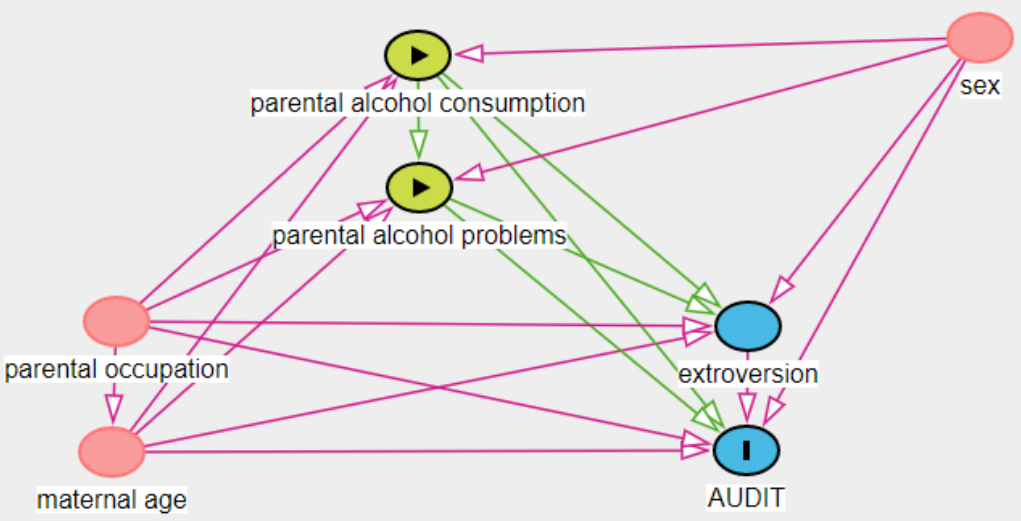


Figure 6-7: DAG for Kendler et al (2013)

For the variables which were not covered in the primary studies but which were added during the synthesis stage, such as neighbourhood availability of alcohol or sexuality, Translation took place after the conceptual I-DAG had been developed. In the first instance, this was conducted in reference to the more general notion of parental influences, and in the second instance to some of the specific concepts. For example, parental drinking and parenting were assumed to be influenced by neighbourhood availability of alcohol. It was noted that there was the possibility of a reverse relationship, (e.g. parents moving the family to a new neighbourhood based on alcohol availability) but that this was probably negligible.

Table 6-7: Causal roles of all recombined concepts from conceptual I-DAG

Role in conceptual I-DAG	Socio-ecological grouping	Recombined concepts
Intermediates	Interpersonal	Peer drinking; peer anti-social behaviour; number of friends; peer smoking; peer substance use; sibling effects; extended family effects; experience of child abuse or neglect
	Intra-personal	Alcohol attitudes; early alcohol initiation; smoking; substance use; risk personality; general personality; beliefs; behaviour; mental health; physical health (BMI); economic activity;
	Institutional	Academic performance; school engagement; neighbourhood attachment
Baseline confounders	Intra-personal	Physical health (birthweight, gestational age, disability, foetal exposure to tobacco or alcohol)
	Institutional	Neighbourhood SES; urbanity; alcohol environment
	Socio-demographic	Parental SES; parental education; family history of alcohol problems; maternal age; family structure; ethnicity; sex; sexuality

6.4 Discussion

This chapter described how further inclusion criteria were applied to the 183 studies from the systematic reviews to restrict the rapid ESC-DAGs procedures to a more workable 50 studies. The concepts from these 50 studies were extracted using the Mapping process and tabulated above in contrast to the concepts that were defined by the Recombination process. This reduced the number of concepts from 200 to 51. The conceptual I-DAG that was produced using the Synthesis process was then presented and interpreted. Thus, this chapter contributed to MRO 3 (apply and demonstrate ESC-DAGs) by describing key outputs from ESC-DAGs up to the Synthesis stage. It also contributed to SRO 3 (build DAGs of parental influences on adolescent alcohol harm) by positing a set of concepts which confound the exposure-outcome relationship, and a second set of concepts which are intermediates between parental influences and adolescent alcohol harm (i.e. the conceptual I-DAG).

No claim is made here that the conceptual I-DAG definitively represents the causation of adolescent alcohol harm, rather it is a succinct graphical abstract of the 50 studies that contributed to this application of rapid ESC-DAGs. Nevertheless, the breadth of these concepts is notable, especially given the intention to use each as a covariate, and how systematic the ESC-DAGs approach to selecting and inter-relating concepts was. As expected given its evidence synthesis emphasis, the conceptual I-DAG produced from this approach was entirely in line with what was suggested by the review of systemic reviews in the previous chapter – parental influences shape the intra-personal, interpersonal and institutional aspects of adolescent life that pertain to adolescent alcohol harm, and parental influences themselves are contextualised against socio-demographic and historical influences.

This chapter represents only part of the ESC-DAGs application in this thesis and as such has not closed any research objectives. Nevertheless, as the Mapping, Translation, and Recombination processes are in an advanced stage by the time the conceptual I-DAG is produced, discussion on the rapid ESC-DAGs process up to this stage is worthwhile.

6.4.1 Mapping and Translation

The Mapping and initial Translation of each primary study served several purposes. First, this was how each concept was identified. Second, it identified the directed edges that would go on to define the conceptual I-DAG. Third, it also commented on some of the modelling practices that are common in the research community – both undercontrolling for confounders and overcontrolling for mediators were common. Chapter 2 discussed how these approaches are likely to 'open backdoor paths' or otherwise fail to remove bias. Chapter 8 demonstrates these biases using ESC-DAGs-based analysis of the ALSPAC data. Thus, in producing the conceptual I-DAG, the rapid ESC-DAGs method can offer a somewhat unique and illustrative critique of the literature. However, the primary purpose of ESC-DAGs is to design a DAG for analysis rather than as a bias appraisal tool. This is covered further in Chapter 10's discussion on using ESC-DAGs as the basis for new bias appraisal tools.

The above demonstrated some of the key patterns that emerged during the Translation process. One of the most notable was the relative lack of relationships that are uni-directional. In other words, when working with causal relationships in social science, there is almost always some degree of bi-directionality between time-varying concepts inherent to the same individual. Relatedly, this study benefited from the nature of the research question – the idea that parents influence children is not controversial. However, if the exposure of interest had been an intra-personal aspect for the adolescent, such as behaviour, then the delineation between exposures and intermediates would have

been a somewhat artificial consequence of that research question. For example, to posit anti-social behaviour as an intermediate for mental health would be heavily caveated by the reverse relationship. This is not to say that the Translation process cannot be used to generate carefully thought-out causal hypotheses in such circumstances – a bi-directional relationship is a valid causal hypothesis.

6.4.2 Recombination

There were instances in which the level of over- or under-generalisation applied between similar themes appears inconsistent. For example, parental drinking, parental smoking, and parental substance use were each used as individual concepts while attitudes towards alcohol and illicit substances were combined into one concept (called substance attitudes). Likewise, alcohol-specific rules and parental supply of alcohol are treated separately, despite the fact that the latter is clearly an aspect of the former. In both cases, the explanation lies in how the concepts were treated in the literature – there are numerous studies on parental supply of alcohol with fewer using the more general concept of 'alcohol-specific rules'. Similarly, attitudes featured much less prevalently than actual behaviours (for example, attitudes to smoking did not feature in any study). However, the ESC-DAGs approach to Recombination (consider conceptual equivalence or nesting, and consider how the two nodes relate to the rest of the integrated DAG) are among the least systematic across the protocol as a whole. There is thus an innate tension in the ESC-DAGs processes between positing causal hypotheses that are believed to be 'true' and causal hypotheses that are investigated in the literature. It should be noted that, while ESC-DAG is a method for reviewing *evidence*, the self-discipline required on the part of the reviewer to avoid making the small but meaningful leap to positing what they assume to be 'true' can be difficult to sustain.

Chapter 7 ESC-DAGs review 2: Data I-DAG

Chapters 4, 5 and 6 introduced ESC-DAGs, reviewed the literature on parental influences on adolescent alcohol use, and applied the ESC-DAGs methodology to that literature to produce the conceptual I-DAG. This chapter corresponds the step of replacing the conceptual nodes in the data I-DAG with variables using data from ALSPAC. It primarily pertains to MRO 3 (apply and demonstrate ESC-DAGs) and SRO 3 (use ESC-DAGs to build DAGs of parental influences on adolescent alcohol harm). Accordingly, as the data I-DAG is the culmination of the ESC-DAG processes, and as the ESC-DAGs processes are the primary novel contribution of this PhD, this chapter is the fulcrum of this thesis.

This chapter considers the data requirements needed for the I-DAG before describing the ALSPAC data source in some detail. It then demonstrates how the ALSPAC data was used to 'measure' the conceptual I-DAG from Chapter 6. Basic descriptive statistics and associations are provided. Finally, this chapter closes by reflecting on the relationship between DAGs and data, and on the analytical models posited by the data I-DAG.

7.1 Data requirements

Given how the methodological and substantive objectives of this thesis are concerned with causal inference, the data requirements were relatively strict. The data source should:

1. Be based in the UK (to assist with generalising to the UK population).
2. Be recent (to further improve external validity for generalising to the recent UK population).
3. Have a large sample size (statistical power).
4. Be a longitudinal data source (to account for reverse causation).

5. Have a variable catalogue of sufficient breadth to span the socio-ecological model of health/conceptual I-DAG (to assist with confounding bias).
6. Have adequate measurement of adolescent alcohol behaviour (to assist with measurement bias on the outcome).
7. Be an inter-generational cohort with self-report on parents' and adolescents' behaviours (to assist with measurement bias, specifically misclassification of the parental influences).

These criteria automatically excluded several high-profile longitudinal UK-based surveys. For example, household surveys such as Understanding Society are not birth cohorts; the 1970 British Cohort Study and 1958 National Child Development Study are not recent – generalising from the adolescent alcohol experience in the 1970s and 1980s to current UK adolescents is not ideal. ALSPAC, however, meets each of these requirements. It is UK-based, recent, longitudinal, and has a large sample size. The range of variables in ALSPAC is broad enough to offer sufficient coverage of the socio-ecological framework and conceptual I-DAG, including detailed survey data on intrapersonal factors (e.g. demographics, personality, mental health, substance behaviours), interpersonal factors (e.g. parent-child, parent-parent, adolescent-peer), and institutional factors (e.g. parent-neighbourhood, adolescent-neighbourhood, adolescent-school). It has detailed information on adolescent drinking, including use of validated instruments with clinical relevance. It collects data on the parents as well as on the adolescents, meaning that researchers can use parental self-report rather than rely on proxy measures from the adolescent. It is also notable that several studies using the ALSPAC data have been argued to be strong examples of causal inference from analysis of observational data (Fraser et al., 2013).

Other suitable data sources included the Millennium Cohort Study (MCS) (Hansen et al., 2012). ALSPAC was selected over MCS firstly because it started following participants during pregnancy rather than at 9 months old, offering a wider selection of potential confounders. Secondly, when this PhD started in 2015, members of the MCS were as young as 13 and thus not as suitable as ALSPAC participants for research on alcohol use and harm in adolescence.

7.2 The ALSPAC data

ALSPAC is a prospective population-based survey of a birth cohort and their mothers which started data collection in 1991. It originated at a WHO meeting in Moscow in 1985 which called for longitudinal birth cohorts to be established across Europe to investigate modifiable influences on child health and development (Golding, 1990, Golding et al., 2001). ALSPAC is one of several such

surveys collectively known as the 'European Longitudinal Study of Pregnancy and Childhood' (ELSPAC) (Golding, 1989). ALSPAC is based in the Avon area in the South-West of England, with focus on the city of Bristol. It has received funding from multiple sources, mainly the UK Medical Research Council, the Wellcome Trust, and the University of Bristol. The University of Bristol is also the institution responsible for data collection, maintenance, and provision, and it maintains a comprehensive website on the resource available at http://www.bristol.ac.uk/alspac/. Access to the data is via an online research proposal on the same website. The website also catalogues all peer reviewed publications based on the ALSPAC data (over 1500).

ALSPAC follows the children born to 13,761 women between 1st April 1991 to 31st December 1992 (Fraser et al., 2013). Data collection commenced during pregnancy. Over the duration of the survey, data collection has mainly taken the form of postal survey questionnaires, although occasional clinic visits were also involved for the child from the age of 7 onwards (ages 7, 11, 12.5, 13.5, 15.5, and 17.5). The mothers were followed closely during childhood, but less so as the child entered adolescence. Data on the child was exclusively collected via the mother until 8 years of age, after which questionnaires were increasingly designed for self-completion by the child. It should be noted that, while data were also collected from the mothers' partners, their participation rates were less than half that of the mothers by the time the child was aged 7. Table 7-1 below summarises the time points for measurement of parental and child data. Data is still collected from the children on a near-annual basis, even now that the oldest participants are in their late twenties (a third generation has recently been integrated into ALSPAC, although the sample size is less than 1,000). While ALSPAC was initially known as the 'Avon Longitudinal Study of Pregnancy and Childhood', as the cohort aged beyond childhood the survey was renamed as the 'Avon Longitudinal Study of "Parents and Children"' to emphasise the inter-generational nature of the data (Golding, 1989). ALSPAC is also known as the 'Children of the 90s study'.

Table 7-1: Number of surveys per subject in ALSPAC

Timepoint	Subject of survey			
	Mother	Partner	Child (completed by mother)	Child (completed by child)
Pregnancy	5	1	0	0
Age 0 – 11	12	12	24	9
Age 12 - 17	2	2	2	13
Total	19	15	26	22

Initial ethical approval for ALSPAC was granted by Bristol and Weston Health Authority in 1989. Research proposals for studies which do not use human tissue data, NHS premises, or ionizing radiation require ethical approval only from the ALSPAC Ethics and Law Committee (i.e. not from an NHS Research Ethics Committee). The College of Medical, Veterinary, and Life Sciences at the University of Glasgow also does not require further ethical approval on projects working with ALSPAC data which do not use human tissue data, NHS premises, or ionizing radiation. The data proposal for this PhD was submitted to the ALSPAC executive board on the 29th of April 2016 and approved on the 31st of May 2016.

ESC-DAGs was still in development in 2016, as such the conceptual I-DAG in Chapter 6 did not exist. Instead, the socio-ecological model of health was used to inform variable selection processes for the data proposal. The ALSPAC variable catalogues were searched using indexed searching with assistance from a 'data buddy' based at the ALSPAC data centre at the University of Bristol's Medical School. After ESC-DAGs was used to develop the conceptual I-DAG, the same processes were repeated to ensure adequate coverage of the concepts in the conceptual I-DAG. Further, the author also benefited from a month-long placement at the ALSPAC data centre in June 2018, working with researchers who were experts on using the data source for analysis of alcohol outcomes, and funded by a Supplement Award from the Medical Research Council Doctoral Training Program.

7.2.1 External validity of ALSPAC

Like with any other data source, inferences from ALSPAC must be caveated by how similar the sample is to the reference population, in this case UK families with children born in 1990/91. Table 7-2 below is excerpted from Fraser et al's 2013 update on the cohort profile, which compares ALSPAC data measured at 8 months postpartum to the 1991 census for women with an infant <1 year of age. The fifth column was generated by the current author to further compare the cohort at age 16.5 in 2008/09 when the outcome used in this thesis was measured (discussed in more detail below). Overall Table 7-2 suggests that the ALSPAC participants are more affluent than the general UK population. The fifth column indicates that the cohort may become more dissimilar to the UK population over time. Some of these differences are substantial, for example the cohort has 20% higher home and car ownership than the GB average. While, this does not account for myriad exogenous influences over the 16 year interval (e.g. fluctuations in house and car prices), compared to the 2011 census, ALSPAC has a higher rate of car ownership than any local authority in England and Wales (95.6% in ALSPAC compared to 89.7% for East Dorset (RAC Foundation, 2012)) and a much greater than average rate of home ownership (85.3% in ALSPAC compared to 64.3% average in England and Wales (Office for National Statistics, 2011)). Note, however, that the 2011 census is for all households, not those with children born in 1990/91.

Table 7-2: Socio-economic generalisability of ALSPAC

Characteristic	UK population, 1991 census (%)	Avon, 1991 census (%)	ALSPAC at age 8 months in 1991/92 (%)	ALSPAC at age 16.5 years in 2008/09 (%)
Owner occupier	63.4	68.7	79.1	85.3
1+ person/room	30.8	26.8	33.5	25.6
Car in household	75.6	83.7	90.8	95.6
Married couple	71.8	71.7	79.4	76
Non-white mother	7.6	4.1	2.2	3.4

As such, while it can be argued that ALSPAC started out as fairly similar to the UK population as a whole, it appears that this has become less so over time. Table 7-3 below further summarises the discrepancy by comparing offspring participation rates on sociodemographic characteristics at key time points: before the infant's first birthday; adolescent age 12; and adolescent age 16.5. At baseline, 13.7% of the mothers had degrees; 68% were in their first marriage; 97.4% were white; their mean age was 28 years; and 51.4% of their children were male. The sample was notably different by the time the adolescent was 16.5 years old: the percentage of mothers who had degrees at baseline increased to 21% (+7.3%); the percentage of mothers who were in their first marriage at baseline increased to 76% (+8%); ethnicity remained largely the same at 96.6% white (-0.8%); their mean age at the time of birth had increased to 29.4 years (+1.4 years); and the percentage of males had decreased to 40% (-11%). On average across these characteristics, the retention rate from baseline was 34.2% at adolescent age 16.5 years.

Table 7-2 (above) and Table 7-3 (below) together indicate that ALSPAC is subject to patterns of attrition that are common to longitudinal data surveys, in that those more likely to be retained are more affluent and have better interpersonal support networks and material resources (Schmidt and Woll, 2017, Wolke et al., 2009) (e.g. female participants whose mothers were university educated, owned their own homes, were older at baseline, and were in their first marriage). Thus, the socio-economic differences between the ALSPAC sample and the UK population at baseline are exacerbated by the socio-economic patterning of sample attrition. Another concern for generalisability is that polygenic risk of mental illness measured at baseline in ALSPAC has been shown to be associated with later attrition (Munafò et al., 2017). Given the widely understood associations between mental health, socio-economic position, and alcohol use, this could imply that the attrition rate in ALSPAC is problematic for studies focusing on alcohol use, such as this thesis. This is the collider or selection bias phenomenon introduced in Chapter 2.

Table 7-3: Sample characteristics in ALSPAC at adolescent age 0, 12 and 16.5 years old, and in imputed data

Variable	Birth		12 years old		16.5 years old		Imputed data	
	Participation (% of 15,445)	Descriptives	Retention (% of first measurement)	Descriptives	Retention (% of first measurement)	Descriptives	Retention (% of first measurement)	Descriptives
Infant sex	14,854 (96.2)	51.4% male	6616 (44.5%)	50.2% male	4769 (32.1%)	40% male	7959 (51.5%)	48.3% male
Maternal ethnicity	12,246 (79.3)	97.4% white	6238 (51.0%)	97.2% white	4469 (36.5%)	96.6% white	7959 (51.5%)	96.2% white
Maternal education	11,717 (76.9)	13.7% have degree	6034 (51.5%)	18.2% have degree	4310 (36.8%)	21% have degree	7959 (51.5%)	17.3% have degree
Maternal age	14,076 (91.1)	mean = 28	6361 (45.2%)	mean = 29.4	4558 (32.4%)	mean = 29.4	7959 (51.5%)	mean = 29.1
Maternal marital status	13,559 (87.8)	68% in first marriage	6250 (46.1%)	75.7% in first marriage	4472 (33.0%)	76% in first marriage	7959 (51.5%)	75% in first marriage

The MICE procedures described in Chapter 3 were thus used to improve the external validity of the data used in this thesis. Table 7-3 also compares participation rates at key time points to the imputed data. It indicates that the MICE processes were largely successful in making the analytical sample more representative of the baseline sample, most notably the shifts from 40% male to 47%, and from 21% having a degree to 16.4%. This goes some way towards reducing the limitations of the data's generalisability. Indeed, as noted in Chapter 5, the Visser et al systematic review argued that high attrition rates were an important source of bias in the studies they reviewed and that future studies on parental influences on adolescent alcohol harm should do more to reduce this problem. Please see Appendix C for further descriptives of the imputed data.

7.3 Using ALSPAC to measure the conceptual I-DAG

This section describes the variables selected to replace the concepts in the conceptual I-DAG, thus generating the data I-DAG (i.e. by replacing conceptual nodes with variable nodes). It includes descriptive summaries and unadjusted associations between variables. Due to the volume of concepts and variables under discussion, the outcome and parental influences, as the substantive focus of this thesis, are given more space here than the confounders and intermediates. However, before describing specific measures in any detail the following paragraphs explicate how the timing of data collection in ALSPAC fundamentally shaped the data I-DAG, and the implications this has for the analysis plan set out in Chapter 3.

The timing of the outcome measurement was key in developing the data I-DAG. Under the rules of d-separation discussed in Chapter 2, no variable measured after the age of 16.5 years could be included in the analyses in this thesis. As the focus of investigation is on estimating the effect of parental influences during adolescence, this meant that all parental influences had to be measured between the onset of adolescence and the age of 16.5 years. This was an especially pertinent issue for mediation modelling. Box 7-1 below is a summary of the analytical plan included here for ease of reference. After the age of 12 years (child), the ALSPAC parents did not complete any surveys on their own behaviour until the adolescent was aged 17. As such, age 12 was the only valid timepoint with data on self-report of parental behaviour (drinking, smoking, substance use, etc.). An implication of this was that the parent-child relationship and parenting (as reported by the child) would not be measured before parental behaviour, limiting the possibility that they could be posited as causes of parental characteristics in the data I-DAG. However, with few notable exceptions, parenting and the parent-child relationship during adolescence were mostly assessed at age 15.5 via adolescent self-report. This necessitated a mediation model in which parenting and the parent-child relationship mediate the

effect of parental behaviour. An additional consequence in terms of both mediation analysis and the average effects analysis is that the effect of parental behaviours cannot be estimated while controlling for parenting and the parent-child relationship if the temporal ordering of the data is to be respected. As identified in the review of systematic reviews (Chapter 5) and ESC-DAGs review 1 (Chapter 6), the parental characteristic which has the strongest evidence in the literature as a cause for adolescent alcohol harm is parental drinking (asides from parental supply of alcohol). Accordingly, parental drinking was identified as the parental behaviour to act as the exposure of interest for the mediation analysis, with parenting and the parent-child relationship as the mediators (as noted below, parental supply of alcohol was not measured in ALSPAC). Thus, the analysis plan to investigate mediation within parental influences as a conceptual domain was updated to investigate parental and non-parental mediators of parental drinking specifically.

Box 7-1: Summary of thesis analysis plan

> The analysis plan for this thesis was set out in Chapter 3. The first stage is to conduct average effects analysis on a primary group of exposures (i.e. each parental influence) and a secondary group of exposures (i.e. non-parental domains of the socio-ecological model). The results from these analyses will then inform the subsequent mediation models. This will involve two general sets of mediation models; one set which investigates mediation within parental influences as a conceptual domain; and one set which investigates mediation of parental influences via the non-parental domains of the socio-ecological model.

Note that the members of the birth cohort itself will mostly be referred to as the 'adolescent(s)', regardless of their age. The main exception is the parent-child relationship, which maintains its nomenclature for consistency. Note also that all reference to specific ages (e.g. variable *v* measured at *y* years old) pertains to the age of the adolescent in years and not the mother or partner unless otherwise stipulated

7.3.1 Outcome measure

The outcome of adolescent alcohol harm is measured at age 16.5 using the Alcohol Use Disorders Identification Test (AUDIT) (de Meneses-Gaya et al., 2009, Saunders et al., 1993). The timing at age 16.5 years was in keeping with the impetus of investigating underage adolescent drinking. An AUDIT questionnaire is included as Appendix G for reference. AUDIT is commonly used both in clinical practice and in research. It is a 10-item questionnaire, each with a minimum score of 0 and maximum of 4, with a resulting range of 0 to 40. The first three questions pertain to alcohol

consumption. They can be used as part of an individual's total AUDIT score or excerpted in the form of the 'AUDIT-C' scale, thus focusing exclusively on alcohol consumption. The other seven questions are concerned with alcohol harm, including indicators of dependence. An AUDIT score conceptualises an individual's drinking across a spectrum of risk - it does not focus on alcohol abuse, alcohol dependence, or alcohol use disorders (a medical diagnosis that subsumes alcohol abuse and dependence) to the exclusion of low risk drinking. This makes it a useful tool for studies interested in the full distribution of alcohol use and harm, or when the sample is unlikely to have high proportions of high-risk or dependent drinkers, such as the adolescents in this analysis. However, it also has clinically informed and validated thresholds used to characterise alcohol use disorders (Moehring et al., 2019). Thresholds are for low risk alcohol consumption (between 0 and 7 out of 40), hazardous risk (between 8 and 15), and harmful risk (≥16). Scores above 20 indicate alcohol dependence. Figure 7-1 below demonstrates the right skew and the low, hazardous and harmful grouping in the ALSPAC cohort at age 16.5. Categorically, of the 4,769 respondents, 3,030 (63.54%) had low risk scores, 1,457 (30.55%) had hazardous scores, and 282 (5.91%) had harmful scores. Thus, 1,739 (36.5%) had AUDIT scores that were hazardous or harmful. Notably, even at an age when purchasing alcohol is illegal, adolescents had similar probability of being in the harmful use category (N=282, 5.9%) as to score 0 (N=247, 5.2%).

Like with most subjective self-report measures of alcohol use and harm, there are notable limitations to AUDIT:

- Very different drinking profiles can have similar AUDIT scores. For example, it is possible for an individual who binge drinks every weekend to have a similar score to another who drinks 1-2 glasses of wine every weeknight. As these profiles are associated with different types of alcohol harm, similar AUDIT scores can be associated with different outcomes.
- AUDIT quantifies consumption in a way that is less specific than other options. For example, rather than asking for a specific number of drinks, the second question asks "how many drinks containing alcohol do you have on a typical day when drinking", with available responses being 0-2. 3-4. 5-6, 7-9, and 10 or more. This lack of specificity makes AUDIT comparatively weaker at measuring consumption (e.g. units per week compared to official guidelines). This is compounded by how AUDIT, like all self-report measures, depends on respondents sharing somewhat consistent definitions of what constitutes 'a drink'.
- AUDIT questions refer to behaviour over the last year, which has strengths (e.g. less susceptible to short term fluctuations in drinking behaviour) as well as weaknesses (e.g. recall bias).
- While the thresholds of low, hazardous and harmful drinking are clinically informed and are generally found to be valid, there is evidence that their sensitivity may be too low (i.e.

because the thresholds are too high) to detect risky drinking in specific social groups such as women (Bradley et al., 2003), older adults (Moore et al., 2002, Philpot et al., 2003) and adolescents (Reinert and Allen, 2007).

- Other forms of bias are also possible due to reliance on subjective self-report. For example, besides typical issues like audience effects, it has been observed that the sequence in which questions are answered by participants may influence the AUDIT score in that earlier questions will be more likely to be given a higher score than later ones (Bischof et al., 2005).

However, AUDIT has been validated in numerous settings on a range of populations. For example, Reinert and Allen (2007) conducted a review of such studies and found that AUDIT in general out-performs similar tools, concluding that "AUDIT can be used with confidence in a variety of settings". This perception may explain why it is the only available measure of alcohol harm used with the ALSPAC birth cohort. Indeed, ALSPAC transitioned to using AUDIT for parental questionnaires rather than the bespoke quantity-frequency measures administered at baseline. As AUDIT is the only alcohol use scale employed in ALSPAC for the adolescent participants, it is important for the current study that Santis et al (2009) verified the AUDIT scale as suitable for use with adolescents. Overall, AUDIT is a fitting option for this thesis as it is a widely used numerical measure with clinical relevance than has been validated on the population of interest.

The entire distribution of AUDIT is used as the outcome measure for the majority of analyses in this thesis (rather than categorical or binary alternatives). This is justified on several grounds. Firstly, while AUDIT has been verified for use with adolescents despite being originally designed for use with adults (Santis et al., 2009), the risk thresholds have not (as far as the author is aware). A score of 8/40 is the lower threshold of the hazardous risk grouping. Assuming that adolescents are more socially and biologically susceptible to alcohol consumption relative to adults, this threshold could be lower. Alternatively, they could be less susceptible as they are less likely to be dependent or to have comorbid medical issues. Therefore, while the categorical outcomes are informative, there is greater uncertainty in their interpretation compared to the numerical measure. Secondly, the statistical methods used in this thesis, especially the causal mediation analysis and propensity score-based methods, perform much better with numerical outcomes. Third, and crucially, it is the normality of the distribution of residuals that is assumed for linear regression, not the normality of the outcome measure. Regressing AUDIT on maternal drinking is the underlying model for analysis in this thesis, and as Figure 7-2 below demonstrates, the residuals are mostly normally distributed. Finally, again as explained in Chapter 3, the propensity score-based methods used here are doubly robust.

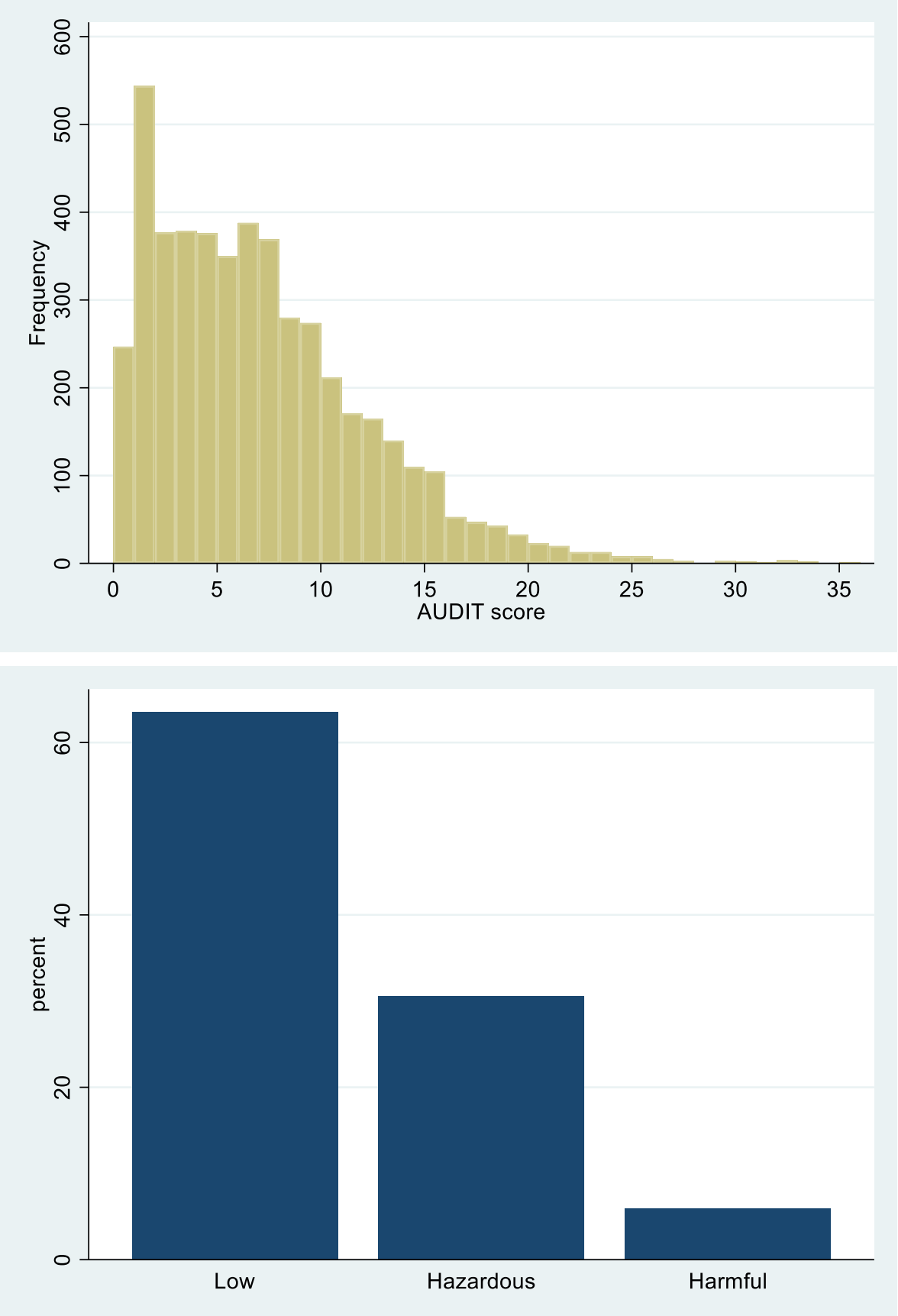


Figure 7-1: Distribution of AUDIT scores at age 16.5 years

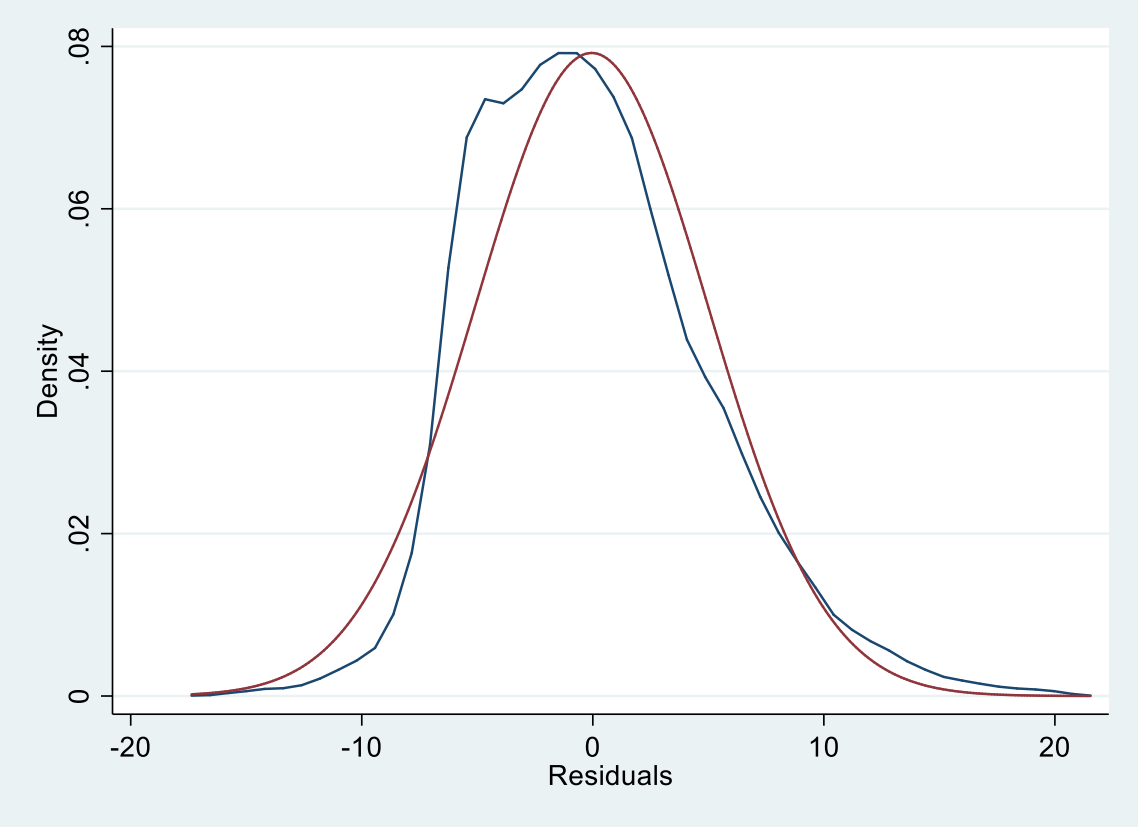


Figure 7-2: Distribution of residuals for maternal drinking on AUDIT score

7.3.2 Using ALSPAC to measure the parental influences

Of the 14 parental influences identified in the review of systematic reviews and ESC-DAGs review, five were not measured sufficiently in ALSPAC: parental supply of alcohol; alcohol-specific communication; parental support; family coherence/conflict; and parental substance use. The nine parental concepts that were adequately measured in ALSPAC were: parental drinking; parental smoking; parental mental health; alcohol-specific rules; parent-child relationship; parental monitoring; parental discipline; parental involvement; and general communication. They are summarised in Table 7-4.

Each variable is described in some detail below, including: the process of developing binary variables from the raw data; basic descriptives of the recoded variables focusing mainly on the proportions of exposed and unexposed; and unadjusted associations with AUDIT scores at age 16.5. The latter were calculated by regressing each parental influence on the outcome using the 'complete case' data (CC) - i.e. only participants who responded to both the outcome and the exposure are included. This was performed using linear regression on the full AUDIT scale. Odds ratios for the adolescent having an AUDIT score of 8 or higher (hazardous or harmful risk) were also calculated using logistic regression. Each parental influence was found to have an unadjusted association with AUDIT scores. Table 7-5

shows sample sizes and unadjusted associations. As noted in Chapter 3, binary variables are preferred because they have stable IPWs, because of the high volume of exposures being analysed, and because of the lack of normally distributed continuous exposures.

Table 7-4: Summary of parental influences and the ALSPAC variables used to measure them

Parental influence	ALSPAC variable	Exposure status	% Exposed
Parental drinking	Maternal drinking	Mother drank > 14 units of alcohol last week	30.5
Parental smoking	Maternal smoking	Mother smokes	17.2
Parental mental health	Maternal depression	Mother had depression in last two years	22.5
Alcohol-specific rules	Parental permissiveness	Adolescent has tried alcohol with parents' permission	7.4
Parental discipline	Parental punishment	Adolescent is punished "often" or "very often"	33.2
Parental monitoring	Parental monitoring	Parents know less about how adolescent spends time	28.7
General communication	Secret-keeping	Keeps secrets from parents "often" or "very often"	26.7
Parental involvement	Time shared at weekends	Spends less than an hour with parents on weekends	31.3
Parent-child relationship	Parent-child tension	Argues with parents "most days"	25.2

Parental Behaviours: parental drinking

Parental drinking was measured using maternal data, as the participation rates of the partners (N=3,196) were much lower that of the mothers at age 12 years (N=6,616) at around 48%. The mothers were asked how many drinks they had on each day 'last week' across several different beverage types: half pints of beer; glasses of wine; measures of spirits; measures of sherry; ready-mixed beverages; and low strength beverages. Compiling days into weeks showed that wine was by far the most popular beverage, with a median of 2 glasses (and interquartile range of 6 glasses) compared to a median of 0 for each of the rest (the next highest interquartile range was 1 for half pints of beer). Each beverage had right skewed distributions, most notably for wine with a 95th percentile of 15 glasses and 99th percentile of 23. These weekly variables were converted into UK standard units (8g alcohol) informed by NHS alcohol support guidance available at https://www.nhs.uk/live-well/alcohol-support/calculating-alcohol-units/.

Table 7-5: Summary of relationship of parental variables to AUDIT score

Parental influence	N for variable	Age	Exposure status	N for CC regression	OLS Coefficient	Odds Ratio
Maternal drinking	6,604	12	Mother drank > 14 units of alcohol last week	3,800	1.18	1.57
					(0.83, 1.53)	(1.36, 1.80)
Maternal smoking	6,851	12	Mother smokes	3,905	1.06	1.48
					(0.60, 1.51)	(1.24, 1.78)
Maternal mental health	6,954	12	Mother had depression in last two years	3,955	0.62	1.28
					(0.23, 1.02)	(1.09, 1.50)
Parental permissiveness	4,202	12.5	Adolescent has tried alcohol with parents' permission	2,341	1.57	1.76
					(0.69, 2.47)	(1.22, 2.54)
Parental punishment	5,348	15.5	Adolescent is punished "often" or "very often"	3,461	1.21	1.55
					(0.85, 1.57)	(1.34, 1.79)
Parental monitoring	5,387	15.5	Parents know less about how adolescent spends time	3,484	3.07	2.76
					(2.70, 3.43)	(2.37, 3.22)
Secret-keeping	4,553	15.5	Keeps secrets from parents "often" or "very often"	3,014	1.93	1.86
					(1.52, 2.35)	(1.57, 2.21)
Parental involvement	5,376	15.5	Spends less than an hour with parents on weekends	3,476	1.60	1.72
					(1.22, 1.97)	(1.48, 2.00)
Parent-child tension	5,370	15.5	Argues with parents "most days"	3,476	1.59	1.65
					(1.19, 1.98)	(1.41, 1.93)

Glasses of wine were weighted as 2 units, half pints of beer as 1 unit, measures of spirts and ready-mix drinks as 1.4 units, measures of sherry as 1 unit, and low strength drinks as 0.5 units. Units per week for each beverage were then compiled to equal the total number of units consumed last week for the mothers. The UK Chief Medical Officer's (CMO) recommendation for low risk female alcohol consumption is 14 units or less per week (Department of Health, 2016). Of the 6,604 mothers 2,014 drank above 14 units (30.5%); 1,356 (20.5%) drank nothing; and 3,234 drank between 1 and 14 units (49%). Figure 7-3 below describes the numerical and categorical distributions graphically.

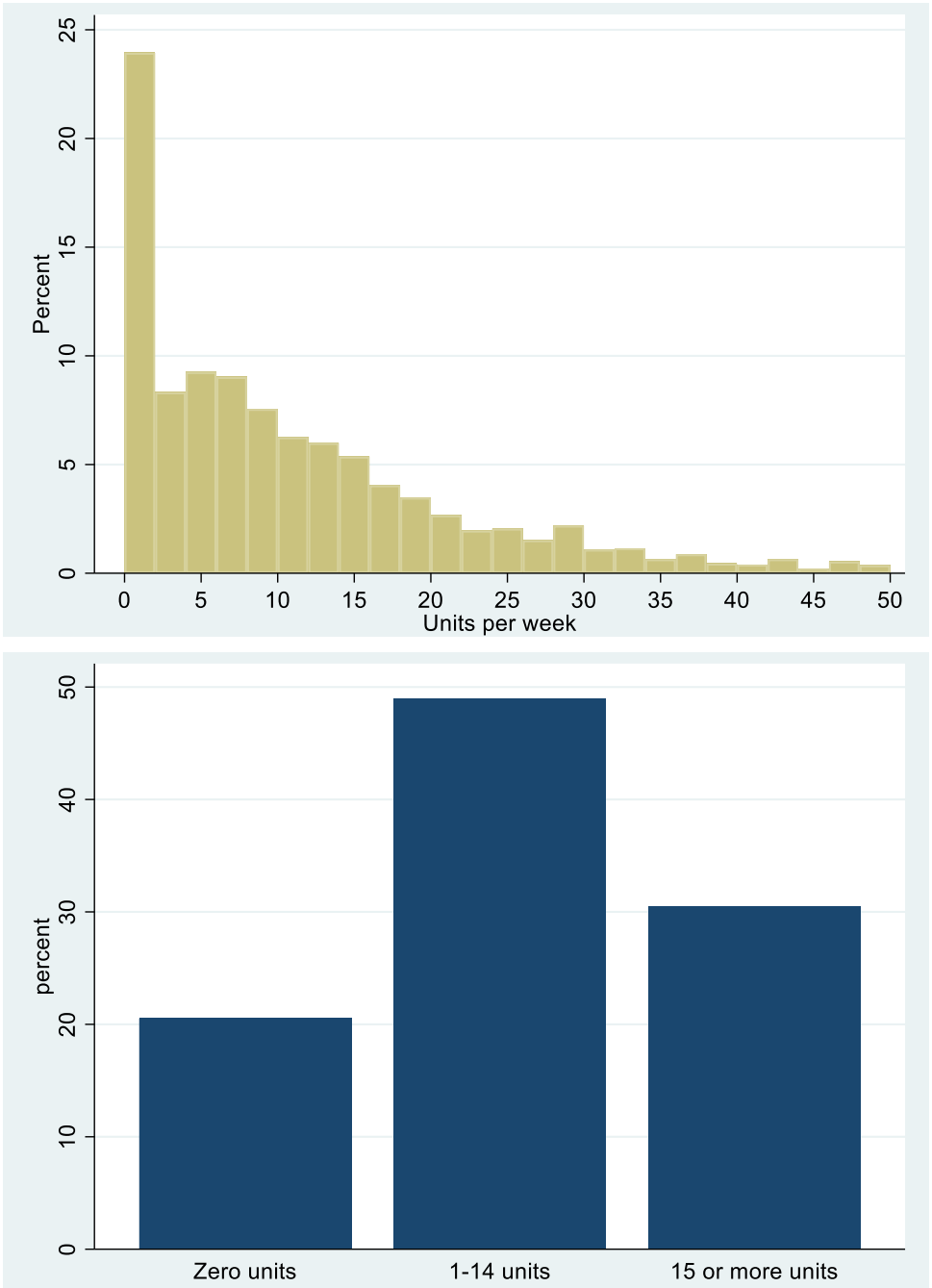


Figure 7-3: Distribution of maternal drinking in units per week

The total number of units was then dichotomised around the CMO recommendations such that 15 or greater units was the exposure. Maternal drinking was found to have a harmful unadjusted association with AUDIT scores, with adolescents whose mothers drank more than 14 units per week having an increased AUDIT score by 1.18 points on average with a 95% confidence interval between 0.83 and 1.53. Adolescents whose mothers drank more than 14 units per week were also 1.57 (1.36, 1.80) times more likely to have hazardous AUDIT scores at age 16.5.

The measure of father's drinking at adolescent age 12 years was the same as for the mothers (units consumed last week). However, the participation rate was decided to be too low to allow for the variable to be included in the imputations as around 60% of the data would have been imputed – 10% more than the variable with the second highest level of missing data (alcohol-specific rules). Multiple imputation was discussed in Chapter 3, and Appendix A. Nonetheless it was possible to use father's drinking for some CC analysis in Chapter 8. The raw data was treated in the same way as for the mothers. While the comparability of the samples is caveated by the participation differential, it was unsurprising to find that fathers tended to drink more than mothers – the median for fathers was 13 compared to 8 for mothers. Of the 3,192 cases, 327 consumed zero units (10.2%), 1,287 consumed between 1 and 14 units (40.3%), and 1,578 consumed 15 or more (49.4). The CMO guidelines for men changed in 2016 to be the same as for women at 14 units. However, a total of 994 (31.14%) fathers consumed over the previous CMO guidelines of 21 or fewer units of alcohol per week, which were the guidelines at the time of measurement. Adolescents whose fathers drank 15 or more units were 1.69 times (1.41, 2.03) as likely to have hazardous AUDIT scores. The odds ratio went up to 1.9 (1.57, 2.30) under the previous guidelines of 22 or more units.

The maternal units-per-week exposure presents implications. While the right skew indicates problematic alcohol behaviour for some mothers, the nature of these behaviours, and the biological and social consequences they present, can vary greatly. The right skew could represent individuals who: engaged in a single occasion of heavy episodic drinking with acute consequences; have clinical alcohol dependence with chronic consequences; or who are 'home drinkers'. The latter is a term that has been used to describe how typically higher SES individuals consume alcohol (wine especially) on a daily basis without ever consuming enough to become intoxicated, as a "source of cultural capital" (Brierley- Jones et al., 2014) (thus risking lagged chronic consequences such as or various forms of cancer). This is especially notable given that this form of drinking is associated with higher socio-economic strata, and that the ALSPAC participants data are relatively affluent.

Parental behaviours: parental smoking

For similar reasons as the above, maternal smoking at age 12 was used to cover parental smoking. This was a continuous variable counting the number of cigarettes smoked on the average day. The

variable was dichotomised into 'smoker' (N=1,177, 17.2%) and 'non-smoker' (N=5,674, 82.8%). Of those who smoked, the median amount was 10 per day, and the 95th percentile was 20. Maternal smoking at adolescent age 12 had a harmful unadjusted association with AUDIT scores at age 16.5. Adolescents whose mothers smoked had AUDIT scores 1.06 higher on average (0.60, 1.51), and were 1.48 times (1.24, 1.78) as likely to have hazardous or harmful AUDIT scores.

Parental mental health: depression

Mothers were asked at age 12 if they have been depressed in the last two years, and if so, had they consulted a doctor. The variable was dichotomised to compare having depression in the last two years (N=1,750, 22.3%) to not (N=5,396, 77.7%). Maternal depression had an unadjusted associated with AUDIT score at age 16.5 of 0.62 (0.23, 1.02) and adolescents whose mothers had experienced depression in the last two years were 1.28 times (1.09, 1.50) as likely to have hazardous or harmful AUDIT scores. It is worth noting that there was little variation in AUDIT score for adolescents whose mother had depression and consulted a doctor and those who had depression and did not consult a doctor. The proportion with hazardous or harmful AUDIT scores was 57% (227/395) for adolescents whose mother had consulted a doctor, and 61% (251/413) for those who had not.

Parenting: alcohol-specific rules

Parental alcohol-specific rules were measured in response to the binary question "have you ever tried alcohol with your parents' permission?" This was asked at ages 12.5 and 13.5 with low variation for both - at age 12.5 only 7.4% answered "no" (312/4,195), and less than 5.7% at age 13.5 (350/6,110). Age 12.5 was selected to leverage the higher variation and because this variation reflected how a number of those answering 'no' at age 12.5 had changed to 'yes' in the interim (N=198). Adolescents who had tried alcohol with their parents' permission had higher AUDIT scores of 1.57 on average (0.69, 2.47), and were 1.76 times as likely to have hazardous or harmful AUDIT scores (1.22, 2.54).

Parenting: discipline

Parental discipline was measured at age 15.5 years by asking the adolescent how often their parents punished them in different ways. A total of 5,348 responses were recorded. The 5 questions were answered on a Likert scale. Types of punishment were: "telling off"; "grounding them"; "stopping pocket money"; "stopping them from seeing friends"; and "in some other way". Unfortunately, because smacking is assumedly subsumed by the "in some other way" category, it was not possible to use this variable as a proxy for child abuse, which was not measured (see below). Variables were compiled and further recoded such that adolescents who were punished very often or often in any way were assigned as the exposure (N=1,774, 33.2%). More frequent parental punishment at age 15.5 had a harmful association with AUDIT scores at age 16.5. AUDIT scores were 1.21 (0.85, 1.57) higher

on average for adolescents who were punished often, and they were 1.55 times as likely to have hazardous AUDIT scores (1.34, 1.79).

Parenting: monitoring

Parental monitoring was measured at age 15.5 by asking the adolescents how often their parents had information on how they spend their time away from home. There were 4 questions asking how often parents knew who they were with; where they were going; what they were doing; and what time they were supposed to come home. Rather than Likert scales, the 4 available responses were "never", "sometimes", "usually", and "always". Each variable was coded such that "never" was 0, "sometimes was 1", "usually" was 2, and "always" as 3. They were then compiled into one variable with scores ranging from 0 to 12, with a strong left skew. This variable was then dichotomised such that lower scores (between 0 and 7) were treated as the exposure group (N=1,545, 28.7%). Adolescents in the low monitoring group had AUDIT scores 3.07 higher on average (2.70, 3.43), and were 2.76 times as likely to have hazardous AUDIT scores (2.37, 3.22).

Parent-child relationship: parent-child communication

The closest conceptual match to general communication was disclosure of secrets from the adolescent to their parents. Adolescents were asked at age 15.5 how often they kept secrets. Responses were again measured on a Likert scale and recoded into often (N=1.208, 26.7%) vs not often (N=3,325, 73.4%). Higher frequencies of secret-keeping had harmful associations unadjusted with AUDIT scores. Adolescents in the high frequency group had AUDIT scores 1.93 higher (1.52, 2.35), and were 1.86 times as likely to have hazardous AUDIT scores (1.57, 2.21).

Parent-child relationship: parental involvement

Parental involvement was measured at age 15.5 years by asking the adolescent how many hours they spent with their parents on the average weekend. Options were "None" (7.9%), "Up to 1" (22.3%), "Up to 2" (28.1%), "Up to 4" (25.6%), and "Greater than 4" (16.1%). The variable was dichotomised such that spending an hour or less with parents each weekend was treated as the exposure (N=1,680, 31.3%). Adolescents in this group had AUDIT scores 1.60 higher on average (1.22, 1.97), and were 1.72 times more likely to have hazardous AUDIT scores (1.48, 2.00).

Parent-child relationship: parent-child tension

Finally, parent-child tension was measured by asking the adolescent how often they argue with their parents across 6 different topics: "tidiness of room"; "what they do when they go out"; "what time they come home"; "who they go out with"; "what clothes they wear"; and "other things". Responses were "hardly ever/never"; "less than once per week; "at least once per week"; and "most days". Arguing with parents on any topic "most days" was treated as the exposure (N=1,355, 25.2%).

Adolescents in this group had AUDIT scores 1.59 higher on average (1.19, 1.98), and were 1.65 times as likely to have hazardous AUDIT scores (1.41, 1.93).

7.3.3 Suitability of measures of parental influences

There are two measures that are potentially problematic. The first is the alcohol-specific rules variable, measured by asking if the adolescent had tried alcohol with their parents' permission by age 12.5 years. While the variable captures the concept neatly, the low levels of variation could be problematic for analysis. For example, when calculating the average treatment effect of parental permissiveness towards adolescent alcohol use, if very wide confidence intervals are returned there is no way to tell how much of the uncertainty is due to the low variation in the exposure, and how much is due to lack of effect. Whether or not this problem is realised depends on the results. Further, while it is certainly possible to build a balanced propensity score for an exposure with low variation, it is more challenging to do so.

The other parental influence that was potentially problematic was general communication. Firstly, it overlapped with the parental monitoring variable. This issue was touched upon in Yap et al's 2017 systematic review, in which they noted that isolating general communication as a construct is very difficult due to the innate role communication plays in parenting and in aspects of the parent-child relationship. Secondly, 'secret-keeping' does not offer full conceptual coverage of general parent-child communication. Arguably then this measure does not capture 'general communication' but rather a specific aspect of communication. Note however that the use of the term 'general' in this instance is primarily to differentiate communication into alcohol-specific communication, which was unmeasured, and 'other' communication. Indeed, the concepts of adolescent disclosure and secret-keeping were used to cover general communication in three of the 50 primary studies used to build the conceptual I-DAG (McCann et al., 2016, Cable and Sacker, 2008, Raninen et al., 2018). As such this variable was retained as the best available option for analysing parent-child communication, due to its relationship to the concept of interest, and precedent in the literature.

7.3.4 Unmeasured parental influences

The lack of measurement for five of the 14 parental influences was concluded to be mostly unproblematic as each has at least some degree of conceptual overlap with the nine parental influences that were measured. The clearest examples are the related unmeasured concepts of parental support and family conflict. Notably, Ryan et al used the quality of the parent-child relationship to subsume parental support; family conflict; parental involvement (measured) and

parent-child tension (measured). Parental support, and family conflict can thus be argued to receive some degree of conceptual 'cover' from parental involvement and parent-child tension. Parental substance use arguably overlaps with parental smoking and parental drinking, acknowledging the distinction between illicit and commercial substances. Perhaps less convincing are the unmeasured concepts of alcohol-specific communication and parental supply. While both are clearly related to alcohol-specific rules (measured as having tried alcohol with parental permission at age 12.5), none of the reviews considered them as being within the same conceptual grouping. Further, alcohol-specific rules are noted above as one of the more problematic measures. Indeed, given that one of the key findings of the review of systematic reviews was that parental supply was a risk factor with convergent evidence of relatively strong effect, lack of any direct measurement of this concept limits subsequent analysis.

7.3.5 Using ALSPAC to measure the non-parental influences (intermediates)

While the intermediate variables are not as central to this thesis as the parental influences, they are important in terms of intermediate confounding and as mediators. As such their measurement and relationship to the outcome is discussed here in some detail. The ALSPAC data offered very good coverage of the intermediates: of the 22 concepts posited in the conceptual I-DAG only three were not measured (beliefs, economic activity, and extended family effects). Two more were only measured before the exposure (adolescent BMI and sibling effects) and were thus reassigned as baseline confounders. In terms of the quality of the measurements used, intra-personal and interpersonal effects had much better matches than the institutional effects of school and neighbourhood. Table 7-6 below summarises pertinent information on the intermediate variables, including exposure status, proportion exposed, age at measurement, sample size, and unadjusted associations with the outcome. Each of the non-parental intermediates that were measured had harmful unadjusted associations with AUDIT at 16.5 years of age, except for IQ which had a protective unadjusted association. The below discusses measurement of the intermediate variables very briefly. Box 7-2 below briefly summarises intermediate variables for ease of reference.

Box 7-2: Intermediate variables reminder

As noted in Chapter 2 and Chapter 3, intermediate variables have specific profiles defined by the directed edges between them and the exposure, other intermediates, and the outcome. For any intermediate variable, directed edges must either be received from the exposure, or not exist between it and the exposure. A variable which 'sends' a directed edge to the exposure is not an intermediate but a confounder. Intermediates which do not receive an edge from the exposure are mediator-outcome confounders. Intermediates which do receive an edge from the exposure are defined in terms of how they relate to the mediator of interest. If we want to estimate the effect through an intermediate which receives a directed edge from the exposure, then it is of course the mediator. If an intermediate which receives a directed edge from the exposure in turn sends a directed edge to the mediator, then it is an exposure-induced mediator-outcome confounder (EIMOC). If it receives a directed edge from the mediator and the exposure then should not be factored in the model at all. These are sometimes referred to as 'nuisance mediators' (Dunn et al., 2015).

7.3.5.1 Intra-personal effects

The conceptual I-DAG identified the following intrapersonal intermediates: attitudes to alcohol; early alcohol initiation; smoking; other substance behaviours; risk-personality; general personality; beliefs; general behaviour; mental health; physical health; and economic activity. Most were measured adequately in ALSPAC. Attitudes to alcohol were measured at age 14 by asking the adolescent how 'harmful or helpful' alcohol consumption is to physical and mental wellbeing. Early alcohol initiation was measured at age 12.5 by asking the adolescent how old they were when they first had a whole drink of alcohol. Smoking was measured at age 13.5 by asking the adolescent if they had ever smoked. Other substance behaviour was measured at age 14 by asking adolescents if they had ever tried any of a comprehensive list of illicit substances. Risk personality was measured using the Arnett's sensation-seeking scale at age 13.5, which produced a score of sensation seeking tendency. General personality was measured using IQ at age 15.5. This measure was the closest match to general personality, as other questions tended to overlap strongly with mental health (e.g. psychosis). General behaviour was measured as anti-social behaviour at age 14 by asking if the adolescent had ever engaged in one of several anti-social behaviours (theft, vandalism, verbal assault, etc.). Mental health was measured using the Short Mood and Feelings questionnaire at age 13.5 years, giving a 'score' for current depressive symptoms. Physical health was measured using BMI and pubertal timing. However, BMI was measured before maternal drinking and so was reassigned as a confounder. Pubertal timing was measured by assessing the adolescent's development of body hair at

age 13. The only intra-personal intermediate concepts that were not measured before the outcome at age of 16.5 were economic activity and beliefs.

7.3.5.2 Interpersonal effects

Interpersonal intermediates suggested by the conceptual I-DAG were: number of friends; peer drinking; peer smoking; peer substance use; peer anti-social behaviour; sibling effects; relationships with teachers; and extended family influences. No information on the latter was available in the ALSPAC data. Number of friends was measured at age 13.5 by asking the adolescent how many 'close friends' they had. Peer drinking and peer smoking were also measured at age 13.5 by asking if the adolescent's friends had smoked tobacco or drank alcohol without parental permission. Peer substance use was asked at age 14 years by asking if the adolescent's friends had tried any of the same list of illicit substances asked of the adolescent as above. Similarly, peer anti-social behaviour was measured in the same way as the adolescent's own, by asking if the adolescent's friends had engaged in one of several anti-social behaviours. Sibling effects could only be measured by 'number of older siblings' at age 7, which was reassigned as a confounder. Relationships with teachers were measured by asking how much the adolescent agreed that their teachers were "OK" at age 16.5. As such, it cannot be used as an EIMOC for any of the other mediators, as it is measured after them.

7.3.5.3 Institutional effects

School engagement was measured by asking the adolescent at age 14 how strongly they agreed on a Likert scale with the statement that school is a "place where they really like to go each day". Academic performance was measured using number of GCSEs between A* and C (taken in year 12 when pupils are aged between 15 and 16). Neighbourhood attachment was measured by asking the adolescent how safe they usually felt in their neighbourhood. The only one of these three variables that is not problematic is school enjoyment. Number of GCSEs is an ideal fit for academic performance conceptually, but the timing of its measurement was after all parental exposures. Similar to relationships with teachers, this meant it could not be treated as an EIMOC, but rather only as a potential mediator. Neighbourhood safety does not offer adequate coverage of community engagement. Nonetheless, both of these variables were retained. No non-parental family effects were used as intermediates – sibling effects, family structure, family transitions, and family history of alcohol were all assigned as confounders.

Table 7-6: Summary of intermediate measures

Concept	Variable	Exposure	Age	N	% Exposed	Unadjusted OR (95% C.I.)
Intra-personal						
Early alcohol	Early alcohol initiation	Before age 10	12.5	3,765	29.1	1.31 (1.09, 1.58)
Risk personality	Sensation seeking	Upper quartile on Arnett's sensation seeking scale	13.5	6,045	30.0	2.23 (1.76, 2.37)
Mental health	Depression	Score from Short Mood and Feeling questionnaire, 1= >8	13.5	6,078	28.7	1.40 (1.20, 1.62)
Attitudes towards alcohol	Positive alcohol attitudes	Does not think regular alcohol consumption or binge drinking are "very harmful"	14	6,089	30.1	1.68 (1.50, 1.93)
General behaviour	Anti-social behaviour	Has engaged in anti-social behaviour (e.g. theft, vandalism)	14	6,178	36.3	2.40 (2.10, 2.74)
Smoking	Smoking	Has smoked	14	5,908	25.9	2.94 (2.53, 3.42)
Substance use	Substance use	Has tried an illicit substance	14	5,911	20.2	2.51 (2.14, 2.94)
General personality	IQ	+1SD on IQ score	15.5	4,370	19.8	0.77 (0.65, 0.92)
Other interpersonal						
Number of friends	Number of close friends	7 or more	13.5	6,108	22.8	1.33 (1.13, 1.56)
Peer smoking	Peer smokers	Friends have smoked cigarettes without parental permission	13.5	5,297	55.1	2.44 (2.12, 2.81)
Peer drinking	Peer drinkers	Friends have drunk alcohol without parental permission	13.5	5,501	49.3	2.57 (2.22, 2.98)

Peer anti-social behaviour	Peer anti-social behaviour	Friends have engaged in anti-social behaviour (e.g. theft, vandalism)	15.5	5,357	45.5	3.04 (2.64, 3.51)
Peer substance use	Peer substance use	Friends have tried an illicit substance	15.5	5,348	53.9	3.21 (2.77, 3.73)
Relationships with teachers	Relationships with teachers	Does not "strongly agree" that most of their teachers are "OK"	16.5	5,284	63.2	1.65 (1.43, 1.90)
Institutional						
Neighbourhood attachment	Neighbourhood safety	Adolescent usually doesn't feel safe in their neighbourhood	14	5,449	6.7	1.55 (1.20, 2.01)
School engagement	School engagement	Adolescent does not like attending school	14	5,552	35.2	1.24 (1.08, 1.43)
Academic performance	Academic performance	>8 GCSEs grade A*-C	16	12,064	62.4	1.23 (1.08, 1.40)
Reassigned as confounders						
BMI	BMI	BMI at age 11	11			
Siblings effects	No. of older siblings	Has older siblings	7			

7.3.6 Using ALSPAC to measure the baseline confounders

The baseline confounders suggested in the conceptual I-DAG were age of the adolescent, ethnicity, adolescent sex, adolescent sexuality, physical health (especially birthweight and foetal substance exposures), maternal age, parental socio-economic status, parental education, experience of abuse, family structure, family history of alcohol problems, neighbourhood socio-economic status, neighbourhood availability of alcohol, neighbourhood space, and urbanity. Of these, only neighbourhood availability of alcohol, neighbourhood space, and experience of abuse were not available. The former is a notable limitation, especially given that there was no measure of interpersonal alcohol availability or supply. Each of the other confounders was measured adequately before the exposure at age 12 years. The exception was adolescent age, in that it is captured by the longitudinal nature of the data, rather than as a variable which can be modelled on an outcome, etc. As mentioned above, adolescent BMI and sibling effects were reassigned from intermediates to baseline confounders due to the timing of measurement at 11 and 7 years old respectively.

Several baseline confounders did not have associations with AUDIT score in unadjusted regression models: sex; urbanity; maternal age; ethnicity; maternal religiosity; and neighbourhood deprivation. In the case of ethnicity, this may be explained by the low variation intimated above – only 3.7% of the adolescents who completed the AUDIT questionnaire were not white. Sex and maternal religiosity were the only baseline confounders that were also not associated with the exposure of maternal drinking. An additional variable of neighbourhood effects was also measured by asking the mother what her impression was of her neighbourhood when the adolescent was aged eight years old, but this was also not associated with the outcome or exposure. Table 7-7 below summarises the timing and measurement of the baseline confounders. Unlike the non-parental intermediates, the effects of the baseline confounders are never estimated during any of the causal analyses, meaning they did not need a dichotomous version.

In some cases there were several measures that pertained to the concept of interest in some way. For adolescent mental health, for example, the Short Mood and Feelings questionnaire at age 13.5 and also a 'life events' scale at age 16. In such cases, there were four characteristics that were compared to identify which measure to integrate into the data I-DAG: presence of an unadjusted association with the outcome; size of the unadjusted association with the outcome; sample size; and timing of the measurement. This last criterion operated differently depending on whether the variable pertained to parenting or the parent-child relationship (in which case close to the outcome were preferred, thus allowing for mediation models with greater control of EIMOCs); another intermediate (in which case close to the exposure were preferred; thus emphasising their role as EIMOCs over mediators) or a baseline confounder (again, close the exposure was preferred). There were no ties for any concept

when weighting these criteria equally. For example, the life events scale had a similar unadjusted association with the outcome as did the Short Mood and Feelings questionnaire, but it had a lower sample size and was measured too late to be used as an exposure-induced mediator-outcome confounder (EIMOC). Concepts put through this process were mental health, anti-social behaviour, school engagement, neighbourhood engagement, parental monitoring, parental involvement, social class, and maternal mental health. Very few concepts were measured on multiple occasions using the same questions, limiting the scope for repeated measures analysis or time-varying confounding analysis, for example.

Table 7-7: Summary of measures for baseline confounders

Concept	**Measurement**	**Age (years)**
Age of adolescent	Inherent in data	-
Maternal age	Measured in years at birth	0
Adolescent sex	Measured at birth	0
Adolescent's weight at birth	Measured in grams	0
Maternal smoking during pregnancy	Smoker vs non-smoker	<1
Ethnicity	White vs non-white	<1
Maternal SES	Maternal NS-SEC	<1
Maternal education	Highest level of education	<1
Family structure	Mother in first marriage vs other	<1
Family history of alcohol problems	Parental/grandparental alcohol dependence	<1
Maternal stress	Number of stressful life events in last two years	<1
Older siblings	Has older siblings	7
Urbanity	Lives in a city, town, village, or hamlet	10
Neighbourhood effects	IMD quintiles	10
Maternal religiosity	Attends church more than monthly	10
Adolescent BMI	Measured in clinic at age 11	11

7.4 The data I-DAG

As per rapid ESC-DAGs, the ALSPAC variables described above were used to replace concepts in the conceptual I-DAG. Specifically;

1. Variables that were measured cross-sectionally to one another were collected into discrete groups.
2. These groups were then reorganised in the I-DAG's virtual space such that groups that were measured earlier were always to the left of groups that were measured after.
3. This graph was then 'forward saturated' such that directed edges were drawn from all variables in earlier groups to all variables that occurred after.
4. Each posited edge, but not the reverse, was then assessed (in practice, for parental, intra-personal, interpersonal, and institutional variables pertaining to an adolescent, it is very unlikely that any of these directed edges will be deleted and technically impossible that they will be reversed).
5. Each group of cross-sectional variables was then saturated such that all nodes within the group were connected by a directed edge.
6. Finally, each of these intra-group directed edges was assessed.

For example, neighbourhood safety, school engagement, adolescent anti-social behaviour, alcohol attitudes, adolescent smoking, and adolescent substance use were all measured at age 14. Omitting the baseline confounders for now, all nodes at age 12, 12.5, 13.5 had directed edges drawn to each node at age 14. Each directed edge was then assessed. Only the first two causal criteria were used. The outcome of each assessment was to retain the directed edge as posited. Next, each node at age 14 had a directed edge drawn from it to the variables measured at age 15.5, 16, and 16.5. These were assessed in the same way, and similarly none were deleted or reversed. Finally, the group of variables at age 14 was saturated and each directed edge was assessed using the full Translation process. None were deleted, meaning that each variable was still either influenced by or influenced all others. The resulting causal order was; neighbourhood safety → school engagement → adolescent anti-social behaviour → alcohol attitudes → adolescent smoking → adolescent substance use, in which each preceding node had a direct effect on all subsequent nodes. This ordering is discussed below. Once this process is completed for each timepoint in the data, the data I-DAG is complete.

The data I-DAG was too complex to replicate usefully on paper. Figure 7-4 below is a highly simplified version in which all baseline confounders were reduced to a single node. Henceforth any reference to 'the data I-DAG' is specific to this graph. Clearly, however, even this simplified model is too complex to be illustrative in a visual sense. As such the following paragraphs deconstruct and interpret it before providing a more intelligible alternative.

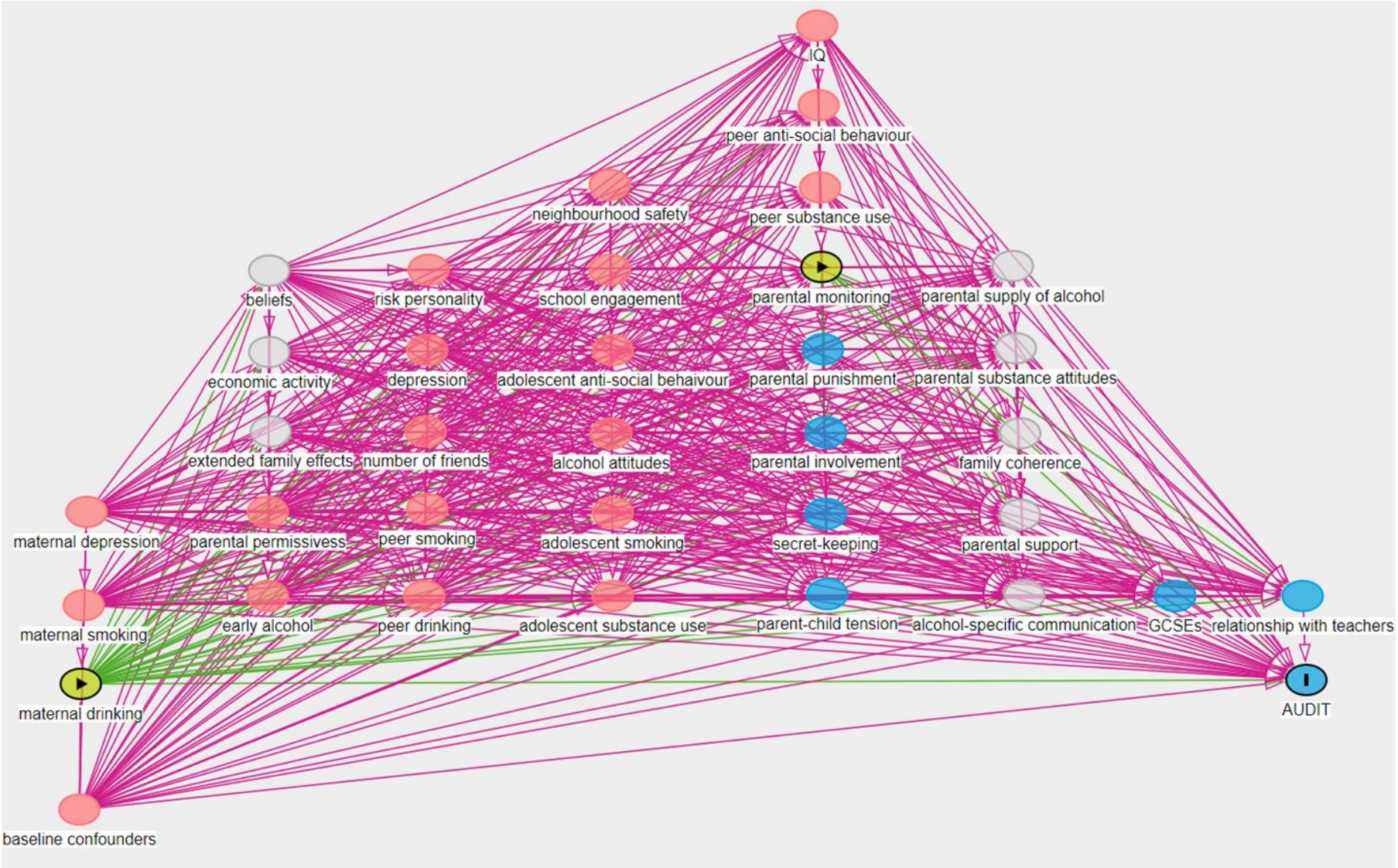


Figure 7-4: Simplified data I-DAG

Figure 7-5 below uses a DAG to reduce the data I-DAG down to ALSPAC waves from which variables were taken. In other words, each node in this DAG represents the ALSPAC timepoints from which variables were taken to replace concepts in the conceptual I-DAG.

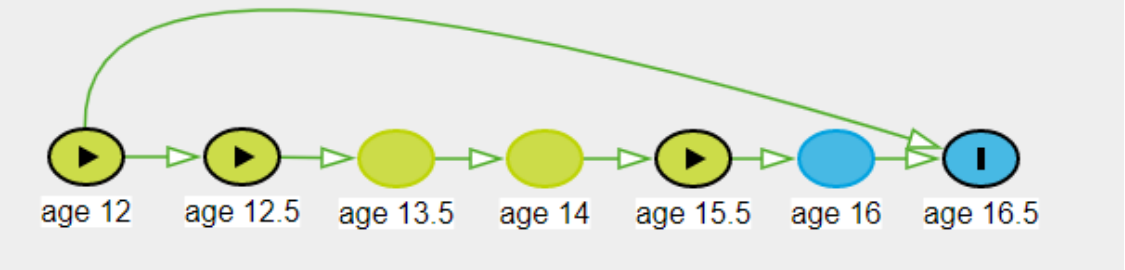


Figure 7-5: Causation between ALSPAC surveys

This DAG demonstrates the underpinning causal model for the data I-DAG. The variables at age 12 influence the variables at age 12.5, which influence the variables at age 13.5, and so on up to the outcome at age 16.5. The baseline confounders are omitted for illustration, as were the direct effects between all nodes and any non-adjacent nodes, except for the direct effect of the age 12 variables on age 16.5. The yellow-green nodes with triangles at ages 12, 12.5 and 15.5 represent ALSPAC waves for which at least one parental exposure was measured. The outcome is represented by the blue node with the vertical bar. This DAG succinctly describes how the timing of measurement in ALSPAC was used to determine the direction of edges for time-dependent variables in the data I-DAG. However, timing of measurement cannot assist in determining the direction of causation for cross-sectional variables. Figure 7-6 below is an excerpt of the data I-DAG, focusing on the variables measured at age 13.5. In this graph causation flows from top to bottom, as well as from left to right. This is indicated by: 1) the directed edge entering the diagram from the left at the top; 2) another leaving the diagram at the bottom to the right; 3) the directed edges connecting each node to its immediate ancestor above it and its immediate descendent below it; and 4), the curved directed edges from each node to all less immediate descendants.

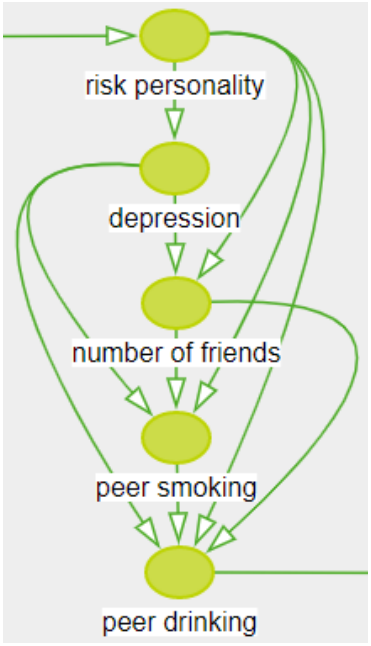


Figure 7-6: DAG focusing on variables at age 13.5

The ESC-DAGs Translation processes determined the causal order of risk personality → depression → number of friends → peer smoking → peer drinking. Of course, for this graph, each directed edge represents a relationship that cannot technically be reduced to a uni-directional arrow. In a conceptual sense, it is implausible to assume that depression does not influence personality, for example. Further, in a statistical sense, reverse causation between any two cross-sectional variables cannot be ruled out. Indeed, this is the case for each cross-sectional variable from the data I-DAG in Figure 7-4. For example, none of the parenting variables at age 15.5 can be assumed to have a uni-directional effect on any of the others (i.e. parental monitoring, parental punishment, parental involvement, secret-keeping, and parent-child tension). However, as noted throughout and demonstrated in the following two Results chapters, the benefit of using a DAG to direct analysis does not lie in determining the direction of effect between cross-sectional variables, but in taking a structural and transparent approach to sensitivity analysis of that relationship.

Figure 7-7 below is a version of the data I-DAG which focuses on the longest possible causal path from the exposure of maternal drinking to the outcome through each of 36 possible intermediates, measured or unmeasured. It thus keeps the same nodes as the data I-DAG, but uses a minimal amount of directed edges.

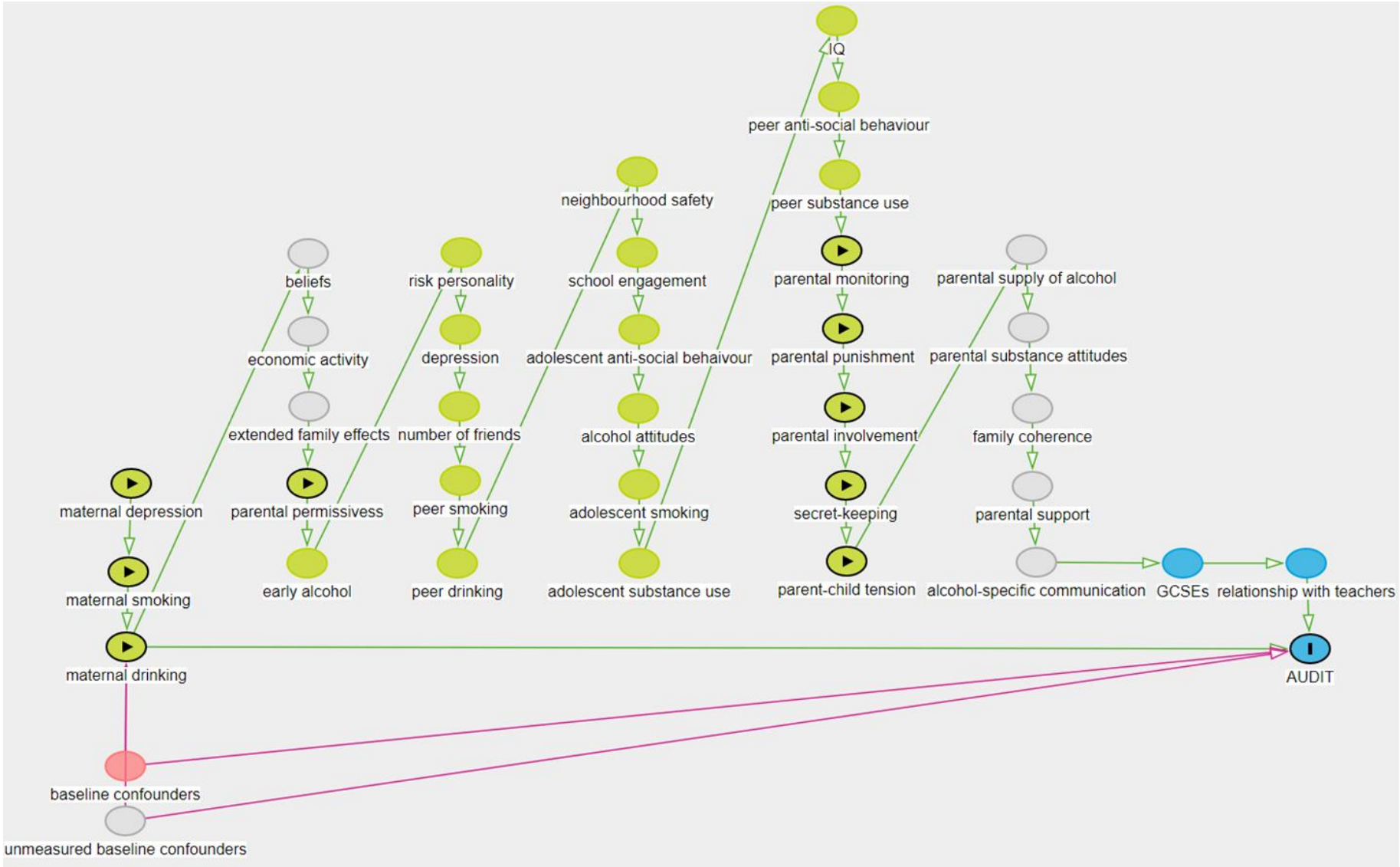


Figure 7-7: Further simplification of data I-DAG

Despite being very simplified, this graph is a useful heuristic as it explicates some key characteristics of the full data I-DAG in a more transparent way. Firstly, similar to the data I-DAG and Figure 7-6, causation flows from left to right, and then from top to bottom. Secondly, each column in this graph represents a timepoint in ALSPAC, similar to the data I-DAG and Figure 7-5. Indeed, despite the simplification, this graph is sufficient to correctly inform all analytical models in the subsequent chapters given the following informal assumptions:

1. If node V1 is above or to the left of node V2, then V1 influences V2
2. Directed edges between variables in the same 'column' as each other can be reversed for sensitivity analysis
3. All baseline confounders influence all other variables

Assumptions 3 is worth brief discussion. Figure 7-8 below represents the baseline confounding structure in reference to the exposure of maternal drinking, a generic EIMOC node, the parenting mediators at age 15.5, and the outcome - each receiving a directed edge from all baseline confounders. Causation flows in the same way as the data I-DAG.

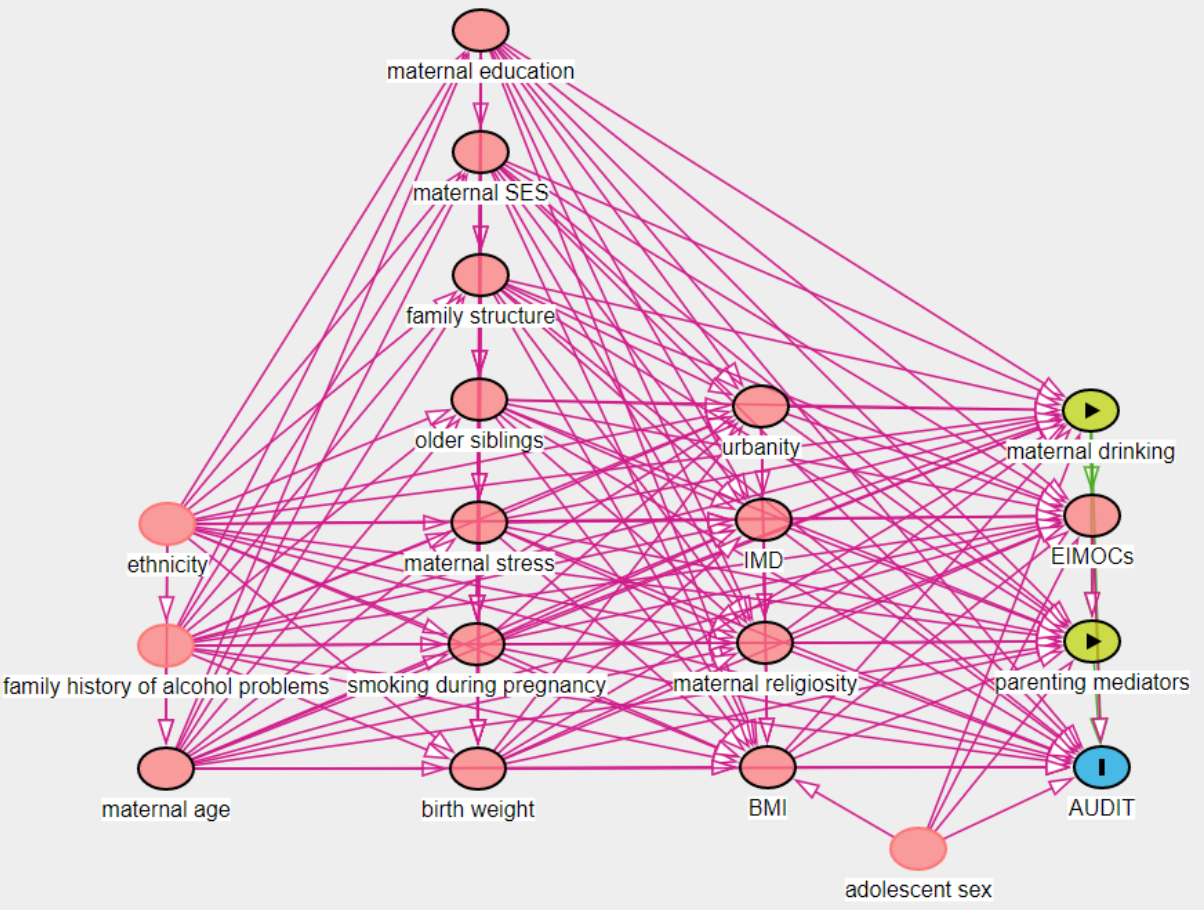


Figure 7-8: Simplified data I-DAG focusing on confounders

This DAG thus encodes assumption 3, that all variables measured from age 12 in the data I-DAG (i.e. the parental influences, intra-personal variables, interpersonal variables, and institutional variables) are influenced by all baseline confounders. It also encodes a further assumption that the biasing influence of each of these confounders is not fully captured by the indirect effects from each baseline confounder to each post-exposure variable. In other words, there is still a direct effect from each baseline confounder to each subsequent covariate. For example, take maternal age at offspring's birth, the variable in the bottom-left of Figure 7-8. This DAG states that at least some aspect of maternal age's influence on all post-exposure variables is not captured by any of the baseline confounders that it influences (i.e. maternal education through to BMI). Similarly, the effect of birthweight is not captured entirely by urbanity, IMD, maternal religiosity, or BMI. Because of this, when estimating the effect of maternal drinking, the EIMOCs, or the parenting mediators, this DAG suggests that all confounders must be adjusted for. This is the conservative approach to building DAGs discussed in Chapter 4. It can be problematic for some statistical methods, and the topic is revisited in Chapter 10.

Using the data I-DAG to directly inform statistical models is made much more practicable by using the DAGitty software to select the effect of interest. For example, to estimate the average causal effect (ACE) of maternal drinking on AUDIT (or indeed the total causal effect (TCE)), the maternal drinking node can be set as the sole exposure. The data I-DAG (i.e. the simplified data I-DAG in Figure 7-4) then suggests that the IPW for estimating the ACE should feature all baseline confounders and all variables measured at age 12 that are above maternal drinking (maternal depression and maternal smoking). These are all the variables that influence both maternal drinking and AUDIT (i.e. that confound the exposure-outcome relationship) and which can be adjusted for without introducing collider bias or overcontrolling for a mediator. The below discussion covers this in more detail and Chapter 8 demonstrates how the data I-DAG can be used to inform specific statistical models.

7.5 Discussion

This chapter justified choosing the ALSPAC data, outlined the process of replacing concepts from the conceptual I-DAG with ALSPAC variables, provided corresponding descriptives, and presented and interpreted the data I-DAG. The following paragraphs reflect on the post hoc suitability of the ASLPAC data and on the process of measuring a conceptual I-DAG to produce the data I-DAG. This chapter closes by considering the data I-DAG in contrast to the research objectives.

7.5.1 What is measured: conceptual coverage in ALSPAC

The vast majority of the concepts in the conceptual I-DAG were matched with a variable. The conceptual I-DAG had 52 concepts in total. Of these, 42 (80.1%) were measured in the ALSPAC data, and 10 were either not measured at all or were not measured in the appropriate time-frame. However, six of the concepts that were measured were done so in a problematic fashion. Two variables were problematic in terms of timing of their measurement (GCSEs and relationships with teachers), one had very low variation (alcohol specific rules measured as parental permissiveness towards adolescent alcohol), and three were potentially problematic in terms of how well they matched the concept (neighbourhood attachment was measured using adolescent perception of neighbourhood safety; parent-child communication was measured via 'secret-keeping'; and birthweight was used to measure child health/illness/disability). Each was retained as the best available option.

Several parental exposures were not measured (alcohol-specific communication; parental supply; parental support; parental substance attitudes; and family coherence/conflict). This is at least partially attributable to the focus on parental influences from the ESC-DAGs review and was assumed to be largely unproblematic due to the conceptual overlap between the parental influences as a whole. Three unmeasured concepts were non-parental intermediates (adolescent beliefs; economic activity, and extended family effects) and three were baseline confounders (neighbourhood availability of alcohol, neighbourhood space, and experience of abuse). While adolescent beliefs were not measured, maternal religiosity was used as a proxy. Despite being one of the additional concepts, child abuse and neglect is an unfortunate omission, as it has been cited as an important determinant (Maniglio, 2009). However, the most notable omission was supply of alcohol to the adolescent, which was not measured either in an interpersonal sense (i.e. supply from parents or from someone else such as peers, extended family, siblings, etc.) or an environmental sense. As the review of systematic reviews found parental supply of alcohol to be one of the parental influences with stronger evidence, the omission of supply in general from the data I-DAG is a notable limitation. Nonetheless, in terms of what is measured and what is not, the data I-DAG was deemed sufficient to enact the data analysis plan.

7.5.2 When is it measured: timing in ALSPAC

As described in Chapter 4, the conceptual I-DAG was not saturated for the pragmatic reason that many of the directed edges posited in such a model would not be possible depending on the data. This is what differentiates ESC-DAGs from rapid ESC-DAGs. It is also a point that comments on the

nature of secondary data. Consider the ideal data for building the data I-DAG: it would have repeated measures at multiple intervals for each time-varying concept identified as important by the researcher(s). Of course, data like this is unrealistic for numerous reasons. Instead, even in birth cohorts such as ALSPAC which are relatively comprehensive in terms of the breadth and depth of their measures, it is the case that most concepts are measured infrequently. For example, ALSPAC surveys tends to focus on specific aspects of the adolescent's life (e.g. social life, wellbeing, etc.). A consequence of this is that the data I-DAG comes to mirror the timing of the surveys in ALSPAC as much as it does the conceptual I-DAG.

In ALSPAC the variables selected to measure the parental influences and the non-parental intermediates (i.e. those variables for which causal effects are estimated in the subsequent analyses) are measured over seven different timepoints (adolescent ages 12; 12.5; 13.5; 14; 14.5; 15.5; and 16). With few exceptions, all of the concepts were measured only once during the interval between age 12 and 16.5. Accordingly, the lack of repeated measures at multiple intervals means that the data shapes the question that the DAG can ask. The main implication for subsequent analysis in this thesis is that there are some instances where concepts considered as EIMOCs cannot be employed as such. For example, it is reasonable to expect that parenting and the parent child relationship are influenced by the school effects of adolescent relationships with teachers and academic performance (and vice versa). However, as these school effects are not measured until after parenting and the parent-child relationship, any confounding effect of the underlying school concepts cannot be controlled for without inducing other forms of bias (collider bias and overcontrol for downstream effects). Note however that this does not preclude sensitivity analysis between these variables.

In short then, the timing of the surveys in ALSPAC determines the causal models in the data I-DAG because causation between surveys should technically obey temporality (at least for time-varying variables). Accordingly, a data I-DAG is best understood as a compromise between the conceptual I-DAG and the availability of measures in the selected data source(s).

7.5.3 Suggested analytical models

The pivotal causal model posited in the data I-DAG is that, controlling for the baseline confounders, maternal drinking at adolescent age 12 has a causal effect on adolescent drinking at age 16.5. The data I-DAG further posits that much of this effect is through numerous intermediates between age 12 and 16.5, mainly: parenting; the parent child relationship; adolescent behaviours; adolescent mental health; peer effects; and institutional effects (school, neighbourhood). If estimating the direct effect of maternal drinking, none of the intermediates should be controlled for, including the parenting and

parent-child relationship variables. Similarly, indirect effects through any of the intermediates, including the parenting mediators, can be estimated by controlling the EIMOCs suggested by the data I-DAG.

As such the data I-DAG is compatible with SRO 4 and SRO 5, and the analysis plan. SRO 4 was to test average effects of the parental influences and their intermediates as determined by the data I-DAG. Chapter 3 determined that using inverse probability weighting (IPW) to estimate average causal effects (ACEs) is a suitable method for doing so. The role of the data I-DAG in this regard is simply to identify the variables that are to be controlled for when estimating the ACEs (i.e. the variables factored in the propensity scores for the IPWs). In practice this means controlling for every node with an input into the variable being assessed. For example, using DAGitty to select the exposure, the data I-DAG suggests that the effect of parenting and the parent-child relationship (measured at age 15.5) can be estimated by controlling for the baseline confounders, maternal drinking, parental permissiveness towards adolescent alcohol use, and various non-parental intermediates. In contrast, the data I-DAG suggests that only the baseline confounders (and potentially maternal smoking and maternal mental health) need to be factored into the IPW for maternal drinking.

SRO 5 was to use the information from the average effects analysis to investigate mediators of parental influences. Now that the data I-DAG has identified maternal drinking as the exposure of interest, SRO 5 can be refined. Thus SRO 5.1 becomes 'investigate indirect effects of maternal drinking via parenting and the parent-child relationship'; and SRO 5.2 becomes 'investigate indirect effects of maternal drinking mediated via non-parental influences'. Note however that both SRO 5.1 and SRO 5.2 are dependent on the analysis of average effects. Chapter 3 selected mediational g-computation as a suitable method for investigating these effects, mainly due to its ability to account for EIMOCs. Essentially, when estimating indirect effects, the data I-DAG is used to identify the EIMOCs that need to be modelled for the mediator in question. For example, when estimating the indirect effect of maternal drinking through the parental monitoring measured at age 15.5, the data I-DAG suggests identifying every intermediate measured between age 12 and 15.5 as a potential EIMOC (the simplified data I-DAG in Figure 7-4 is set to this specific question). As such the mediation analyses corresponding to SRO 5.1 may be assumed to deal with the issue of exposure-induced mediator-outcome confounding in a comprehensive fashion. However, due to the timing of the data, some of the intermediates measured at earlier ages cannot be said to account for EIMOCs as comprehensively, including the effect of parental permissiveness towards adolescent alcohol use measured at age 12.5. This is a limitation discussed further in Chapter 10.

As noted in Chapter 2 and Chapter 3, when researchers are less certain about the presence or directionality of edges in a DAG, they should be free to transparently and systematically manipulate these edges in an attempt to gain some indication of their effect. While the restrictions imposed by temporal ordering between the surveys in ALSPAC arguably preclude this type of sensitivity analysis between timepoints, the cross-sectional groupings are entirely compatible with this approach. This is noted when relevant over the subsequent analyses.

7.6 Conclusion

This and the previous two chapters have met MRO 3 – apply and demonstrate ESC-DAGs. The process of developing the data I-DAG has been detailed from the earliest steps. The review of systematic reviews was used to determine the important parental influences on adolescent alcohol harm. The ESC-DAGs review then systematically extracted and inter-related these concepts through the Translation and Synthesis stages, thus producing the conceptual I-DAG. Finally, this chapter has demonstrated how the concepts in the conceptual I-DAG are replaced by variables from a data source, and reflected on some of the tensions between DAG-based approaches and real-world data. The above discussion also concluded that the data I-DAG is suitable to meet the analytical research objectives.

The data I-DAG is a complex diagram that integrates multiple components of the socio-ecological model of health. For example, the total number of causal paths in the data I-DAG, even with reducing all confounders to one node, is 382,507,508. As such, it is categorically more comprehensive than DAGs that are generally used in the literature. Additionally, by its nature the data I-DAG can readily direct causal analysis. Thus, it is the position of this chapter that the data I-DAG, as the final product of the ESC-DAGs process for this thesis, further supports the argument that ESC-DAGs meets MRO 2 – to develop a method for building DAGs (an argument made in Chapter 4 and supported by peer review).

Chapter 8 Results 1: Causes of adolescent alcohol harm

This chapter is the first of two to perform inferential analysis of the ALSPAC data using the POF statistical methods described in Chapter 3 and the data I-DAG produced in Chapter 7. It mainly pertains to MRO 4 (demonstrate how the method can be used to direct conventional data analysis), MRO 5 (use the method to direct data analysis under the potential outcomes framework), and SRO 4 (test the causal effects of parental influences and their intermediates on adolescent alcohol harm). In terms of the analysis plan this chapter thus pertains to the first aspect - producing causal effect estimates for all parental influences and their intermediates which can subsequently inform mediation models (Chapter 9). Box 8-1 is included below as a brief recap on the thesis thus far.

Box 8-1: Recap on thesis so far

A foundation in causal inference was established in Chapters 2 and 3, focusing on counterfactuals, DAGs, and POF statistical methods, especially the IPWs for estimating ACEs, and mediational g-computation for estimating natural in/direct effects. Chapter 3 also set out a data analysis plan, to estimate causal effects of parental influences on adolescent alcohol harm (and their intermediates) and to use these effects to inform mediation models. Chapter 4 then introduced the novel ESC-DAGs method. A review of systematic reviews was conducted in Chapter 5 that identified 12 parental influences of adolescent alcohol harm that appeared to be important across the literature. The ESC-DAGs protocol was then applied in Chapter 6 to a subset of 50 studies and identified a further two parental influences and differentiated between confounders and intermediates. This produced the conceptual I-DAG. The ALSPAC data was introduced in Chapter 7, key measures were explored in some detail, and the ESC-DAGs protocols were further applied to convert the conceptual I-DAG into a data I-DAG. The data I-DAG identified nine parental influences and 17 non-parental intermediates.

The first section of this chapter is concerned with MRO 4. It takes a subset of the parental influences and subjects them to DAG-based sensitivity analysis. Thus, this section demonstrates the utility of DAGs outside POF statistical methods, generates estimates which can be usefully contrasted with later estimates from POF models, and comments on some of the key strengths and limitations of using DAGs to determine adjustment strategies. The second section is different. It has both the methodological aim of demonstrating how I-DAGs can inform statistical analysis using POF methods and the substantive aim of generating causal estimates for parental influences on adolescent AUDIT score (thus informing the subsequent mediation models). It uses the I-DAGs and the IPW methods described in Chapter 3 to estimate ACEs for each of the parental influences on adolescent AUDIT scores, as well as their intermediates.

8.1 Contrasting adjustment strategies using multiple regression

Figure 8-1 is included below as a simple reference point going forwards. It is an excerpt of the data I-DAG that focuses exclusively on the parental influences and the outcome, and only shows directed edges between the waves of the data. This demonstration uses the parental substance behaviours that were adequately measured in ALSPAC. As per Figure 8-1 and as discussed in Chapter 7, these were maternal drinking and maternal smoking, both measured at adolescent age 12. The following investigates both these variables in the complete case data (CC), and in the multiply imputed data (MI). However, father's drinking is also used, although this was limited to CC analysis. Despite not being included in the data I-DAG due to comparatively low participation rates, its inclusion was for two reasons; firstly, the review of systematic reviews suggested that father's drinking was an important determinant of adolescent alcohol harm and thus is worth exploring. Secondly, the review of systematic reviews further indicated that father's drinking and maternal drinking are often studied in tandem. Thus, it was important to determine the effect of including father's drinking in models which estimate the effect of maternal drinking on adolescent AUDIT score (e.g., it would be problematic if father's drinking explained the majority of the effect of maternal drinking). Father's smoking was not used because this would further restrict the number of cases available for CC analysis, and because its relationship to maternal drinking is not as important as father's drinking. Illicit substance use was not used at all due to prohibitively low variation, as per Chapter 7.

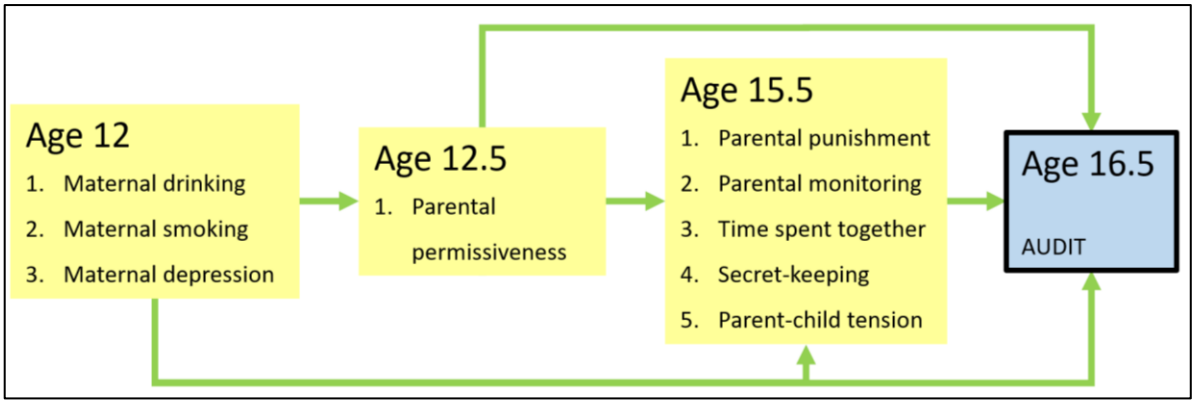


Figure 8-1: Except of data I-DAG focusing on parental influences of adolescent AUDIT score

This particular subgroup of parental substance behaviours was selected for two main reasons. First, the I-DAGs do not posit as many confounders for these variables compared to the parental influences measured at age 15.5, and thus multiple regression models can avoid problems with overfitting. Only nine confounders were used for these models (birthweight, maternal smoking during pregnancy, maternal SES, maternal education, family structure, family history of alcohol problems, maternal stress, older siblings, and BMI). The rest (sex, urbanity, maternal age, ethnicity, maternal religiosity, and neighbourhood deprivation) were excluded because they did not have associations with the outcome, and thus cannot meet the technical definition of confounders set out in Chapter 2. Secondly, the subsequent mediation models all investigate indirect effects of maternal drinking on adolescent alcohol harm. Thus, given its crucial role in subsequent analyses, a more detailed analysis of how this variable affects adolescent AUDIT score was deemed as necessary.

8.1.1 I-DAG-based regression modelling

There are four 'sets' of results differentiated by adjustment strategies. The analysis strategy is summarised in Table 8-1 below and the results are presented in Table 8-2. Set 1 and Set 2 both use I-DAG-based adjustment strategies, except the first set does not include the other parental substance behaviours in its adjustment strategy while the second does. For example, when estimating the effect of maternal drinking, maternal smoking and father's drinking are not included in the Set 1 model but are included in the Set 2 model. A contrast between this first pair of sets comments on the limitations of DAGs when dealing with reverse causation in settings with suspected bi-directional relationships in cross-sectional data. The resulting coefficients also act as comparators for Set 3 and Set 4 which use adjustment strategies designed to emulate common modelling errors. Specifically, Set 3 includes mediators in its adjustment strategies, and Set 4 only adjusts for key sociodemographic variables. In other words, Set 3 looks at a case of 'overcontrol' and Set 4 at a case of 'undercontrol'. Except for father's drinking which was not imputed due to sample size issues as discussed in Chapter 7, the

effect of each parental substance use behaviour on adolescent AUDIT score was estimated in the CC data and then replicated using the MI data (i.e. across imputed 50 datasets). There are 20 models altogether.

Table 8-1: Summary of OLS models

Data	Adjustment Strategy	Model	Maternal drinking/smoking	Father's drinking
Complete Case	Set 1 - DAG, excl other substances	A	✓	✓
	Set 2 - DAG, incl other substances	B	✓	✓
	Set 3 - Overcontrol for mediators	C	✓	✓
	Set 4 - Undercontrol for confounders	D	✓	✓
Imputed	Set 1 - DAG, excl other substances	E	✓	
	Set 2 - DAG, incl other substances	F	✓	
	Set 3 - Overcontrol for mediators	G	✓	
	Set 4 - Undercontrol for confounders	H	✓	

These analyses are designed to meet MRO 4 (demonstrate how the method can be used to direct conventional data analysis). However, part of this research objective is concerned with *demonstrating* the task of using the I-DAGs inform the specific models. When estimating the effect of maternal drinking, variables that the I-DAGs suggested to confound the relationship between maternal drinking and the outcome were simply added to the regression model as covariates. In a more practical sense, when performing this task with the DAGitty software, once the 'exposure' was set to maternal drinking, the g-separation algorithm was performed, and the confounders are highlighted (in red/pink). Unsurprisingly, given the conceptual similarity between these substance behaviours and how they were measured cross-sectionally, the I-DAGs suggested the same sets of variables as confounders for each, as above.

The interactive DAGitty versions of the I-DAGs can then be altered to reflect the different adjustment strategies demonstrated here. For Set 1 and Set 2 the directed edges between the different parental influences can be actively changed to match the models. Figure 8-2 below demonstrates this using a simplified DAG with maternal drinking as the exposure and maternal smoking as the covariate (father's drinking was omitted for demonstration). Figure 8-2(a) corresponds to Set 1; it recommends not including maternal smoking in the model for maternal drinking, as maternal smoking is a mediator in this case. The situation is different in Figure 8-2 (b), in that maternal drinking is now downstream from maternal smoking. In other words, the directed edge between maternal drinking and maternal

smoking has been reversed, and so this DAG recommends that maternal smoking should be entered as a covariate alongside the confounders when estimating the total effect of maternal drinking on AUDIT score. Thus Figure 1-1(b) corresponds to Set 2.

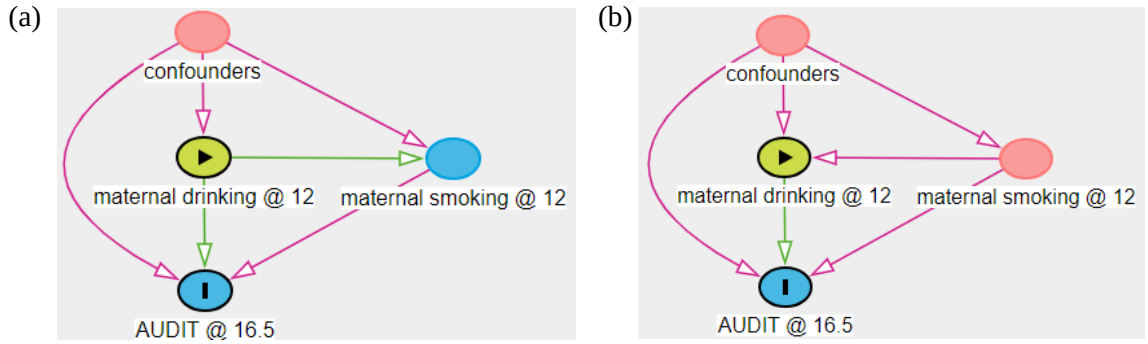


Figure 8-2: DAGs of maternal drinking for Set 1 and Set 2

There were 25 variables from the I-DAG that might have been valid mediators (i.e. have an indirect effect as opposed to not having an indirect effect). Thus, in an attempt to ensure that the Set 3 analysis was a valid demonstration of overcontrol for mediators, the intermediates that were used here were those that were tested as mediators of the effect of maternal drinking on adolescent AUDIT score in Chapter 9 and were found to have indirect effects. In other words, the Set 3 regression models were performed at the chronological end of this study's analysis. The mediators included were adolescent attitudes to alcohol, adolescent smoking, peer alcohol use, and peer smoking. The DAG in Figure 8-3 below describes the regression model used for Set 3. The d-separation algorithm suggests modelling the confounders as covariates and leaving the mediators out of the model altogether. If the intermediates are controlled for in the DAG (recall that any node can be 'set' to 'controlled' in the DAGitty software), then the total effect of maternal drinking would not be identified. As such, the regression models for Set 3 include each of these intermediates as covariates in order to demonstrate the consequences of overcontrol for a mediator.

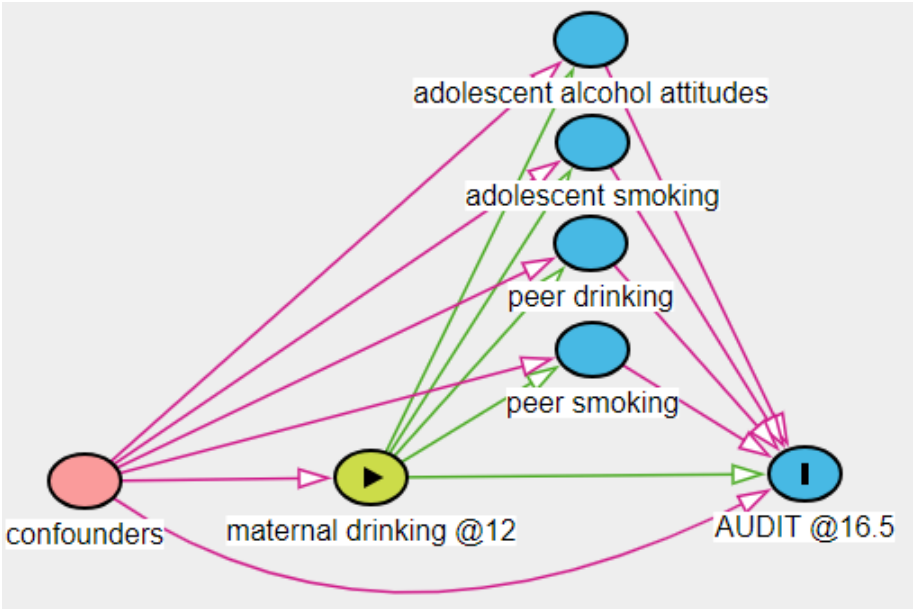


Figure 8-3: DAG of maternal drinking for Set 3, showing overcontrol for mediators

The sociodemographic variables used in the models that undercontrolled for confounders were class, ethnicity, sexuality, and sex - omitting all others. Figure 8-4 below demonstrates the corresponding DAG, again working with the example of maternal drinking. It shows how, even assuming that the rest of the confounders are all downstream of the key sociodemographic confounders, novel confounding bias caused by the other confounders would not be accounted for just by controlling for key sociodemographics (a 'controlled' node is represented by a white fill).

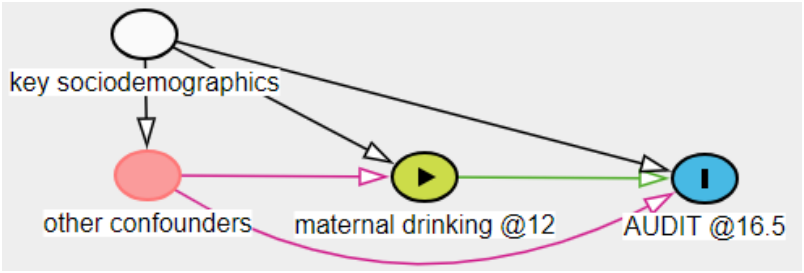


Figure 8-4: DAG of maternal drinking for Set 4, showing undercontrol for confounders

8.1.2 Findings and interpretation

Linear regression coefficients and 95% confidence intervals for all 20 models are presented in Table 8-2 below. To aid interpretation, each row in the table has been labelled A through H, as per Table 8-1. Thus, maternal drinking and maternal smoking both have eight models A through H, and father's drinking has four models A through D. In other words, models A through D correspond to Set 1 – 4 in

the CC data for each parental substance behaviour, and models E through H correspond to Set 1-4 in the MI data for maternal drinking and maternal smoking.

8.1.2.1 Cross-sectional sensitivity analysis

For maternal drinking in the CC data without controlling for maternal smoking or father's drinking (i.e. model A), adolescents whose mothers drank more than 14 units had AUDIT scores that were 1.01 (0.58, 1.44) higher on average (higher scores represent greater risk of harm). This decreased to 0.93 (0.31, 1.55) when maternal smoking and father's drinking were introduced in model B. The effect was similar in the MI data - when maternal drinking was modelled without maternal smoking in model E the effect was 1.06 (0.72, 1.39) which decreased to 1.02 (0.69, 1.36) in model F when maternal smoking was included. Because maternal drinking is the exposure of interest for the mediation models, this range of coefficients from 0.93 to 1.06 is a useful benchmark for the rest of the analyses in this chapter and Chapter 9. The decrease in effect size when introducing maternal smoking and father's drinking could indicate that they mediate part of the effect of maternal drinking on AUDIT score. However, while controlling for a variable on the causal pathway between the exposure and the outcome can reduce the coefficient between models, it could equally indicate the opposite - that maternal drinking lies on the pathway from maternal smoking and father's drinking, and that including them simply removes some of the previously uncontrolled confounding. In other words, it is impossible to tell from simply including the other substance behaviours whether they cause or are caused by maternal drinking. There is thus no way to know if Figure 8-2(a) or Figure 8-2(b) is correct. This is a key limitation of DAGs that is sometimes misunderstood. DAGs can only represent hypotheses and data, they cannot be used to objectively uncover causal relationships.

If any of these variables had been measured before or after the others, then it could be assumed that the effect of later variables was not being conflated with reverse causation. Further, if these variables were repeated measures, approaches which take account of feedback mechanisms in the form of time-varying confounding could be used to verify the bi-directional relationship given repeated measures data (Daniel et al., 2011). However, as the parental substance behaviours are measured cross-sectionally in ALSPAC, there is no practical way to do so with this data. Instead, the only available option is to test the robustness of the estimates to different assumptions regarding causal order with sensitivity analysis, which is what has been done here. Accepting various caveats including sample size, attrition, and modelling method, these models show that maternal smoking and father's drinking fail to notably attenuate the effect of maternal drinking on AUDIT score, even using the rigorous ESC-DAGs approach to selecting the wider adjustment set. It is also notable that these coefficients are a slight reduction on the unadjusted association of 1.18 in the CC data from Chapter 7.

Table 8-2: Changes in AUDIT score for parental substance behaviours using different adjustment strategies

Data	Adjustment Strategy	Model	Maternal drinking		Maternal smoking		Father's drinking	
			N	β (95% C.I.)	N	β (95% C.I.)	N	β (95% C.I.)
Complete Case	Set 1 - DAG, excl other substances	A	2,287	1.01 (0.58, 1.44)	2,334	0.76 (0.12, 1.14)	1,339	1.60 (1.00, 2.15)
	Set 2- DAG, incl other substances	B	1,238	0.93 (0.31, 1.55)	1,238	1.00 (0.06, 1.96)	1,238	1.11 (0.49, 1.74)
	Set 3 - Overcontrol for mediators	C	1,574	0.53 (0.03, 1.02)	1,613	0.38 (-0.36, 1.13)	975	1.54 (0.92, 2.15)
	Set 4 - Undercontrol for confounders	D	2,420	1.18 (0.76, 1.59)	2,481	0.87 (0.27, 1.46)	1,442	1.75 (1.21, 2.29)
Imputed	Set 1 - DAG, excl other substances	E	7,959	1.06 (0.72, 1.39)	7,959	0.72 (0.22, 1.22)	-	-
	Set 2- DAG, incl other substances	F	7,959	1.02 (0.69, 1.36)	7,959	0.63 (0.13, 1.13)	-	-
	Set 3 - Overcontrol for mediators	G	7,959	0.71 (0.38, 1.03)	7,959	0.18 (-0.32, 0.68)	-	-
	Set 4 - Undercontrol for confounders	H	7,959	1.09 (0.75, 1.43)	7,959	0.92 (0.45, 1.38)	-	-

For maternal smoking in Table 8-2, modelling without maternal drinking or father's drinking in the CC data (model A) produced an estimate of 0.76 (0.12, 1.41). Surprisingly, when maternal drinking and father's drinking were added (model B) the effect increased to 1.00 (0.06, 1.96). The range of the confidence interval also increased markedly from 1.29 to 1.90, although this may be explained by the reduction in sample size from N=2,334 in model A to N=1,238 in model B. This was due to missing data on father's drinking. The models in the MI data, however, were more in line with expectations as the effect of maternal smoking decreased from 0.72 (0.22, 1.22) in model E to 0.63 (0.13, 1.13) when modelled with maternal drinking in model F. There could be several explanations for the differences between model B in the CC data and model F in the MI data. First, the difference in sample size is relatively large at 6,721. Second, the analysis in the MI data did not factor father's drinking. Third, assuming that the imputations were successful in tackling selection bias via sample attrition (as argued in Chapter 7 and Appendices A-C), the CC sample of N=1,238 is likely to be subject to more selection bias than the MI data. Thus, the higher effect size of maternal smoking in this model may be due to selection bias, rather than the addition of the maternal drinking and father's drinking variables. Acknowledging the limitations of CC analysis, there is an indication that maternal drinking and father's drinking either confound or mediate the effect of maternal smoking on adolescent AUDIT score.

Finally, the effect of father's drinking, while larger than maternal drinking, decreased from 1.60 (1.00, 2.15) in model A to 1.11 (0.49, 1.74) in model B. This is a substantial reduction of 31% that was unlikely to be due to changes in the sample: sample size was fairly stable for father's drinking between model A (N=1,339) and model B (N=1,238). Thus, maternal substance behaviours appear to either confound or mediate the effect of father's drinking on adolescent AUDIT score. Again, however, note that the sample size in model A was notably small for father's drinking compared to maternal drinking or maternal smoking at N=1,339 (compared to N=2,287 and N=2,334 respectively).

Before continuing to look at overcontrol for mediators and undercontrol for confounders, it is helpful to consider the meaning of the coefficients in terms of adolescent alcohol harm. For example, what does maternal drinking's increase of +1.06 on the AUDIT scale mean for adolescent alcohol harm at age 16.5 in the ALSPAC cohort? Compared to the observed measures of central tendency (median of 6, mean of 6.59), an increase of +1.06 is substantial, and indicates a higher degree of average consumption and consequence, especially when considered against the AUDIT threshold for hazardous drinking. Acknowledging that uncertainty around the hazardous drinking threshold for adolescents in AUDIT is an important caveat, logistic regression was used to model the probability of adolescents having hazardous or harmful AUDIT scores. This was done for the same parental substance behaviours using the same DAGs from Set 1 and Set 2 in the CC data (i.e. model A and B with adjusted logistic regression and a binary version of the AUDIT score). Results support the above

coefficients in finding that adolescents exposed to maternal drinking of over 14 units per week, father's drinking of over 21 units per week, and maternal smoking are markedly more likely to have hazardous AUDIT scores than their referents. These effects also reduce or have wider confidence intervals when the other parental substance behaviours are added to the models. Thus, while the regression coefficient of +1.06 on the AUDIT score may have unclear implications for adolescent alcohol harm, the logistic regression models indicate that hazardous drinking is more likely in adolescents who are exposed to parental substance behaviours.

Table 8-3: DAG-based odds ratios for hazardous AUDIT scores in complete case data

	Set 1		Set 2	
	N	OR (95% C.I.)	N	OR (95% C.I.)
Maternal drinking	2,287	1.48 (1.23, 1.78)	1,238	1.48 (1.12, 1.96)
Maternal smoking	2,334	1.69 (1.29, 2.22)	1,238	1.71 (1.29, 2.26)
Father's drinking	1,339	1.91 (1.49, 2.45)	1,238	1.8 (1.17, 2.77)

8.1.2.2 Overcontrol for mediators

Regression models can be applied to identify the total effect on an exposure on an outcome, in this case of each parental substance behaviour on AUDIT score. As discussed in Chapter 2 and Chapter 3, a total effect is constituted of a direct effect and an indirect effect. In other words:

$$direct\ effect + indirect\ effect = total\ effect$$

As such, including a mediator in an adjustment set, thus stopping any of its variation from being included in the model, removes the indirect effect through that mediator from the equation for the total effect. The estimate from such a regression model is thus more likely to be analogous to the estimate of the direct effect than the total effect.

As per Table 8-2, controlling for mediators attenuated the effect of each parental substance behaviour. For maternal drinking in the CC data (model C), the effect size was 0.82 (0.31, 1.32), down from 1.01 (0.58, 1.44) in model A. Controlling for mediators of maternal drinking in the MI data (model G) had an effect of 0.93 (0.73, 1.6), compared to 1.06 (0.72, 1.39) in model E. Overcontrolling for mediators for maternal smoking resulted in wider confidence intervals that cross the null in both the CC and MI data. Father's drinking dropped only slightly between model A and model C, from 1.60 (1.00, 2.15) to 1.54 (0.92, 2.15). These attenuated coefficients and widened confidence intervals support the view that controlling for mediators systematically underestimates the total effect of the exposure on the

outcome by removing part of the indirect effect. This is a key strength of DAGs. In using the ESC-DAGs methodology, researchers are explicitly required to assess the direction of effects between any potential confounder and the exposure and the outcome. The crucial result of this is that mediators and confounders will be differentiated by considering how the exposure relates to the covariate. Thus, instances such as that demonstrated here, in which mediators are statistically treated as confounders, can be avoided (and the corresponding bias in estimates).

8.1.2.3 Undercontrol for confounders

As alluded in Chapter 2, a useful way to understand confounding bias is in how the estimate of the exposure's effect on the outcome could in fact represent some of the confounder's effect on the outcome *through* the exposure. In other words, there is effectively a mediation chain from the confounder to the outcome through the exposure, or path C → X → Y in Figure 8-5 below. Thus, in the presence of an additional directed edge C → Y, a confounder can attenuate the effect of an exposure if it has the opposite direction of effect, or amplify if the direction of effect is the same.
In both the CC and MI data, the pattern was similar for each parental substance variable, in that there was a slight increase between the Set 1 models, and the Set 4 models (which only controlled for sociodemographics). The range of these increases was between +22% for maternal smoking in the MI data and +3% for maternal drinking in the MI data.

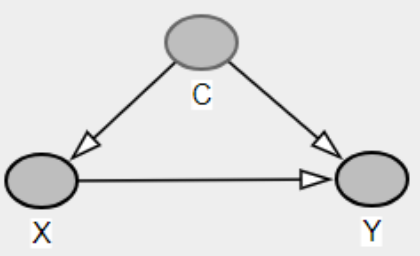


Figure 8-5: Exposure mediates the effect of the confounder on the outcome

Thus, the sociodemographic confounders appear to be largely ineffective at removing confounding bias for parental substances. Indeed, for maternal drinking, the unadjusted coefficient from Chapter 7 was unchanged from 1.18. Rather it is the confounders that were identified from the review of systematic reviews that appear more consequential (e.g. family structure). This example neatly demonstrates that simply controlling for sociodemographics can bias estimates – more thought, of course, is required to identify important confounders. However, it also demonstrates how useful DAGs can be for allocating confounders correctly when modelling.

8.1.2.4 Summary of contrasted adjustment strategies

In terms of reverse causation and differentiating mediators from confounders more generally, this section overall has demonstrated that DAGs are useful tools for directing conventional analysis. The above patterns can be summarised succinctly. Firstly, when other parental substances were controlled for, the parental substance behaviour under investigation saw either a reduced coefficient, a widened 95% confidence interval or both. Secondly, when mediators were controlled for, the coefficient reduced and confidence intervals widened. Finally, when only key sociodemographics were controlled for, effect sizes were slightly higher compared to the I-DAG based models. Forest plots were used to capture these patterns graphically for each parental substance behaviour. They are presented below. Figure 8-6 pertains to maternal drinking, Figure 8-7 to maternal smoking, and Figure 8-8 to father's drinking. As well as demonstrating the above patterns, they further support how this section has found quite strong evidence for the harmful effect of parental substance behaviours on adolescent AUDIT score. For example, even after using ESC-DAGs to account for confounding bias, the only occasion in which confidence intervals crossed the null was when maternal smoking was intentionally misspecified to demonstrate overcontrol for mediators.

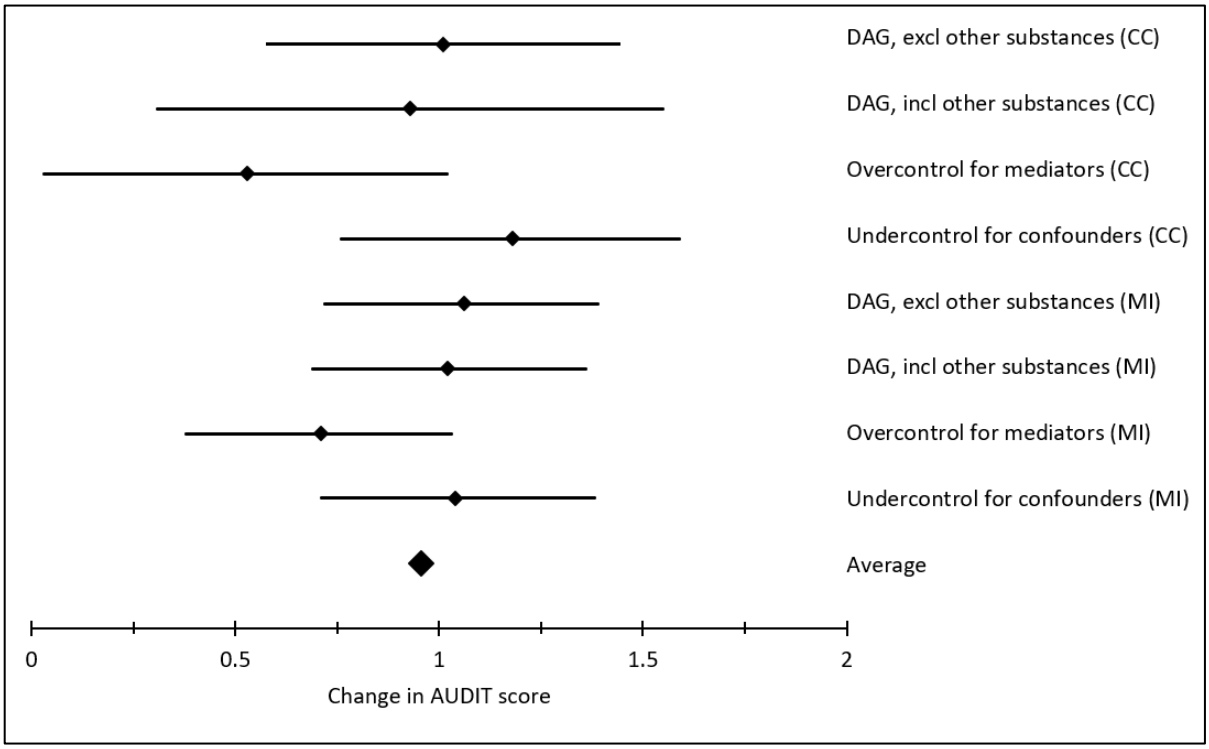


Figure 8-6: Forest plot of linear regression coefficients of AUDIT scores on maternal drinking using different adjustment strategies

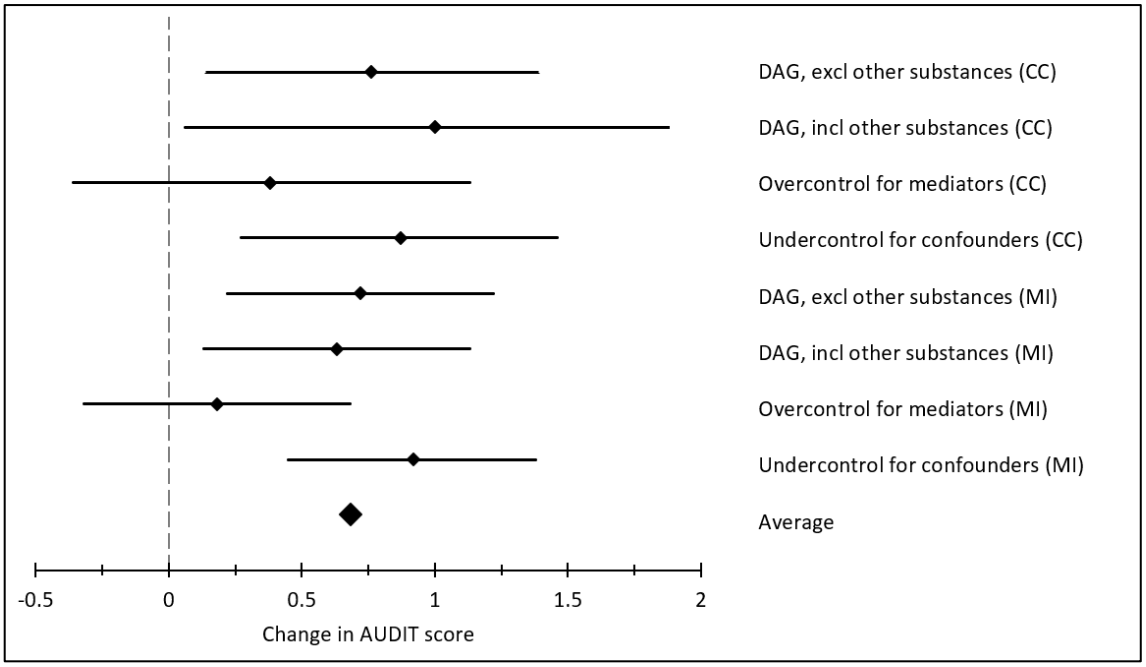


Figure 8-7: Forest plot of linear regression coefficients of AUDIT scores on maternal smoking using different adjustment strategies

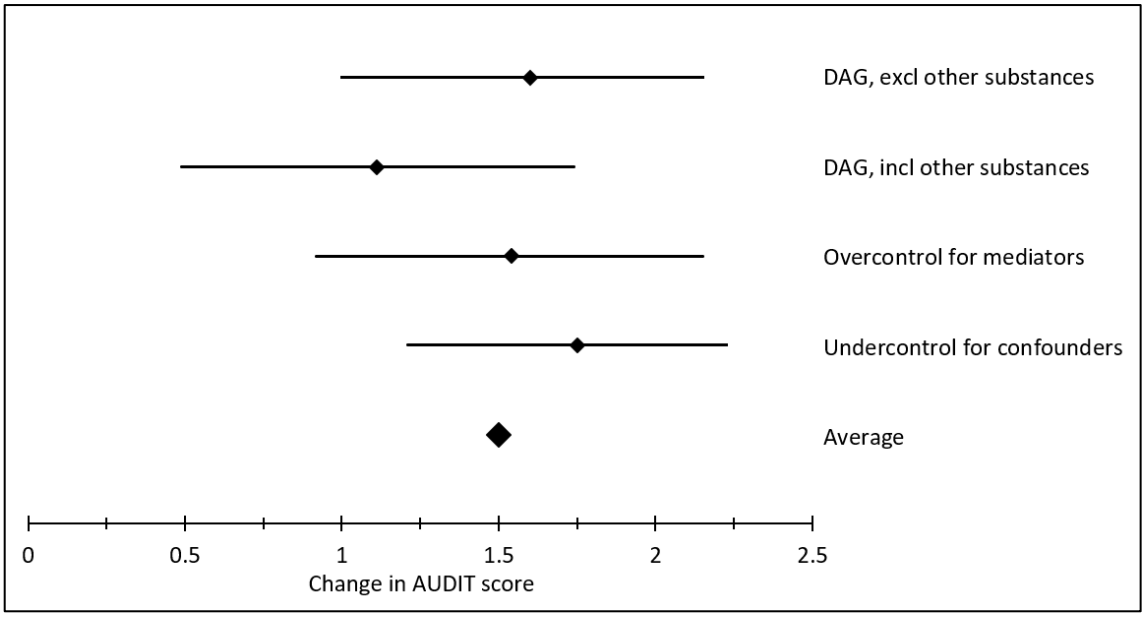


Figure 8-8: Forest plot of linear regression coefficients of AUDIT scores on father's drinking using different adjustment strategies in complete case data

8.2 Average causal effects of parental influences

This section estimates average causal effects (ACEs) for each of the parental influences under investigation in this thesis. As was discussed in greater detail in Chapter 3, this is achieved in several steps: first, a balanced propensity score was generated for each participant based on all confounders of the exposure-outcome relationship currently under analyses, as identified by the I-DAGs; then the MI data was reweighted by the inverse of the propensity score to create an 'unconfounded pseudo population' (Thoemmes and Ong, 2015); finally a regression model of AUDIT score on the exposure was fitted to this data (this regression model also factored the propensity score as a covariate, making the model doubly robust). All of these analyses, including balance checks for the propensity scores, were performed using the `teffects` suite in Stata. For each exposure, estimates were combined across the 50 models using Rubin's Rules via the `mi estimate` suite, again in Stata.

Before discussing the results, some technical aspects of the modelling process are worth explication and consideration, especially reverse causation and how the I-DAGs were operationalised.

8.2.1 Reverse causation

In instances where parental influences were measured cross-sectionally to one another, an ACE was first estimated without controlling for the cross-sectional measures, and then a second ACE was estimated including them. Similar to the above demonstration on Sets 1 and 2, this is effectively a form of sensitivity analysis which tests the robustness of estimates to assumptions of no reverse causation. There were two such instances wherein parental influences were measured cross-sectionally to one another and which were treated in this way: at age 12 (maternal drinking, maternal smoking and maternal depression) and at age 15.5 (secret-keeping, parental monitoring, time spent together, parental punishment, and parent-child tension). Further, there were two instances in which *non*-parental measures were cross-sectional to parental influences. At age 12.5 parental permissiveness was cross-sectional to early alcohol initiation, and the parental influences at age 15.5 were cross-sectional to peer anti-social behaviour, peer substance use, and adolescent IQ. Because parental permissiveness and early alcohol initiation are arguably very closely related, parental permissiveness was estimated both with and without early alcohol initiation in the propensity score, similar to the above. However, the parental influences at age 15.5 were all estimated while controlling for peer anti-social behaviour, peer substance use, and adolescent IQ.

8.2.2 Using DAGs to help estimate ACEs

It is worth further explaining how the I-DAGs informed the models. When estimating the ACE for any parental influence, any other parental influence that was measured prior was included as a confounder in the propensity score. In other words, the exposure from one model could be a confounder in another question. For example, the effect of parental permissiveness towards adolescent alcohol use was estimated while controlling for maternal drinking, maternal smoking, and maternal depression. It was also the case that any variable that was treated as an intermediate in regards to a prior parental influence, but that was measured *before* the current parental influence under analysis, was also factored as a confounder. For example, using highly simplified DAGs, Figure 8-9 below demonstrates how estimating the effect of parental monitoring requires adjusting for maternal drinking, another prior parental influence (parental permissiveness towards adolescent alcohol use at age 12.5), and another intermediate of maternal drinking (adolescent depression at age 13.5). Thus, as long as it was in keeping with the DAGs, the coding for the models was done sequentially, in that once an earlier parental influence was modelled as the exposure, it was added as a confounder for the next parental influence. The baseline confounders used were the same as for the above regression models. Unmeasured confounders included the alcohol environment, the nature of space in the local neighbourhood, and child abuse or neglect.

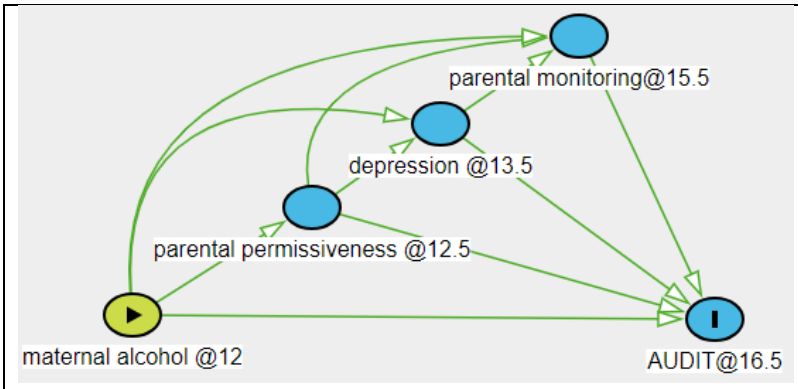
8.2.3 Results and interpretation

Table 8-4 below lists the ACE and 95% confidence intervals for each parental influence that was measured adequately in ALSPAC. It also differentiates between models which did and did not factor in the cross-sectional parental influences. The sample for each was the analytical sample from the multiple imputations (N=7,959). Monte Carlo errors from the imputations were acceptably low and are documented in Appendix H.

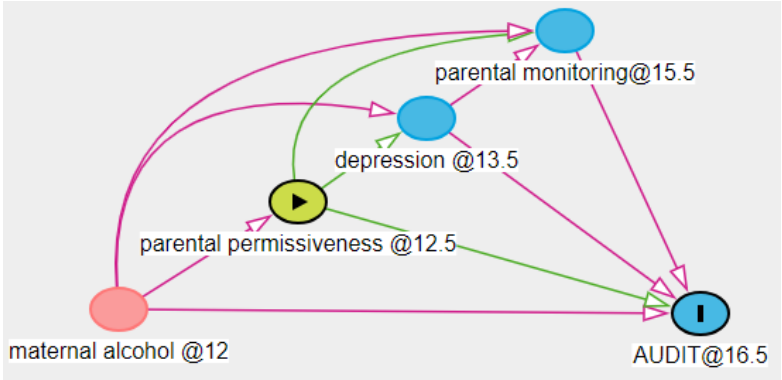
8.2.3.1 Parental influences with higher ACEs

Parental drinking, parental supply of alcohol, and low parental monitoring were the risk factors for adolescent alcohol harm that had the highest ACEs. Thus, the results from the IPW models were consistent with the review of systematic reviews in Chapter 5 (i.e. Yap et al's 2017 systematic review found that they were the risk factors with the most 'convergent' evidence). Moreover, the ACEs for

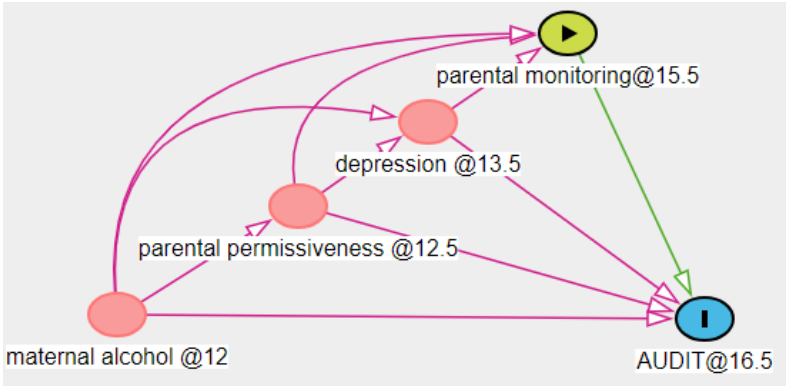
these variables were notably robust to sensitivity analysis on reverse causation between cross-sectional variables.



Maternal drinking as the exposure – no other variables should be controlled



Parental permissiveness as the exposure – maternal drinking should be controlled



Parental monitoring as the exposure – maternal drinking, permissiveness, depression controlled

Figure 8-9: Changing roles of parental influences in different models

Table 8-4: Average causal effects of parental influences on AUDIT score

Parental variable	Age	Cross sectional control?	ACE (95% C.I.)
Maternal drinking	12	No	1.08 (0.73, 1.42)
		Yes	1.04 (0.70, 1.39)
Maternal smoking	12	No	0.79 (0.28, 1.29)
		Yes	0.71 (0.20, 1.22)
Maternal depression	12	No	0.43 (0.05, 0.81)
		Yes	0.56 (-0.0, 0.75)
Parental permissiveness	12.5	No	1.67 (0.98, 2.36)
		Yes	1.58 (0.85, 2.32)
Secret-keeping	15.5	No	0.78 (0.36, 1.20)
		Yes	0.57 (0.16, 0.98)
Parental monitoring	15.5	No	1.41 (1.02, 1.79)
		Yes	1.26 (0.87, 1.65)
Time spent together	15.5	No	0.52 (0.17, 0.88)
		Yes	0.30 (-0.06, 0.67)
Parental punishment	15.5	No	0.68 (0.34, 1.02)
		Yes	0.46 (0.09, 0.82)
Parent-child tension	15.5	No	0.65 (0.26, 1.04)
		Yes	0.31 (-0.15, 0.76)

On average, participants whose mothers drank above 14 units per week had an AUDIT score 1.08 (0.73, 1.42) higher than adolescents whose mothers drank below 14 units per week. When maternal smoking and maternal depression were included in the model, there was only a slight decrease to 1.04 (0.7, 1.39). The robustness of the coefficient and the thorough ESC-DAGs approach to confounder selection thus suggest maternal drinking to be an important determinant of AUDIT score. Again, this is consistent with the wider evidence base. It also further supports the analysis plan to use maternal drinking as the exposure for mediation models. Although the effect size was notably smaller, the pattern was similar for adolescents whose mothers smoked, at 0.79 (0.28, 1.29) without cross-sectional control and 0.71 (0.20, 1.22) with cross-sectional control for maternal drinking and maternal depression.

Adolescents who had tried alcohol with parental permission by age 12.5 had AUDIT scores 1.67 (0.98, 2.36) higher on average than those who did not. This was the largest effect size for the parental influences and, if taken as a proxy for parental supply of alcohol, is in keeping with recent high-

profile findings from Mattick et al 2018 that parental supply of alcohol is a risk factor for adolescent alcohol harm (Mattick et al., 2018). Thus, this result (cautiously) comments on the policy debate as to whether parents should introduce their children to alcohol as a harm reduction strategy (Sharmin et al., 2017b). In short, it suggests that a zero-tolerance approach could be more effective, as further discussed in Chapter 10.

Participants subjected to lower levels of parental monitoring (scores of between 0 and 7 on the parental monitoring composite variable) had AUDIT scores 1.41 (1.02, 1.79) higher on average than those subjected to higher levels of parental monitoring. This figure reduced to 1.26 (0.87, 1.65) when cross-sectional parental influences were modelled to test for reverse causation (these were parental punishment, time spent together at weekends, secret-keeping, parent-child tension). This a is notably small decrease, especially considering the high volume of confounders used. This indicates that parental monitoring could be an important determinant of adolescent AUDIT score and is in keeping with findings from the review of systematic reviews.

8.2.3.2 Parental influences with lower ACEs

The review of systematic reviews found fairly strong evidence for an effect of the parent-child relationship in general, although effect sizes were smaller than for the above and there were several instances of studies finding no effects for different aspects. The ACEs were largely consistent with this more mixed evidence. For example, participants who reported higher levels of parental punishment had AUDIT scores 0.68 higher on average (0.34, 1.02) that attenuated to 0.46 (0.09, 0.82) when controlling for cross-sectional parental influences (parental monitoring, secret-keeping, parent-child tension, and time spent together at weekends). While the effect appears robust, the relatively small effect size may reflect how the review of systematic reviews found somewhat mixed evidence - in Ryan et al's systematic review, 3 of the 9 studies found no effect (Stice et al., 1998, Coker and Borders, 2001, Latendresse et al., 2008). Although, Stice et al's study had a low sample of N=216. Similarly, secret-keeping, used to represent general parent-child communication and measured by comparing adolescents who kept secrets from their parent 'often' or 'very often' to the rest, had an effect of 0.78 (0.36, 1.79) that attenuated to 0.57 (0.16, 0.98) under cross-sectional control (for parental monitoring, parental punishment, parent-child tension, and time spent together at weekends). In general, the systematic reviews were inconclusive on communication, although Yap et al indicated that there was weak evidence of a harmful effect of poor communication.

The remaining parental influences (parental involvement, parent-child tension, and maternal depression) each had a small effect that attenuated to null when the cross-sectional parental influences

were included. Parental involvement, measured by comparing adolescents who spent less than an hour with their parents over a normal weekend to the rest, had an effect size of 0.52 (0.17, 0.88) before controlling for parental monitoring, parental punishment, parent-child tension, and secret-keeping. Parent-child tension, measured by comparing adolescents who argue with their parents "most days" to the rest, had an effect size of 0.65 (0.26, 1.04) before controlling for parental monitoring, parental punishment, parental involvement, and secret-keeping. Finally, maternal depression, had a small effect with a 95% lower confidence interval close to zero (0.43 (0.05, 0.81) before controlling for maternal drinking and maternal smoking.

8.2.3.3 Summary of evidence of effects for parental influences

Maternal drinking, maternal smoking, parental permissiveness towards adolescent alcohol use and low parental monitoring were the parental influences with the most notable effects and were robust to sensitivity analysis on reverse causation. Secret-keeping and parental punishment had smaller effect sizes but were robust to cross-sectional control for other parental influences. Maternal depression, parent-child tension, and parental involvement all had smaller effects that attenuated to null when the cross-sectional parental variables were included in the model. In other words, they were not robust to sensitivity analysis of reverse causation with other parental variables. As noted earlier, it is impossible to tell whether the attenuation to the null under cross-sectional control was because the cross-sectional control models were erroneously controlling for downstream effects, or correctly controlling for confounding (upstream effects). Nonetheless, the variables that were robust to this form of sensitivity analysis matched the variables with the strongest evidence from the review of reviews, although, even using the rigorous ESC-DAGs approach to confounding bias, there was some evidence for almost all parental influences having some degree of effect on adolescent AUDIT score.

8.2.4 Sensitivity analysis of transformed outcome

As discussed in Chapter 3 and Chapter 7 the skewed nature of the outcome presents some challenges to these analyses. Thus, to keep the ACEs from this chapter in line with the estimates of total effect from the forthcoming mediation analyses, the decision was made to use the continuous outcome for the ACE models as well, albeit with a doubly robust approach (i.e. adjusted IPW models with the propensity score included as a covariate in the marginal structural model). However, as a robustness check on the above ACE models, each was repeated using z-scores of a normalised version of the AUDIT score. This variable was transformed by the 5^{th} root, and while not meeting the technical specification of a normal distribution, it was the closest fit after extensive testing (the skewness -0.24, and kurtosis was 2.38). Results are presented in Table 8-5 below. Note that, as transformations are

generally not recommended in MI data (as their analysis will yield 'biased estimates' (von Hippel, 2009)) these analyses were performed by using the imputed covariates on the observed outcome.

Table 8-5: Average causal effects in standard deviations from transformed AUDIT score

Parental variable	Age	N	Cross sectional control?	ACE (95% C.I.)
Maternal drinking	12	2,173	No	0.22 (0.13, 0.31)
		2,162	Yes	0.21 (0.12, 0.30)
Maternal smoking	12	3,106	No	0.18 (0.07, 0.29)
		3,076	Yes	0.16 (0.05, 0.28)
Maternal depression	12	3,121	No	0.06 (-0.04, 0.15)
		3,071	Yes	0.06 (-0.04,0.16)
Parental permissiveness	12.5	2,721	No	0.35 (0.15, 0.55)
		2,705	Yes	0.30 (0.08, 0.52)
Secret-keeping	15.5	2,167	No	0.13 (0.03, 0.24)
		2,139	Yes	0.11 (-0.00, 0.22)
Parental monitoring	15.5	2,167	No	0.27 (0.17, 0.39)
		2,139	Yes	0.24 (0.13, 0.35)
Time spent together	15.5	2,167	No	0.03 (-0.06, 0.12)
		2,139	Yes	0.00 (-0.09, 0.10)
Parental punishment	15.5	2,167	No	0.13 (0.04, 0.22)
		2,139	Yes	0.06 (-0.30, 0.16)
Parent-child tension	15.5	2,167	No	0.11 (0.01, 0.22)
		2,139	Yes	0.05 (-0.07, 0.17)

The coefficients in this model are measured in units of standard deviations. For example, even when controlling as per the I-DAG and for maternal depression and maternal smoking, maternal drinking has an average causal effect +0.21 (0.13, 0.31) standard deviations on the AUDIT scale. However, as products of a transformation of the observed distribution, these results are hard to interpret in terms of their implications for adolescent alcohol harm in ALSPAC at age 16.5. Indeed, they are difficult to reconcile with the above regression coefficients, odds ratios, and ACEs. However, the overall pattern across these models is markedly similar to the untransformed analyses. For example, maternal drinking, maternal smoking, parental permissiveness towards adolescent alcohol use, and parental monitoring again had the largest coefficients with tight confidence intervals, and the variables that attenuated to null above did so again here (although, unlike with the above ACEs, maternal depression had no effect regardless of cross-sectional control). As such, while these models are not informative in terms of the degree of alcohol harm that is associated with the exposures, the parental influences

that were the most harmful in the untransformed models were the most harmful in the z-score models as well. This consistency suggests that the skewed outcome variable is not generating spurious results.

8.3 Average causal effects of intermediates

Table 8-6 below lists the ACEs for the intermediates, using the same methods and sample as for the parental influences. These results are a more holistic examination of the wider causal system and thus help identify exposure-induced mediator-outcome confounders (EIMOCs) for the subsequent mediation analyses in Chapter 9. For example, if any intermediate is found to have no effect in these analyses, this suggests that the directed edge from that intermediate to AUDIT score does not exist in this data. As such, it cannot confound the relationship between a later mediator and AUDIT score, and thus cannot be an EIMOC. Figure 8-10 below uses DAGitty to explain how deleting the directed edge from an EIMOC (L) to the outcome (O) removes any confounding bias (alphabetical labels for the nodes were used for simplicity).

Table 8-6: Average causal effects of intermediate variables on AUDIT score

Socio-ecological grouping	Intermediate variable	Age	ACE (95% C.I.)
Intra-personal	Early alcohol initiation	12.5	0.87 (0.40, 1.35)
	Sensation seeking	13.5	1.76 (1.38, 2.13)
	Depression	13.5	0.68 (0.30, 1.05)
	Anti-social behaviour	14	1.20 (0.86, 1.54)
	Positive alcohol attitudes	14	1.01 (0.70, 1.31)
	Smoking	14.5	1.70 (1.21, 2.20)
	Substance use	14.5	1.10 (0.59, 1.61)
	IQ	14.5	-0.38 (-0.89, 0.14)
Inter-personal	Number of friends	13.5	0.47 (0.74, 0.86)
	Peer smoking	13.5	1.86 (1.53, 2.19)
	Peer drinking	13.5	2.11 (1.73, 2.48)
	Peer anti-social behaviour	15.5	1.52 (1.13, 1.90)
	Peer substance use	15.5	1.36 (0.95, 1.77)
	Positive relationships with teachers	16	0.56 (0.23, 0.89)
Institutional	Neighbourhood safety	13.5	0.21 (-0.11, 0.53)
	School engagement	13.5	0.29 (-0.04, 0.63)

	Academic performance	16.5	-0.12 (-0.49, 0.23)

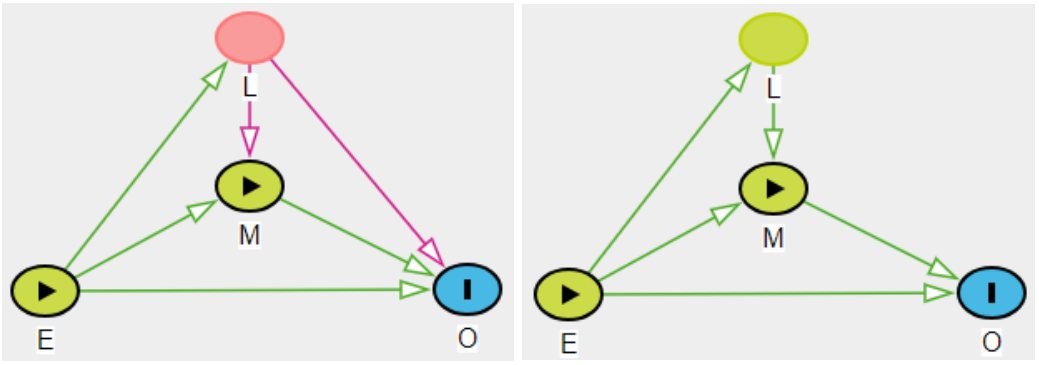


Figure 8-10: Deleting the directed edge from the EIMOC to the outcome removes confounding

Further, finding no ACE for any intermediate also indicates that it is not a mediator. The above figure suggests that there may still be a directed edge from the EIMOC to the mediator, and thus that there is a pathway from the exposure to the outcome through L (i.e. E → L → M → O). However, because an ACE is analogous to a total effect, finding no average causal effect of any intermediate means that there is no total effect to decompose. Thus, using mediation analysis to decompose the total effect of a variable with no ACE into a direct effect and an indirect effect serves no purpose.

As the intermediates are not the main substantive interest of this thesis, and due to the volume of variables involved, they were not treated with the same level of detail as the parental influences above. As such, cross-sectional controls were treated differently depending on how the ESC-DAGs processes related the variables in question. In other words, where the above analyses of parental influences systematically explored uncertainty on the direction of relationships, the models for the intermediates instead use the most plausible DAG for each. This is an efficiency measure, both reducing computational time and limiting the number of results to a more manageable total (for example, there are already 17 models to interpret).

Cross-sectional variables were dealt with in one of two ways. Firstly, the variables in question could be arranged in a causal chain, such that one was posited as a confounder for the other and was modelled accordingly. For example, number of friends was controlled for when estimating the effect of peer drinking and peer smoking, and therefore peer drinking and peer smoking were not controlled for when estimating the effect of number of friends. Other examples include sensation seeking personality being used as a confounder for depression; and positive attitudes towards alcohol being modelled as a confounder for anti-social behaviour. The other form was to not control for specific cross-sectional variables. This was done for smoking and substance use; peer smoking and drinking; peer substance use and anti-social behaviour; and neighbourhood safety and school engagement. This

form assumes that the risk of overcontrol for downstream effects is greater than that of confounding bias. Note that it was applied to concepts which have a large degree of conceptual overlap – substance use, peer substance use, etc. Of course, conceptual overlap cannot be ruled out for the former examples, but the stipulation here is that they are much less obvious. Note also that this approach comments on a limitation of an individual researcher or a small team of researchers performing an ESC-DAGs project: it is inevitable that there will be numerous relationships in the DAGs on which the researcher(s) will not have expert knowledge. This topic is revisited in Chapter 10.

8.3.1.1 Summary of evidence of effects for intermediates

Of the intra-personal effects, early alcohol initiation, sensation seeking, depression, anti-social behaviour, positive attitudes towards alcohol, smoking, and substance use all had harmful effects. Of these, only depression (0.68 (0.30, 1.105)) had an effect size that was notably smaller than maternal drinking. This could indicate that an adolescent's own substance related behaviour and attitudes, as well as personality, have a greater causal effect on AUDIT score than most parental influences. IQ was found to have no effect.

Of the inter-personal effects, positive relationships with teachers had a relatively small effect size of 0.56 (0.23, 0.89). Peer drinking, peer smoking, peer anti-social behaviour, and peer substance use were all harmful with notably high effect sizes. Peer drinking had the highest ACE of any variable tested in this chapter – adolescents who reported that their friends had consumed alcohol without parental permission by age 13.5 had AUDIT scores that were 2.11 (1.73, 2.48) higher on average at age 16.5. This is unsurprising. If we make several assumptions, namely that the social aspect of drinking is important for adolescents and that peers are an important social group in which adolescents drink (Teunissen et al., 2016, Wang et al., 2017, Giletta et al., 2012, Glaser et al., 2010, Huang et al., 2014); and that peer drinking at age 13.5 predicts peer drinking at age 16.5 (Modecki et al., 2014, Tucker et al., 2003, Windle et al., 2008), then it is likely that the quantity, frequency, and overall pattern of peer alcohol consumption will have a strong positive correlation with the adolescent's own consumption. Overall these models suggest that peer influences may be more important than parental influences by age 16.5. However, number of close friends had a notably smaller effect than the others (0.47 (0.74, 0.86)).

None of the institutional variables had an effect. It was notable that, like IQ, academic performance had no relationship with AUDIT score (the unadjusted correlation between them was 0.56 at a .000 significance level). However, while the adolescent's perception of neighbourhood safety was found to have no effect, this measure was an imperfect proxy for attachment to neighbourhood and fails to

comment at all on the alcohol environment. As such, in terms of the wider causal system of adolescent alcohol use and harm, neighbourhood effects cannot be ruled out only from this negative result. Arguably, the same argument can be made for institutional effects in general in these analyses.

8.4 Discussion

This chapter was concerned with MRO 4 (demonstrate how the method can be used to direct conventional data analysis), MRO 5 (use the method to direct data analysis under the potential outcomes framework), and SRO 4 (test the causal effects of parental influences and their intermediates on adolescent alcohol harm).

MRO 4 was concerned with demonstrating that DAGs can help improve quantitative analysis without needing to use statistical techniques from the POF. This aim was met by using regression modelling to contrast DAG-based adjustment strategies to common modelling errors. While there was no objective estimate to compare to, findings suggested that DAG-based adjustment strategies could be less likely to induce biased estimates than other approaches. This was true for overcontrol for mediators and undercontrol for confounders. Thus, using an approach such as ESC-DAGs to direct statistical models could help reduce the prevalence of these forms of bias in health research in general. These analyses also demonstrated that DAGs can be used to structure sensitivity analysis of suspected bi-directional relationships in cross-sectional data. Overall, conclusions on MRO 4 were that the ESC-DAGs protocol can fruitfully direct primary data analysis and sensitivity analysis, and that doing so can result in less biased estimates without needing to use POF statistical methods. Thus, MRO 4 can be considered as being achieved at this point.

MRO 5 (use the method to direct data analysis under the potential outcomes framework) was partially met by using inverse probability weighting to estimate average causal effects. However, MRO 5 is covered in more detail in the forthcoming mediation chapter, which uses causal mediation analysis in the form of mediational g-computation.

The reviews, methods, DAGs, and data all informed this chapter's approach to SRO 4 (test the causal effects of parental influences and their intermediates on adolescent alcohol harm). Overall there was strong support for parental influences being important determinants of adolescent alcohol harm. For example, maternal drinking was found to be harmful across all 8 models in which it was factored as the exposure. Multiple forms of sensitivity analysis were performed on maternal drinking. This included models in the CC data and the MI data, models which controlled for mediators or other

parental substance behaviours, and models using the skewed, normalised, or dichotomised AUDIT score. Maternal smoking, parental permissiveness towards adolescent alcohol use, and low parental monitoring were also harmful across different types of sensitivity analysis, with the latter pair having higher effect sizes than maternal drinking. There was some evidence that maternal depression, parent-child tension, high levels of parental punishment, time spent together at weekends, and secret keeping were harmful, although these effects were not as robust across different analyses. Father's drinking was found to be harmful in CC data analysis.

These results were thus largely in keeping with the overall indication from the review of systematic reviews that a more dysfunctional family dynamic is an important risk factor for adolescent alcohol harm on average (i.e. greater parental substance use, more extreme parenting strategies, and a less healthy parent-child relationship). One superficial if informative illustration in the data was to construct a variable in which participants who had the exposure status for two or more of maternal drinking (a parental substance behaviour), low parental monitoring (a more extreme parenting strategy), or parent-child tension (an aspect of the parent-child relationship with a 'less healthy' value) were assigned as '1' and the rest as '0'. Controlling for baseline confounders in the CC data, the exposed had an odds ratio of 2.57 (2.06, 3.22), a linear regression coefficient of 2.97 (2.45, 3.49), and an ACE of 2.98 (2.40, 3.56). While clearly a flawed post hoc analysis (e.g. only baseline confounders could be controlled for, a compound of harm should always produce more harm due to moderation, composite variables cannot meet the consistency assumption, etc.), these results were very high compared to all other models, perhaps lending some support to the assertion that a more dysfunctional family dynamic is an important risk factor.

SRO 5 (investigate mediators of parental influences on adolescent alcohol harm) was also concerned with the intermediate variables. Of the 17 tested, only 4 had no effect (IQ, educational attainment, neighbourhood safety, and school engagement). Several intermediates had higher effect sizes than maternal drinking, including sensation seeking; anti-social behaviour; smoking; peer smoking; peer drinking; peer anti-social behaviour; and peer substance use. While unconventional for a discussion section to include a table, Table 8-7 below lists effect sizes for the explanatory variables in descending order. Parental permissiveness and parental monitoring were more comparable to the peer effects, but maternal drinking was the weakest effect of the 11 in this table. However, the fact that several other effects were larger was not considered to be problematic for several reasons. Firstly, as noted in the Rossow et al systematic review, the policy interest in the relationship between parental drinking and adolescent alcohol harm is substantial. Secondly, theories such as parental modelling were supported by the weight of evidence from the review of systematic reviews; and lastly, father's drinking appeared to have a larger effect, although this was only explored in the CC data. Thus,

parental drinking on average may have an effect that is more comparable to the peer effects, although this assertion cannot be verified with this data.

Table 8-7: Summary of causes of adolescent alcohol harm

Variable	ACE	C.I. lower	C.I. upper
Peer drinking	2.11	1.73	2.48
Peer smoking	1.86	1.53	2.19
Sensation seeking	1.76	1.38	2.13
Smoking	1.70	1.21	2.20
Parental permissiveness	1.67	0.98	2.36
Peer anti-social behaviour	1.52	1.13	1.90
Parental monitoring	1.41	1.02	1.79
Peer substance use	1.36	0.95	1.95
Anti-social behaviour	1.20	0.86	1.54
Substance use	1.10	0.59	1.61
Maternal drinking	1.04	0.70	1.39

In summary the findings were that parental influences appear to be important determinants of adolescent alcohol harm, more specifically that 'more dysfunctional' family dynamics appear to have a sizeable harmful effect. In terms of effect sizes, peer influences, personality, and the adolescent's own substance behaviours were of equal importance, if not greater. However, SRO 5 was specifically concerned with *causal* effects. Thus, despite how the above analyses used ESC-DAGs to systematically direct holistic analysis of adolescent alcohol harm using POF statistical methods, it is necessary to explicitly consider whether the assumptions for valid causal inference hold.

8.4.1 Valid causal inference

The assumptions for valid causal inference outlined in Chapter 3 were conditional exchangeability (i.e. no unmeasured confounding), positivity (there must always be a positive probability of being exposed for every combination of confounders), no interference (no participant's exposure status should influence that of another), and consistency (the meaning of being exposed should be the same across all of the exposed).

Conditional exchangeability cannot be assumed for the above analyses because several potential confounders were not used, including the alcohol environment, the influence of the extended family, and child abuse & neglect. However, while the thoroughness of the ESC-DAGs approach cannot account for unmeasured concepts, it does lend confidence to the position that whatever confounding bias might remain could be small. One extension to these analyses would be to use simulated data to try and quantify the strength of associations unmeasured confounders would need to have with any given exposure and the outcome to attenuate that exposure's effect (Greenland, 1996).

The positivity assumption was generally not a problem for these analyses due to the ability to investigate unbalanced propensity scores using Stata's `teffects` suite, and respecify them. In other words, the programme allows the researcher to ensure that there is always a positive probability of a participant being exposed at each stratum of the propensity score.

A breach of the no interference assumption requires that that the parental influences on one adolescent measurably influence the parental influences of another adolescent. This could be problematic when adolescents live in the same neighbourhood or attend the same school, which is likely for at least a small number of participants given how ALSPAC is localised to the city of Bristol and its surrounds. It is not the case that the statistical methods used to account for neighbourhood and school effects would attenuate a breach of the no interference effect. Because interference results in clusters of highly correlated data, it could amplify effect sizes and artificially increase precision of estimates. However, this is perhaps more problematic for the peer effects than parental influences. Nevertheless, no interference may be an implausible assumption.

Consistency is the causal assumption that is most problematic for these analyses. Arguably, the consistency assumption is problematic in general when working with survey response data. For example, while the exposure of maternal drinking above 14 units per week is a clinically informed guideline, even this is not 'consistent' in the strictest sense. Recall that the definition of consistency means that the exposure should be 'well defined' such that it is the same for everyone. This is clearly not the case for this variable – drinking 14 units in one week could mean drinking all of them in one sitting or drinking 2 units per day (e.g. a glass of wine). Indeed, while these drinking patterns are considered harmful, their consequences may be characteristically different. Such issues are pervasive across the parental and intermediate variables used in the above analyses. As such consistency cannot be assumed in general. One potential extension to these analyses could be to perform sensitivity analysis on misclassification of the exposures (Greenland, 1996). While this would not be able to ascertain whether or not the exposure in question is 'consistent' in a technical sense, it could give an indication of how sensitive it is to exposed participants being mistakenly measured as unexposed and vice versa.

A causal interpretation of these estimates would require that each of these assumptions is met. As it is plausible that this is not the case, especially for conditional exchangeability and consistency, a causal interpretation may be overly ambitious. Thus, in terms of SRO 5, this thesis does not claim that ACEs are true causal effects. However, given the rigorous approach to confounder selection, the degree of sensitivity analysis, and the sophisticated statistical techniques used, estimates can still be argued to have a useful degree of external validity.

It is worth considering the underlying reasons for why the estimates produced in this chapter are not true causal effects - specifically, the use of ALSPAC as the sole data source for analysis. Several variables were not measured, and those that were tended not to meet the causal assumption of consistency. These are important observations. Firstly, ALSPAC is an excellent data source in a relative sense, especially in terms of the breadth of its measures. Thus, if ALSPAC data cannot be used for valid causal inference, this perhaps indicates that statistical methods may be advancing ahead of the ability to actually employ them in observational data. Secondly, when discussing data in a general sense, it is difficult to avoid the concurrently ubiquitous narrative on 'big data'. Even if it is assumed that ALSPAC's sample size is too small to qualify as 'big data', larger sample size is not a solution to problems of conditional exchangeability and consistency. As such, this chapter's analyses suggest that the rhetoric surrounding big data is perhaps overly optimistic, at least in terms of quantifying valid causal effects. As discussed in Chapter 7, the ideal data for this analysis would have repeated measures of all key variables. Short of this, a partial solution if perhaps unpragmatic solution could be to use several data sources. Then the process of replacing the conceptual I-DAG with a data I-DAG could be repeated for each data source, as well as the subsequent analysis.

8.4.2 Other technical limitations

One of the key limitations of these analyses and of the forthcoming mediation models was how rigidly the timing of measurement was treated, especially in terms of how it shaped the DAGs and subsequent statistical models. For example, while the effect of each intermediate was estimated while controlling for maternal drinking, maternal smoking, maternal depression, and parental permissiveness towards adolescent alcohol use, almost none of them were able to control for the parent-child relationship or parenting variables that were only measured once at age 15.5. As such, it may be that the effect sizes for the intermediates would be attenuated by secret keeping and parental punishment, for example, if earlier measures had been available. In other words, the intermediate models may be subject to unmeasured confounding by aspects of parenting or the parent-child relationship.

It is also worth discussing the implications of working with only binary exposures. For this thesis, the decision to follow this format was made for three reasons. First, weighting approaches are more unstable with non-binary exposures. Second, simply to reduce the complexity of interpreting results for the 27 different explanatory variables under investigation. Third, there was a lack of normally distributed continuous variables. However, this means that anything obfuscated in one analysis is likely to be obfuscated in others. Thus, the decision to dichotomise each exposure may obfuscate some of the variation in the above causal models. An example mentioned above was one study from the review of systematic reviews in Chapter 5 that found that very low levels of parental discipline had a similar association with adolescent alcohol harm as very high levels of parental discipline. However, given how the emphasis of this research is on understanding parental influences on adolescent alcohol harm within a wider holistic causal system, breadth was preferred over depth, thus the focus on average effects of numerous exposures.

Chapter 9 Results 2: Mediators of maternal drinking

This chapter pertains to MRO 5 (use the method to direct data analysis under the potential outcomes framework) and SRO 5 (investigate mediators of parental influences on adolescent alcohol harm). These are the final two research objectives. This chapter also concludes the data analysis plan (estimate ACEs for the hypothesised causes of adolescent alcohol harm and investigate how they mediate the effect of maternal drinking).

As described in some detail in Chapter 3, the technique used for mediation analysis is mediational g-computation, which decomposes the total causal effect (TCE) of maternal drinking into randomised interventional analogues of the natural indirect effect (NIE) and natural direct effect (NDE). This allows for comment on which mediators could potentially be intervened upon to reduce the harmful effect of the exposure (maternal drinking) on the outcome (adolescent alcohol harm), as well as the harmful effect of the mediator. The data I-DAG from Chapter 7 directs the model for each mediator although the specific model is briefly described in each instance. This chapter begins with a preliminary discussion on implementation before presenting results for parental and non-parental mediators of maternal drinking. It closes with a brief interpretation of the results and discussion on the limitations of the methods, which is expanded upon in Chapter 10.

9.1 Mediation models: preliminaries

This section considers the different roles variables could take on across different models (e.g. mediator, EIMOCs, etc.) and some of the logistical problem-solving that was required when applying a user-written Stata command to so many variables across 50 imputed datasets. Box 9-1 is included below for ease of reference as a brief summary of causal mediation analysis.

Box 9-1: Causal Mediation

Mediation: Expanding on Pearl (Pearl et al., 2016), a variable M is a mediator of the effect of X on Y if M depends on X for its value, and Y depends on M for its value: X → M → Y. The potential mediators of maternal drinking on adolescent AUDIT score at age 16.5 were identified by estimating ACEs in the previous chapter. These correspond to M → Y. Thus, as mediation analysis seeks to establish the combined pathway of X → M → Y, the directed edge X → M becomes a prerequisite.

Exposure induced mediator outcome confounders (EIMOCs): An EIMOC is another mediator that is assumed to affect the mediator of interest. Failing to control for an EIMOC allows it to confound the indirect effect of the mediator of interest, and controlling for it biases the direct effect. Mediational g-computation can resolve this paradox.

Nested counterfactuals:

The total causal effect (TCE): $\text{TCE} = E[Y(1, M(1))] - E[Y(0, M(0))]$

The TCE compares the scenario in which all are exposed and the mediator takes the value that it would have taken under exposure, to the scenario in which none are exposed and the mediator takes the value that it would have taken under no exposure. As it contrasts the exposed and unexposed it is analogous to the ACE of maternal drinking of +1.08 on the AUDIT scale from Chapter 8.

The natural indirect effect (NIE): $\text{NIE} = E[Y(1, M(1))] - E[Y(1, M(0))]$

The NIE compares the scenario in which all are exposed and the mediator is set to the value that it would have taken under exposure, to the scenario in which all are exposed and the mediator is set to the value that it would have taken under no exposure. The natural direct effect (NDE) equals the difference between the TCE and NIE.

Randomised interventional analogues: Because the potential outcome $E[Y(1, M(0))]$ can never be observed due to Holland's fundamental problem of causal inference (Holland, 1986), randomised interventional analogues of the NIE and NDE are used instead. This involves taking a random draw from the distribution of the simulated mediators, rather than setting each individual participant to their individual value. Each NIE and NDE in this chapter is estimated this way.

Mediational g-computation: Mediational g-computation is a procedure for sequentially simulating counterfactual distributions of each EIMOC and the mediator based on associations in the observed data to produce a potential outcome. For example, the potential outcome $E[Y(1, M(0))]$ is estimated by simulating each EIMOC and the mediator, based on associations in the unexposed and on simulated versions of the EIMOCs based on their associations with the unexposed.

9.1.1 Candidate mediators

The first stage of the data analysis plan was completed in Chapter 8 by using IPWs to calculate ACEs for all parental and non-parental influences identified in the data I-DAG. Not only did this identify which variables might be causes of adolescent alcohol harm, it also indicated which could not be mediators of the effect of maternal drinking in the ALSPAC data. Assume maternal drinking is assigned as X and each intermediate (i.e. all variables measured cross-sectionally to or after maternal drinking) is assigned as M. Then a statistically significant ACE for M implies that, because the directed edge M → Y is viable, the path X → M → Y might also be viable. A total of 19 intermediates had statistically significant ACEs, meaning that 19 potential mediators of maternal drinking were identified in Chapter 8. This included eight parental influences (maternal smoking, maternal depression, parental permissiveness towards adolescent alcohol use, parental monitoring, parent-child tension, parental punishment, parental involvement and secret-keeping) and 11 non-parental variables which were mainly intra-personal (personality and behaviour) or inter-personal (peers and teachers). Specifically, these were: age of alcohol initiation; sensation-seeking; depression; peer drinking; peer smoking; positive attitudes towards alcohol; anti-social behaviour; smoking; substance use; peer anti-social behaviour; peer substance use; and positive relationships with teachers. Four variables were not tested as mediators because they had no causal effects in Chapter 8, and thus there was no evidence for the presence of the directed edge M → Y for each. These were neighbourhood safety, IQ, academic performance, and school engagement. Unmeasured mediators were beliefs, economic activity, extended family effects, parental supply of alcohol, parental substance attitudes, family coherence, parental support, and alcohol-specific communication.

9.1.2 EIMOCs

The data I-DAG suggested that most mediators could act as an EIMOC for all later mediators. For example, parental monitoring at age 15.5, even ignoring its cross-sectional parental influences, had a pool of 15 potential EIMOCs. As covered in more detail below, this required an unfeasible degree of data simulation for each mediator (e.g. 15 EIMOCs for each potential outcome across 50 different datasets by the number of bootstrap replications). As a compromise a subset of the intermediates were selected to operate as the EIMOCs across all mediation models. The process of determining this subset was conceptual in the first instance in that groups of highly related variables were identified from the data I-DAG in reference to the conceptual I-DAGs recombined concepts (for example, peer effects). The next step involved using the data to select the most appropriate variable(s) from each group. This involved two associations: firstly, the ACE of each intermediate on the outcome of adolescent AUDIT score as estimated in the previous chapter; and secondly, the associations between the intermediates themselves calculated using unadjusted logistic regression models in the complete

case data. Seven conceptual groups were identified: parental characteristics; parenting; the parent-child relationship; personality; alcohol attitudes & behaviours; substances & mental health; and peer effects.

1. Because maternal smoking and maternal depression were both measured cross-sectionally to the exposure, it is not possible to discern how much of any indirect effect through them (a prerequisite for being used as an EIMOC) would represent confounding rather than mediation. As such they were not used as EIMOCs except for a few exceptions noted below.
2. The group of parental influences at age 15.5 were measured too late to be included as EIMOCs for the majority of intermediates. One implication of this was that the parent-child relationship could not be used as an EIMOC. Parenting did have an earlier measure in the form of parental permissiveness towards alcohol use at age 12.5. As such it was included as an EIMOC. It also had the highest ACE of any parental variable at 1.67.
3. As the only measure of personality, sensation seeking personality was included. It also had a notably high ACE of 1.76.
4. Positive alcohol attitudes was used in place of early initiation as it had a larger ACE (1.01 compared to 0.87) and was measured closer to the outcome, thus it was reasoned that the effect of initiation on the outcome would be largely mediated by attitudes.
5. Smoking and substance use were used in place of anti-social behaviour and depression because of the degree of conceptual overlap, because they are more immediately related to drinking, and because of their ACEs (1.70 and 1.10 compared to 1.20 and 0.68). Smoking was associated with both depression (OR: 2.02; 1.74, 2.34) and anti-social behaviour (OR: 4.75; 4.15, 5.44), as was substance use (OR: 2.12; 1.822, 2.50 & OR: 4.33; 3.74, 5.01).
6. Peer drinking was selected in place of other peer effects because it was the peer effect that is most conceptually related to adolescent drinking and because it had the highest ACE (i.e. it had an ACE of 2.11 compared to 0.47, 1.86, 1.52, 1.35 for peer number, peer smoking, peer anti-social behaviour, and peer substance use respectively). It was also associated with each: peer number (OR:1.33; 1.18, 1.52); peer smoking (OR: 10.96; 9.64, 12.46); peer anti-social behaviour (OR: 3.75; 3.30, 4.25); and peer substances (OR: 4.14; 3.64, 4.71).

Thus, the maximum number of EIMOCs for any model was reduced from 15 to 6. They covered parenting; adolescent personality; alcohol attitudes; substances & mental health; and peer effects:

1. Parental permissiveness towards adolescent alcohol use at age 12.5
2. Sensation-seeking personality at age 13.5
3. Peer drinking at age 13.5
4. Attitudes to alcohol at age 14
5. Smoking at age 14

6. Substance use at age 14

Parental characteristics and the parent-child relationship were not included in this group of EIMOCs because of the timing of their measurement (parental characteristics were too early, and the parent-child relationship was too late). However, a degree of flexibility was required due to how intermediates that were measured comparatively early had few potential EIMOCs. For example, parental permissiveness towards adolescent alcohol use was measured at 12.5 – before any of the variables selected to perform as EIMOCs. As such, when estimating the indirect effect through parental permissiveness, maternal depression was factored as an EIMOC. Similarly, maternal depression and maternal smoking were both measured at the same time as the exposure at age 12. As such, the only potential EIMOCs for each was the other. Therefore, when estimating the indirect effect through maternal smoking, maternal depression was used as an EIMOC, and vice versa. One consequence both of this strategy and the timing of measures was that different indirect effects were estimated with varying degrees of control for intermediate confounding. This is revisited in the discussion below. A final consideration was which EIMOCs to use when the indirect effect in question was via one of the six selected EIMOCs (i.e. when an EIMOC was being used as a confounder). The approach taken was to include one of the variables from the same conceptual grouping when possible (for example, peer drinking controlled peer smoking and vice versa). The specific variables used as EIMOCs for the parental mediation models are given in Table 9-2, and for the non-parental mediators in Table 9-6 below.

9.1.3 Baseline confounders

The baseline confounders used were family history of alcohol problems, maternal age, birth weight, maternal education, maternal occupation, family structure, older siblings, maternal stress, smoking during pregnancy, IMD, BMI, urbanity, ethnicity, pubertal timing and sexuality. Two measured confounders were excluded altogether because they were not associated with either the exposure or outcome, and thus cannot meet the definition of a baseline confounder, an EIMOC, or a mediator-outcome confounder. These were adolescent sex and maternal religiosity. Unmeasured confounders included the alcohol environment, the nature of space in the local neighbourhood, and child abuse or neglect.

9.1.4 Dealing with uncertainty

There are two sources of uncertainty that are worth discussion. First, because mediational g-computation uses Monte Carlo simulation to generate counterfactuals and potential outcomes, a degree of simulation error is to be expected. Thus, as noted in Chapter 3, bootstrap resampling is employed to calculate 95% confidence intervals for each estimate to quantify the uncertainty for each mediation model. The default resampling in `gformula` is 1000. However, given that the volume of mediators being tested was large, 1000 replications was prohibitively high in terms of computation time. This is discussed further below.

A second source of uncertainty was due to how Rubin's rules were not applied to combine estimates across imputed datasets (uncertainty between imputations). Stata has an in-built feature for combining estimates across imputed datasets, but `gformula` is not yet compatible. Thus, for each mediator, there were 50 different estimates of the TCE, NDE, NIE and CDE, as well as their confidence intervals. Across the 19 potential mediators, this is a total of 11,400 results. To make this level of information intelligible the mean for each estimate and confidence interval is presented, as well as the proportion of datasets in which the confidence interval crossed the null as a measure of variability. The ability to rely on averages was possible because the simulation techniques inherent in both MICE and `gformula` rely on randomness, thus a normal distribution was approximated for each estimate. Indeed, the `gformula` estimates are averaged across simulated potential outcomes. The Monte Carlo errors for each ACE are included in Appendix H and were found to be low.

9.1.5 Dealing with high computation time

If computation time was not a factor, the ideal approach would have involved estimating the TCE, NDE, NIE, CDE, and their confidence intervals for each of the 19 intermediates identified in the previous chapter across all 50 datasets using 1000 bootstrap replications. However, the estimated computation time given no programming errors was over 4 weeks on a standard single-core desktop PC using Stata 14 SE. Notably, the journal article that introduced `gformula` features a clear warning on computation time (Daniel et al., 2011). As alluded to in Chapter 3 and above, this is largely attributable to how each EIMOC and mediator must be simulated under each nested counterfactual. Reducing the computation time was a non-trivial challenge. Four different tactics were implemented to successfully reduce the computation time for the whole analysis to less than 1 week without compromising the models. Each is briefly explained here.

Hardware and software

After consulting with experts at Stata to ensure that increasing the number of CPU cores would reduce computation time, a 16 core PC and Stata 16 were purchased for this analysis in the first instance, and for use by the MRC/University of Glasgow thereafter. This was jointly funded by the MRC/CSO SPHSU and by the MRC Doctoral Training Programme.

Number of EIMOCs modelled

As noted above the results from the previous chapter suggested that as many as 15 EIMOCs should be modelled when estimating the indirect effect of the parenting mediators at age 15.5, excluding other cross-sectional parental variables. If 1000 bootstrap replications are used across 50 datasets, then each of these 15 EIMOCs as well as the mediator must be simulated 50,000 times for each potential outcome. Each mediation model uses 7 different potential outcomes. Thus, more than 5 million simulated distributions would be required for each mediator (i.e. (15000*50)*7)). However, failing to model for EIMOCs heavily caveats any conclusion, and in fact negates the benefits of using mediational g-computation. As discussed above, the pool of EIMOCs was reduced to a subset of six variables which had high ACEs and were hypothesised to have the highest potential to reduce intermediate confounding bias.

Number of datasets and replications used

Each intermediate has been posited thus far as having a causal effect on the outcome $M \rightarrow Y$. However, this does not necessitate that they will mediate the effect of the exposure $X \rightarrow M \rightarrow Y$. As such, it would have been inefficient to model each intermediate across the full 50 datasets with high replications given the potential that many of them would not mediate the effect of maternal drinking. Instead, the analysis was split into two waves. The first wave assesses which intermediates were worth investigating further by modelling all mediators in the same 20 datasets using 250 bootstrap replications in each. A forgiving approach was taken here in that only the intermediates which failed to mediate in the majority of the datasets were excluded at this stage. Those that did appear to have some degree of mediation were then modelled across the full 50 datasets with 1000 replications. In other words, the approach was to efficiently differentiate mediators from non-mediators before attempting to calculate reliable estimates for the former.

Exposure-mediator interaction

As discussed in Chapter 3, the methods which come under the umbrella of causal mediation use potential outcomes to estimate indirect effects in the presence of exposure-mediator interaction, and in the case of mediational g-computation, in the presence of EIMOCs as well. One direction the current analysis could have taken was to model each mediator twice, once with an exposure-mediator

interaction, and once without. However, this would have effectively doubled the computation time for wave 1. As such, each intermediate was tested for interaction with the exposure in models on the outcome. Only parental monitoring and parent-child tension were found to have an interaction effect with maternal drinking. Testing was extensive and was initially carried out using bi-factorial ANOVA for the binary versions of each intermediate with the binary version of maternal drinking on the continuous AUDIT score. Further testing was done using multiple regression across 3 different versions of maternal drinking (continuous, binary, and a 3-category version of the binary variable in which greater than 21-unit consumption acted as the 3rd category). Any intermediate that had a non-binary version was tested in a multiple regression model using that version as well (e.g. sensation-seeking personality, the composite parenting variables, etc.).

9.2 Parental mediators of maternal drinking

Including how parental monitoring and parent-child tension were both modelled with and without the exposure-mediator interaction term, ten mediation models were performed for the parental influences in wave 1 (250 reps, 20 datasets). Results are presented in Table 9-1 below. As the role of this stage of analysis is effectively to determine which variables might be mediators, these results are simplified to focus on the average NIE alongside the average lower and upper 95% confidence intervals. The number of datasets in which the confidence interval crossed the null for each mediator is also stated. The full results for these models are presented in Appendix I.

Table 9-1: Natural indirect effects of parental influences

Parental variable	Average NIE	Average L C.I.	Average U C.I.	Proportion of models for which confidence intervals crossed the null
Maternal smoking	0.039	0.016	0.062	3/20 (15%)
Maternal depression	0.007	-0.004	0.019	20/20 (100%)
Parental permissiveness	0.018	-0.007	0.044	13/20 (65%)
Parental monitoring	-0.064	-0.147	0.018	13/20 (65%)
Parental monitoring with interaction term	-0.060	-0.138	0.018	12/20 (60%)
Parental punishment	-0.073	-0.115	-0.031	0/20 (0%)
Parent-child tension	-0.039	-0.084	0.007	14/20 (70%)
Parent-child tension with interaction term	-.045	-0.100	0.009	15/20 (75%)
Secret-keeping	-0.017	-0.051	0.017	12/20 (60%)
Parental involvement	-0.001	-0.026	0.023	15/20 (75%)

Using the above strategies on EIMOC selection, very little evidence was found for parental mediators of maternal drinking in the wave 1 mediation models. The only variables that had significant effects were maternal smoking and parental punishment. Each of the other parental influences had average confidence intervals which crossed the null. For each of these the majority of the 20 individual models also had confidence intervals which crossed the null. As per Table 9-2 below, parental monitoring, parental involvement, secret-keeping, and parent-child tension were all modelled on the full set of 6 EIMOCs, parental permissiveness was only modelled on maternal depression, and maternal depression was modelled with no EIMOCs.

Table 9-2: EIMOCs used for each parental mediation model

Posited mediator	Total EIMOCs	Variables
Maternal smoking	1	Maternal depression
Maternal depression	0	NA
Parental permissiveness	1	Maternal depression
Parental monitoring	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use
Parental punishment	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use
Parent-child tension	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use
Secret-keeping	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use
Parental involvement	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use

Maternal smoking had an average indirect effect of 0.039 (3.6% of the TCE) with three models having confidence intervals that crossed the null. It was modelled with only maternal depression as a potential EIMOC as this was the only other intermediate that was cross-sectional to maternal drinking. Parental punishment had an average indirect effect of -0.073 (6.8% of TCE) and none of the 20 models had confidence intervals crossing the null. The implications of a negative indirect effect are discussed in more detail below. It was modelled with the full set of six EIMOCs. Thus, the only parental influences that were carried forward to the wave 2 mediation analysis (1000 reps, 50 datasets) were maternal smoking and parental punishment.

The full results for the wave 2 versions of maternal smoking and parental punishment are presented in Table 9-3 below. The table features the average point estimate and 95% confidence intervals for the TCE, NDE, and NIE, as well as the percentage mediated and the proportion of insignificant models.

Table 9-3: Parental mediators

Parental Mediator		Effect size	Lower C.I.	Upper C.I.	Proportion of models for which confidence intervals crossed the null
Maternal smoking	TCE	1.066	0.817	1.315	3/50 (6%)
	NDE	1.025	0.776	1.273	
	NIE	0.041	0.018	0.065	
	% of TCE	3.8			
Parental punishment	TCE	1.048	0.800	1.295	0/50 (0%)
	NDE	1.123	0.875	1.370	
	NIE	-0.075	-0.116	-0.034	
	% of TCE	7.2			

Neither maternal smoking nor parental punishment mediates a substantial portion of the effect of maternal drinking on AUDIT score. The randomised interventional analogue of the natural indirect effect for maternal smoking, estimated while accounting for intermediate confounding by maternal depression, was 0.041 (3.8% of the TCE of 1.066). This was very similar to the wave 1 result. This indicates that, for adolescents aged 12 whose mother drinks above the CMO guidelines, there is an increased probability of their mother being a smoker, which in turn increases AUDIT score at age 16.5. The NDE, as the difference between the TCE and NIE, was 1.025. Note that the TCE of maternal drinking on AUDIT (1.07), is very similar to the ACE estimated in the previous chapter (1.08), as expected. Of the 50 datasets, 3 estimated a lower confidence interval below zero and the average lower confidence interval was close to zero at 0.018. Accordingly, mediational g-computation indicates that, while maternal smoking may mediate the effect of maternal drinking, the scale of this mediation is small compared to the TCE.

Parental punishment had a negative NIE and thus its NDE was higher than its TCE. This may appear counter-intuitive, but this thesis' definition of a causal mediator is a useful heuristic. Mediation was defined by saying that a variable M is a mediator of the effect of X on Y if M depends on X for its value, and Y depends on M for its value. Based on the ACEs in the previous chapter, the assumption

is made that the directed edge from parent-child tension to AUDIT score has a harmful effect (an increase of 0.68). Thus, Y depends on M for its value in that being exposed to a higher degree of punishment results in a higher average AUDIT score. We also assume that the same is true for being exposed to maternal drinking above 14 units per week, thus Y depends on both X and M for its value (an increase of 1.08). It is thus the directed edge from maternal drinking to parent-child tension that is of interest. The negative NIE implies that being exposed to maternal drinking results in being unexposed to high levels of parental punishment (i.e. lower levels of parental punishment). Thus, M depends on X for its value. Higher levels of maternal drinking appear to cause lower levels of parental punishment which in turn cause lower AUDIT scores. (Lin et al., 2017).

The interpretation of the TCE and NDE are also not as straightforward compared to the maternal smoking model. Firstly, note that the TCE (1.048) is still equal to the sum of the NDE (1.123) and the NIE (-0.075). However, the NDE is now higher than the TCE. This is not because the NDE is estimated by taking the NIE from the TCE – the NDE is estimated separately by contrasting different potential outcomes compared to the NIE. Implementations of mediational g-computation elsewhere have interpreted a higher NDE than TCE as indicating that the seemingly protective mediated effect of the exposure (maternal drinking) through the mediator (parental punishment) may "partially mask the detrimental effect" of the exposure on the outcome (adolescent alcohol harm) (Lin et al., 2017). This articulates the idea of 'suppression', whereby including a third variable (the mediator) increases the relationship between the exposure and the outcome (MacKinnon et al., 2000). It also makes the interpretation of the ratio of the indirect effect to the total effect unclear. Nevertheless, compared to maternal smoking the NIE for parental punishment was large at around 7.2% of the TCE (and 6.7% of the NDE), none of the 50 models had a confidence interval which crossed the null, and the upper confidence interval was further from 0 (-0.034). This indicates that the effect of maternal drinking on AUDIT sees a moderate degree of mediation through parental punishment. Specifically, higher levels of maternal drinking results in lower levels of parental punishment which in turn results in lower AUDIT scores.

9.3 Non-parental mediators of maternal drinking

The other potential mediators were the intra-personal and interpersonal intermediates from the data I-DAG for which there was evidence of an ACE in Chapter 8. As per Table 9-4 below, the wave 1 mediation models (250 reps, 20 datasets) indicated that only attitudes towards alcohol, adolescent smoking, adolescent substance use, peer smoking, and peer drinking mediated the effect of maternal drinking on AUDIT score at age 16.5. For the rest (early alcohol initiation, sensation seeking,

depression, anti-social behaviour, number of friends, peer anti-social behaviour, peer substance use, and relationships with teachers) the average lower confidence interval crossed the null and/or at least half of the models had a confidence interval which crossed the null.

Table 9-4: Natural indirect effects of non-parental influences

Variable	Average NIE	Average L C.I.	Average U C.I.	Count of models for which confidence intervals crossed the null
Early alcohol initiation	0.010	-0.013	0.033	16/20 (80%)
Sensation seeking	0.007	-0.020	0.033	20/20 (100%)
Depression	0.001	-0.011	0.013	20/20 (100%)
Anti-social behaviour	0.046	-0.019	0.112	17/20 (85%)
Alcohol attitudes	0.091	0.048	0.135	0/20 (0%)
Smoking	0.178	0.077	0.279	2/20 (10%)
Substance use	0.148	0.049	0.247	3/20 (15%)
Number of friends	0.015	-0.003	0.034	15/20 (75%)
Peer smoking	0.088	0.027	0.149	1/20 (5%)
Peer drinking	0.193	0.122	0.264	0/20 (0%)
Peer anti-social behaviour	-0.050	-0.096	-0.003	10/20 (50%)
Peer substance use	0.010	-0.028	0.048	20/20 (100%)
Relationships with teachers	-0030	-0.058	-0.002	10/20 (50%)

Table 9-5 below presents the full results for the wave 2 models (1000 reps, 50 datasets). It is immediately clear that the effect sizes for the NIE are larger than for the parenting mediators (parental smoking and parental punishment). Of the intra-personal effects, attitudes towards alcohol had the smallest indirect effect of 0.086 (8.1%). As per Table 9-6 below this mediator was modelled with four EIMOCs (parental permissiveness, sensation seeking, peer drinking, and smoking). None of the 50 models had confidence intervals crossing the null. Thus, positive attitudes towards alcohol appear to mediate the harmful effect of maternal drinking above CMO guidelines on adolescent AUDIT score. Adolescent smoking was modelled with four EIMOCs (parental permissiveness, sensation seeking, peer drinking, attitudes to alcohol) and had a much larger indirect effect of 0.174 (17.4%), although three of the 50 models had confidence intervals crossing the null. This suggests that adolescent smoking, as a substance behaviour that is closely related to adolescent drinking, mediates a substantial portion of the effect of maternal drinking on AUDIT score. Adolescent substance use was similar with a large indirect effect of 0.137 (13.2%). However, 10 of the 50 models had a confidence interval which crossed the null. This may be attributable to how it was modelled with five rather than four EIMOCs (parental permissiveness, sensation seeking, peer drinking, alcohol attitudes, and

smoking). As such, while mediational g-computation suggests that there is a large degree of mediation via adolescent substance use, this finding is less certain compared to alcohol attitudes and substance use.

Table 9-5: Non-parental mediators of the effect of maternal drinking on adolescent AUDIT

Mediator		Effect size	Lower C.I.	Upper C.I.	Count of models for which confidence intervals crossed the null
Positive attitudes towards alcohol	TCE	1.068	0.823	1.317	0/50 (0%)
	NDE	0.982	0.738	1.227	
	NIE	0.086	0.043	0.130	
	% of TCE	8.1			
Adolescent smoking	TCE	1.002	0.752	1.253	3/50 (6%)
	NDE	0.828	0.586	1.070	
	NIE	0.174	0.076	0.272	
	% of TCE	17.4			
Adolescent substance use	TCE	1.035	0.783	1.287	10/50 (20%)
	NDE	0.898	0.654	0.142	
	NIE	0.137	0.038	0.237	
	% of TCE	13.2			
Peer smoking	TCE	1.009	0.764	1.254	2/50 (4%)
	NDE	0.918	0.676	1.159	
	NIE	0.092	0.029	0.154	
	% of TCE	9.1			
Peer drinking	TCE	1.032	0.785	1.278	0/50 (0%)
	NDE	0.840	0.597	1.083	
	NIE	0.192	0.121	0.263	
	% of TCE	18.6			

The interpersonal mediators were peer smoking and peer drinking. Peer smoking modelled for three EIMOCs (parental permissiveness, sensation seeking, and peer drinking) and had an NIE of 0.092

(9.1%). Two of the 50 models found no effect. When modelling peer drinking as the mediator, peer smoking was swapped in as an EIMOC and so modelled for three EIMOCs (parental permissiveness, sensation seeking, and peer smoking). Peer drinking had a very large NIE of 0.192 (18.6%) and all of the 50 models found an effect. Thus, having peers that smoke and/or drink appears to be an important mediator of maternal drinking's effect of AUDIT score.

Table 9-6: EIMOCs used for each non-parental mediation model

Posited mediator	Total EIMOCs	Variables
Early alcohol initiation	1	Parental permissiveness
Sensation seeking	1	Parental permissiveness
Depression	1	Maternal depression
Anti-social behaviour	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use
Alcohol attitudes	2	Parental permissiveness; sensation-seeking
Smoking	4	Parental permissiveness; sensation-seeking; attitudes to alcohol; peer drinking;
Substance use	4	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking;
Number of friends	2	Parental permissiveness; sensation-seeking
Peer smoking	5	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking;
Peer drinking	5	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer smoking;
Peer anti-social behaviour	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use
Peer substance use	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use
Relationships with teachers	6	Parental permissiveness; sensation-seeking; attitudes to alcohol; smoking; peer drinking; substance use

9.4 Discussion

Mediation of the effect of maternal drinking above CMO guidelines at adolescent age 12 on adolescent AUDIT score at age 16.5 was tested across 19 intermediates which were related to parenting, the parent-child relationship, intra-personal effects, and interpersonal effects. These intermediates were identified using the novel ESC-DAGs methodology to generate a conceptual I-DAG which was then converted into a data I-DAG using the ALSPAC data. Indirect effects were calculated as randomised interventional analogues of the natural indirect effect using mediational g-computation via the Stata package `gformula`. An indirect effect was detected for only seven intermediates; maternal smoking at age 12; parental punishment at age 15.5; attitudes to alcohol at age 14; smoking at age 14; substance use at age 14; peer smoking at age 13.5; and peer drinking at age 13.5. Of these, maternal smoking mediated only a small proportion of the TCE (3.8%), peer drinking and adolescent smoking mediated notably large proportions (17.4% and 18.6% respectively) and the remaining four had moderate effects (parental punishment -7.2%; attitude to alcohol 8.1%; adolescent substance use 13.2%; peer smoking 9.1%). This discussion compares these results to knowledge on the indirect effects of maternal drinking and the wider literature on adolescent alcohol harm. It closes in consideration of the methodological limitations.

9.4.1 Other research on the indirect effect

As noted in Chapter 5, few studies from the systematic reviews performed mediation analysis on the effect of parental drinking. Latendresse et al used an extension of the Baron and Kenny method mentioned in Chapter 3 to perform "multivariate path analysis" and found support for parental monitoring and parental discipline as mediators (Latendresse et al., 2008). Mares used structural equation modelling and found an indirect effect through alcohol-specific communication (Mares et al., 2011). Conversely, Van der Zwaluw et al used similar methods and found "unexpectedly" that parenting did not mediate the effect of parental problem drinking (Van der Zwaluw et al., 2008) and Pears et al found no indirect effect through adolescent inhibitions (Pears et al., 2007). Notably, a study by Mahedy et al that was not covered in the reviews but was published in 2018 in *Addiction*, found that parental monitoring and peer anti-social behaviour mediate the effect of parental drinking (Mahedy et al., 2018). This was notable as they also used the ALSPAC data to investigate the indirect effect of maternal drinking at adolescent age 12 through the same mediators but on adolescent AUDIT score at age 18 instead of 16.5. However, they used the Baron and Kenny method rather than a technique that can take account of EIMOCs. Thus, the different conclusions of this thesis – that parental monitoring and peer anti-social behaviour do not mediate the effect of maternal drinking - may be explained by the difference in mediation method and the timing of the outcome. Indeed, none

of the studies from the reviews investigated the indirect effect of parental drinking while accounting for EIMOCs - as far as the current author is aware, the analyses in this chapter are the first to do so. Thus, there has been little research conducted on the indirect effect of parental drinking on adolescent drinking, there is a notable degree of inconsistency across this small evidence base, and none of the research published so far adequately accounts for EIMOCs.

9.4.2 Interpretation

Figure 9-1a posits the causal relationships implied by the results in a simplified DAG and Figure 9-1b further simplifies the diagram into a single causal chain to help with interpretation. Both rely on the timing of the data to determine causation in a similar way to the data I-DAG. Note however that drinking is stated to cause smoking for both parents and peers, and that the chain attitudes → smoking → substance use is posited at age 14.

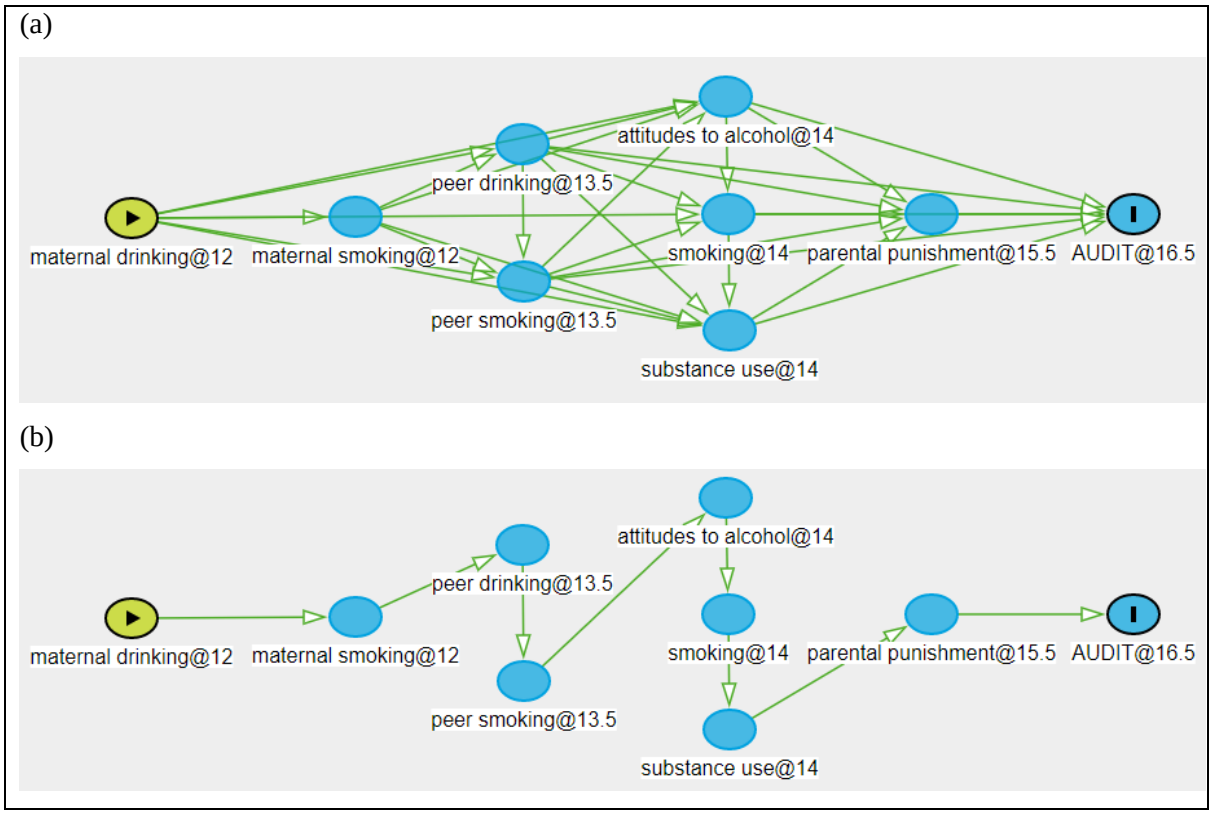


Figure 9-1: Causal-chain through mediators of maternal drinking's effect on AUDIT score

Figure 9-2 below, however, is a more cautious interpretation. It simply states seven mediators of the effect of maternal drinking. Arguably, this is a more accurate representation of the findings than Figure 9-1, given the lack of so-called multiple mediation models (i.e. models which estimate indirect effects through more than one mediator). For example, while Figure 9-1a above states that peer drinking influences adolescent attitudes to alcohol both with a direct effect and indirectly through peer

smoking, this was not tested in any of the mediation models. As such, for the purposes of this discussion, Figure 9-2 is a useful reference.

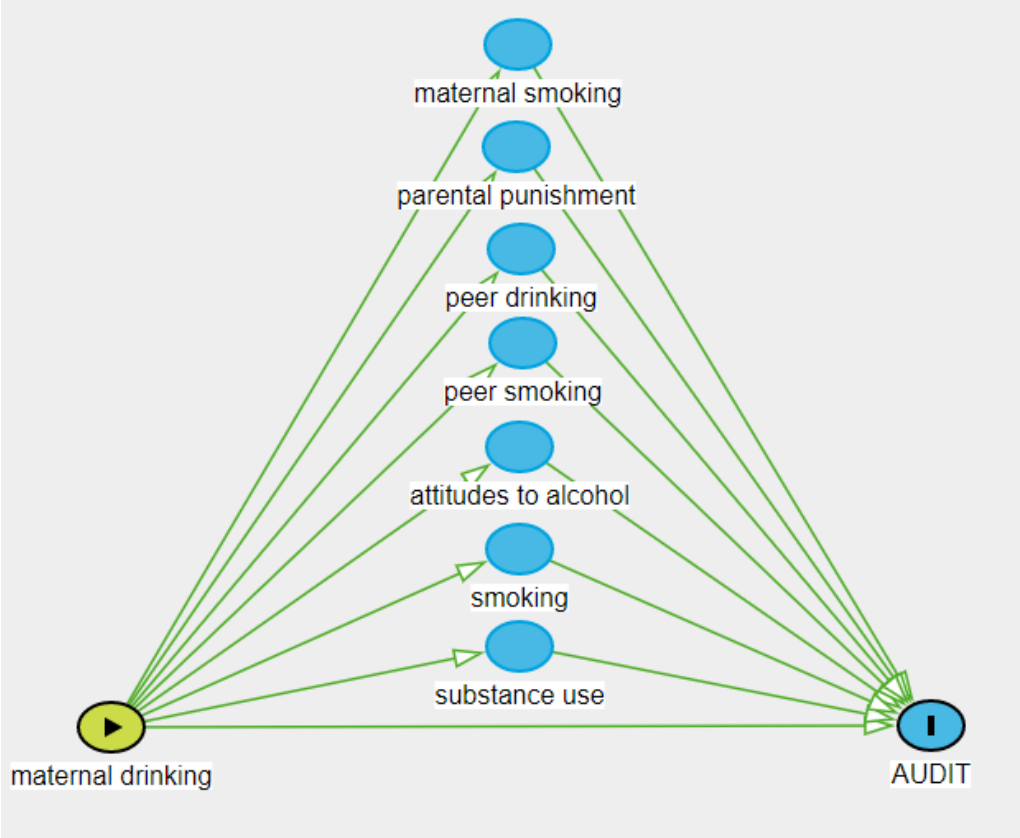


Figure 9-2: Mediators of maternal drinking's effect on adolescent AUDIT score

The most notable trend across the results was that substance-related variables formed the main group of mediators. This was across parental, intra-personal, and interpersonal dimensions. The only variable that was not related to substance use that had an indirect effect was parental punishment. On one hand this may appear somewhat tautological – adolescents who use a substance (alcohol) do so because they have positive attitudes towards this substance and because they use other substances. However, one of the strengths of these analyses is how they are innately social – these are indirect effects from one individual (mother) to another (child) through either of them or even a 3rd party (peers). Thus, these effects are not tautological. Rather, the variables that were mediators were those that had the highest conceptual overlap with the exposure and outcome, and that had higher ACEs on the outcome. This pattern persisted across the mediators themselves - the highest indirect effects were for the mediators that had arguably the strongest conceptual relationships and highest associations with AUDIT scores – adolescent smoking and peer alcohol use. Adolescent substance use was also relatively high. Peer smoking and adolescent attitudes to alcohol were somewhat lower. The parental effects were the lowest.

Parental influences

The relative lack of mediation for parental influences was unexpected. This was especially the case for the parenting variables (parental permissiveness towards adolescent alcohol use, parental monitoring, parental punishment). Firstly, while a high ACE for the mediator (M → Y) does not necessitate a high indirect effect, it does increase the chances that a smaller indirect effect would be detectable and the ACEs for these variables were the highest of the parental influences (parental permissiveness and parental monitoring were higher than maternal drinking itself at 1.67 and 1.41 respectively). Moreover, the evidence from the review of systematic reviews for parenting as a cause of adolescent alcohol harm was strong. Secondly, as alluded to above, a common hypothesis in research on parental influences on adolescent behaviour is that parental alcohol consumption is related to problematic parenting which in turn can result in adolescent alcohol harm (Latendresse et al., 2008, Mares et al., 2011, Pears et al., 2007, Van der Zwaluw et al., 2008, Mahedy et al., 2018). The results from this chapter seem to reject this hypothesis – parental permissiveness and parental monitoring (with and without an exposure-interaction term) were not mediators, and higher levels of parental punishment produced a negative NIE, indicating the opposite of this overall theory (i.e. rather than maternal drinking causing high levels of punishment which result in high levels of adolescent alcohol harm, maternal drinking resulted in lower levels of punishment which caused lower levels of alcohol harm).

No evidence was found for mediation via the parent-child relationship across the dimensions of parent-child tension, parental involvement, or secret-keeping. Despite how there are similar expectations for the indirect effect of parent-child relationship as parenting, this was less surprising. The review of systematic reviews found weaker evidence for their effect and their ACEs were small - indicating that the directed edge M → Y may have been more tenuous than for the parenting behaviours. This was also the case for maternal depression, which was not treated as an exposure in the review of systematic reviews, and which had a smaller ACE.

Maternal smoking was the only other mediator related to parents. This was in line with expectations given the well-established literature on the relationship between smoking and drinking. Particularly in how smoking and drinking are viewed as co-occurring or mutually reinforcing or otherwise existing in a positive feedback loop (Weitzman and Chen, 2005, VanderVeen et al., 2013, Nichter et al., 2010, Room, 2004). Or, as per Nichter et al, "it's a package deal" (Nichter et al., 2010). In other words, they both confound and mediate each other's effect over time. However, the size of the indirect effect was small at 0.041 (3.8% of TCE), especially compared to adolescent smoking (0.174) and peer smoking (0.092).

Intra-personal influences

The idea that having a mother who drinks more leads to positive alcohol attitudes that further lead to higher AUDIT scores is not controversial. There are numerous concepts, theories and perspectives that would anticipate how adolescent attitudes to alcohol might mediate the effect of maternal drinking. For example, a common concept employed in studies of parental effects on adolescents is parental modelling, in which children observe how parents behave and model their attitudes and behaviours on these observations. Indeed, parental modelling was the terminology used to cover parental drinking in the Ryan et al review from the Chapter 5 (Ryan et al., 2010). Similarly, under the assumption that previous behaviour predicts later behaviour, and that smoking and substance use are strongly related to alcohol use, the indirect effects through smoking and substance use were also not surprising. However, while smoking and drinking can be mutually reinforcing, a common theory is that illicit substances can displace either or both as 'drug of choice'. For example, the concern that more expensive alcohol could cause drinkers to displace alcohol with illicit substances was a key issue during the design of minimum unit pricing in Scotland (Katikireddi et al., 2019). However, the relative youth of the ALSPAC participants in these analyses may explain why substance use at age 14 mediates the harmful effect of maternal drinking (with harmful indirect effect), rather than displacing alcohol use at age 16.5 (i.e. a protective indirect effect).

The intra-personal effects that were not mediators included mental health, anti-social behaviour, and sensation seeking. Sensation seeking had a very high ACE. The lack of mediation can be interpreted as personality at age 13.5 not being influenced by maternal drinking at age 12. It may be the case that this aspect of personality is largely determined earlier in the life-course such that by age 13.5 it is not influenced by maternal drinking at age 12. Neither mental health nor anti-social behaviour were mediators, which was unexpected given they are both commonly investigated exposures. However, neither had a high ACE.

Interpersonal influences

The indirect effects through peers are not intuitive as they suggest that some aspect of maternal drinking causes another aspect of peer smoking and drinking. In other words, they suggest that parents have a direct effect on peers via the directed edge X → M. Unpacking this direct effect from parents to peers makes it clear that a parent's child is likely to be the main channel of their influence on the child's peers – there is an indirect effect through the adolescent on the path parents → adolescent → peers. It may also be the case that collider bias inflated these results if the assumption is made that adolescents integrate with peers who are similar to them, and that this effect is not fully captured by the confounders used in the analyses. Similar to the ACEs in the previous chapter, the no-interference assumption for valid causal inference may be breached by how the above analyses fail to account for clustering on school or neighbourhood levels. Finally, the peer influences which had

indirect effects controlled for fewer EIMOCs, meaning that unmeasured intermediate confounding could inflate the results. Another interpretation, however, is that peers mediate the effect simply because the context of the peer group is where the adolescent is likely to drink and smoke most. In other words, it may be the case that peer drinking acts as something of a proxy for previous adolescent drinking.

The lack of an indirect effect for anti-social behaviour was unexpected. While the variable did not measure drinking or smoking, explaining its discrepancy from peer drinking and smoking, it did have a similarly high ACE of 1.52, and some definitions of anti-social behaviour may include underage drinking and smoking (Alati et al., 2014). However, it was notable that peer anti-social behaviour was close to qualifying for the wave 2 analysis (its average confidence intervals did not cross the null, but were very close to zero (-.003) and 10 of the 20 models found no effect). This lack of effect may be explained by how peer drinking was used as an EIMOC. In other words, peer drinking may be the aspect of peer anti-social behaviour which would explain an indirect effect of the latter. This may explain why Mahedy et al found an indirect effect – they were not able to control for intermediate confounding by peer drinking.

9.4.3 Limitations

The limitations that applied in the previous chapter applied here, including the restriction to binary explanatory variables, the timing of the data, unmeasured confounding, and concerns around the assumptions for valid causal inference. The restriction to maternal drinking as a binary measure based on CMO guidelines created distinct limitations for mediation analysis, however. The CMO guidelines are for low risk drinking – the most harmful drinkers and drinkers who consume only one unit more than 'safe drinkers' are treated the same. Thus, the effect of more harmful or dependent maternal drinking on both the outcome *and* the mediators may be obfuscated, perhaps masking an indirect effect. It may be the case that variables which did not mediate under the CMO guidelines threshold may mediate under a different recode, for example a threshold of 21 units per week. Note that this would not solve the problem with the consistency assumption for valid causal inference – 21 units per week could still represent one instance of heavy episodic drinking or one large glass of wine per night (i.e. a large glass of wine is 250 ml and would have between 2.8 and 3.5 units for ABV 11-14%).

The timing of measures was also problematic in a way that was similar to the IPW analyses, in that some indirect effects were estimated while controlling for six EIMOCs, while others were estimated while controlling for only one. The small indirect effect of maternal smoking, for example, may have

attenuated to null if more EIMOCs (such as parenting measures) were used. Similarly, the high effect of peer drinking was detected while only controlling for three EIMOCs. As such, the additional assumption for valid causal inference for mediation analysis (no uncontrolled EIMOCs) was satisfied to different degrees depending on the mediator.

Measurement error and resulting residual confounding are as much a concern for mediation analysis as for analysis of total effects – better measured meditators will have larger indirect effects. Of course, the success of a measure at capturing the nature of the underlying concept is always a concern for any form of analysis. For example, the degree to which secret-keeping matches the underlying concept of parent-child communication, and the degree of participant bias in completing the question (e.g. recall bias, audience bias), will influence results and interpretation. A second concern, which could have influenced the mediation results, and which must be noted given the emphasis on dichotomous variables in this chapter, is that the transformations used (i.e. cut-off points for binary variables) can potentially inflate or suppress results. For example, the small indirect effect of maternal smoking is dichotomised around whether or not individuals are smokers, but the mediator used was a recode of how many cigarettes were smoked on the average day. It is possible that a larger indirect effect could have been found if the variable had been coded such that more extreme participants (e.g. greater than 10 cigarettes per day) acted as the exposure group. Given greater time resource (e.g. fewer restrictions due to data imputation), recoding mediators could be a valid avenue for sensitivity analysis.

Relatedly, there was a small degree of mismatch between the EIMOCs that were used and the mediators. This was largely due to how few mediators were detected. Firstly, using maternal smoking as an EIMOC was potentially problematic as there is no way to know if its indirect effect in fact represented confounding of the exposure-outcome relationship, due to the cross-sectional measurement of maternal drinking and the nature of the relationship between smoking and drinking. As noted above, this was why its use as an EIMOC was restricted. Indeed, it had by far the smallest indirect effect even with controlling for only one EIMOC (maternal depression). Secondly, parental punishment was not used as an EIMOC because its measurement was comparatively late. However, each of the five remaining mediators were used as EIMOCs (attitude to alcohol, smoking, substance use, peer drinking, peer smoking) although peer smoking was only used when modelling peer drinking as the mediator.

9.5 Conclusion

This chapter concluded the data analysis plan by taking the various parental and non-parental influences that appeared to be causal in previous analyses of adolescent AUDIT score and testing them as mediators of the effect of maternal drinking. MRO 5 was met by using mediational g-computation in tandem with the data I-DAG to estimate the indirect effects. A substantial portion of the effect of maternal drinking on adolescent AUDIT score was explained by the substance attitudes and behaviours of the adolescent and their peers, while only a small proportion appears to be mediated by parental influences. Thus, SRO 5 was met. These findings contrasted with expectations from the literature, although these expectations are based on very little empirical research which in turn relied on methods which do not account for EIMOCs. Thus, the analyses in this chapter address a clear gap in understandings of adolescent alcohol harm.

Chapter 10 Discussion

This Discussion is split into three main parts. Part one briefly considers how each research objective was met and thereby sets the context for the rest of the chapter. Part two expands on the methodological research objectives by reviewing the strengths, limitations and future developments of ESC-DAGs. Part three focuses on the results of the substantive analysis, with emphasis on strengths and limitations, future research, and wider implications.

10.1 Research objectives

The five methodological and five substantive research objectives were met in Chapters 2 through 9. A foundation in causal inference was established in the opening chapters (MRO 1) before ESC-DAGs was introduced on the basis of this foundation (MRO 2). A systematic review of reviews then identified key parental influences of adolescent alcohol harm and critically evaluated the literature (SRO 1, SRO 2). ESC-DAGs was then applied to this information to build the conceptual and data I-DAGs (MRO 3, SRO 3). Using the data I-DAG and ALSPAC data, average causal effects (ACEs) were estimated for numerous parental and non-parental influences (MRO 4, MRO 5, SRO 4) before causal mediation analysis was performed on the effect of maternal drinking on adolescent alcohol harm (MRO 5, SRO 5). How each research objective was met is discussed here with sign-posting to pertaining Chapters and sections.

MRO 1 – Establish a foundation in causal methods for this thesis

Chapter 1 identified the need for a method for systematically designing DAGs. This drove the overarching aim for this thesis - design and apply a method for building DAGs, and use it to direct causal analysis. Thus, a foundation in modern causal inference was necessary and was stipulated as MRO 1. This research objective was met across Chapters 2 and 3. Chapter 2 introduced

counterfactual reasoning (page 13), a working definition of cause borrowed from Pearl (page 17), and DAG functions (page 18). Chapter 3 built upon this by introducing formal notation from the potential outcomes framework (page 30); IPW as a statistical approach to estimating ACEs (page 34); causal mediation analysis as an approach to estimating natural in/direct effects (page 38); and an expanded definition of cause (page 39). Thus the 'foundation in causal inference' reduces to: 1) counterfactual reasoning; 2) corresponding statistical methods; and 3) DAG functions. This foundation was sufficient to achieve the rest of the research objectives.

MRO 2 - Develop a method for building DAGs

This research objective represented the foremost novel contribution of this thesis. It was achieved in Chapter 4 with the ESC-DAGs method, which was developed in reference to the DAG functions from Chapter 2 and several other principles of DAGs introduced in Chapter 4 (page 51). The method passed peer review with the *International Journal of Epidemiology* (Ferguson et al., 2019). This was part of the rationale for concluding that MRO 2 was met with ESC-DAGs. The Methodological discussion below expands on the strengths and limitations of ESC-DAGs in some detail.

MRO 3 – Apply and demonstrate the method

ESC-DAGs was applied and demonstrated in detail rather than systematically compared to other approaches, largely due to the lack of any such approach. This was mainly achieved in Chapter 6 which applied ESC-DAGs to the literature on parental influences on adolescent alcohol harm to produce a conceptual I-DAG, and Chapter 7 which used the ALSPAC data to convert the conceptual I-DAG into a data I-DAG. Chapters 8 and 9 also demonstrated how data I-DAGs can be used to direct analysis.

MRO 4 – Demonstrate how the method can be used to direct conventional data analysis

Chapter 1 argued that limiting DAG usage to only POF analyses could prohibit wider health and social science from benefiting from the strengths of DAGs (page 10). As such, MRO 4 was interested in showing that DAGs can be useful tools to other forms of analysis. This research objective was met in Chapter 8 by demonstrating how DAGs can be used to direct multiple regression analysis (page 158). It also demonstrated how DAGs can improve over other model-building techniques by comparing different adjustment strategies.

MRO 5 – Use the method to direct data analysis under the potential outcomes framework

If the method had not produced DAGs that could assist in POF analyses, then it arguably would not be fit for purpose. In order to assess this, notable POF methods were applied using the data I-DAG. In Chapter 8 inverse probability weighting (IPW) was used to estimate ACEs on AUDIT score at age 16.5 of parental influences, intra-personal, interpersonal, and institutional influences (page 172). In

Chapter 9 causal mediation analysis, in the form of mediational g-computation, was used to analyse the indirect effect of maternal drinking (binary measure of whether mothers drink above or below CMO guidelines) on adolescent alcohol harm (AUDIT score at age 16.5). While it is not possible to compare to how these analyses would have progressed in the absence of ESC-DAGs, the data I-DAG was valuable for informing how covariates should be allocated across a high volume of models (e.g. confounders to be controlled, intermediates to not be controlled, etc.).

SRO 1 - Determine the main parental influences on adolescent alcohol harm from the literature

Parental influences on adolescent alcohol harm were the health context chosen for the ESC-DAGs application and subsequent analyses. However, the corresponding review had to be compatible with ESC-DAGs, as well as being a critical summary of the literature. A review of systematic reviews was identified in Chapter 5 as a suitable method (page 79). A total of 12 categories of parental influences were identified across the systematic reviews: parental drinking; provision of alcohol; favourable alcohol attitudes; family conflict; parent-child relationship quality; parental involvement; parental support; parental monitoring; alcohol-specific rules; parental discipline; parent-child communication; alcohol specific communication (page 92).

SRO 2 – Appraise the evidence on parental influences on adolescent alcohol harm

The scope for causal inference was given special attention due to the explicit causal impetus of this thesis. The Bradford Hill criteria and other principles of causal inference and evidence synthesis were employed to evaluate the ability to infer harmful or protective effects for each of the 12 parental influences (page 90). Parental drinking, low parental monitoring, and parental supply of alcohol appeared to be important harmful determinants across the literature. The evidence for other parental influences, such as the parent-child relationship and parent-child communication, was less clear. The review of systematic reviews also identified a gap in the literature on parental influences on adolescent alcohol harm in that there was a dearth of POF analysis. Thus, SRO 2 was met by critically engaging with the most current meta-literature.

SRO 3 – Build DAGs of parental influences on adolescent alcohol harm

The conceptual I-DAG from Chapter 6 (page 110) and the data I-DAG from Chapter 7 (page 146) met this research objective. The conceptual I-DAG had 52 concepts categorised as 15 confounders, 14 exposures, and 22 intermediates. The data I-DAG used the timing of the data in ALSPAC and the ESC-DAGs translation processes to decide on the relationships between the confounders, exposures, and intermediates. A total of 42 variables were included and inter-related in the data I-DAG (15 confounders, 9 exposures, 17 intermediates). As far as the author is aware, the data I-DAG is the only systematic and complex DAG in alcohol studies.

SRO 4 – Test the causal effects of parental influences and their intermediates on adolescent alcohol harm

The analysis plan for this thesis was to identify parental causes and non-parental intermediates of adolescent alcohol harm and then to investigate them holistically using causal mediation analyses. As such SRO 4 was concerned with the process of identifying parental influences of adolescent alcohol harm, and other non-parental intermediates causes. IPW models were used in Chapter 8 to estimate ACEs for this purpose. All of the parental influences were found to have an effect, as did 13 of the 17 intermediates. Maternal drinking, parental permissiveness towards adolescent alcohol use, and parental monitoring had relatively high effects compared to the rest of the parental influences (page 172). This was in keeping with the review of systematic reviews. However, several intermediates had higher ACEs, especially peer drinking and smoking, and sensation-seeking personality (page 176). Thus SRO 4 was met by using the data I-DAG to determine which variables should be used in the propensity score (and thus IPW) when estimating ACEs.

SRO 5 – Investigate mediators of parental influences on adolescent alcohol harm

Maternal drinking was selected in Chapter 7 as the exposure for causal mediation analysis. Of the 21 potential mediators, eight were other parental influences, seven were intra-personal effects, and six were interpersonal. Using mediational g-computation, only seven of these were identified as mediators. These were mainly substance related effects pertaining to the adolescent and their peers (page 197). Notably, parenting (measured as parental monitoring, parental discipline, parental permissiveness towards adolescent alcohol, and parental involvement) did not appear to be an important mediator of maternal drinking (page 194). The data I-DAG informed causal mediation analysis of the effect of maternal drinking on adolescent alcohol use by identifying EIMOCs for each model. These estimates, as well as those pertaining to SRO 4, are methodologically unique in alcohol studies.

10.2 Discussion on ESC-DAGs

The decision to design ESC-DAGs was motivated by the lack of methods for building DAGs. Authors have called for guidelines (Tennant et al., 2017) or a principled approach (Sauer et al., 2013). Problems with DAG usage have included over-simplification; a lack of transparency on the overall decision-making behind selecting directed edges; and variation in practice. ESC-DAG lays out a step-by-step evidence synthesis protocol which focuses on bringing background knowledge to bear in a systematic way, and which was able to direct a variety of analysis in this thesis. As noted above the only approach to building DAGs that was applied in this thesis was ESC-DAGs. As such, there is no

way to comparatively evaluate how successful it is. However, this is a limitation of the literature, rather than the method itself. Neither does it prohibit an evaluative discussion. As such, this section begins with a brief summary of the rapid ESC-DAGs application in this thesis before considering the main strengths of the approach and then closing with a detailed discussion of limitations and future developments.

10.2.1 Summary of rapid ESC-DAGs application

A total of 50 studies were included in the rapid ESC-DAGs review. Of these, 47 were identified from the review of systematic reviews and three from forward citations. Inclusion criteria were that the study was published after 2005 and that the context had similar legal age of purchase of alcohol to the UK. Most studies were Western European, with the Netherlands making the largest contribution (18 of the 50 studies). A DAG was produced for each of these studies by using the Mapping and Translation processes. Mapping involved replicating the narrative conclusions of each study using graph theory. All 'control variables' were mapped as confounders according to the potential outcomes framework definition of a mutual cause of the exposure and the outcome. The final step in Mapping is to saturate the IG such that all concepts are connected directly to all others. The Translation process was then applied to directed edge in each saturated IG to produce a DAG. The causal criteria were temporality, plausibility, and recourse to theory. If a directed edge possessed at least the first two criteria, then a counterfactual thought experiment was conducted. Directed edges could be indexed as bi-directional if necessary. The directed edges that involved any parental influence and/or adolescent alcohol harm were then added to a new graph to produce a conceptual I-DAG. This differentiated parental influences of adolescent alcohol harm from their confounders and intermediates on a conceptual level. Next the concepts were replaced with ALSPAC variables. The timing of the data was then used to organise the variables in the graph in that all earlier variables were posited as causing all later variables. Finally, cross-sectional groupings were saturated and translated. This produced the data I-DAG that was used to guide analysis. Recombination of concepts/variables was performed throughout.

10.2.2 ESC-DAGs strengths

The key strengths of ESC-DAGs included how it was able to direct data analysis in an immediate and clear fashion; the relative comprehensiveness of the DAGs it produced; the overall transparency of the method; the ability to comment on modelling practices in the literature; and the modularity of the approach. The most important of these was how the data I-DAG was able to inform the modelling

processes in a very direct way. For example, it clearly identified which variables should be included as covariates in regression models; which should be included in propensity scores when using IPW to estimate ACEs; and it differentiated between which variables had to be simulated as EIMOCs when estimating natural indirect effects (NIEs) and which variables should be modelled as baseline confounders. This is a crucial strength given that the fundamental purpose for using DAGs is to inform adjustment strategies – indeed it is the clearest indication that the method is fit for purpose.

In terms of the coverage of concepts included and the degree of causal modelling, the data I-DAG was substantially more wide-ranging and holistic than any modelling framework used in any of the 50 primary studies used to build it. Under the assumption that correctly assigning and modelling a high volume of covariates reduces bias (including uncontrolled confounder bias simply by identifying variables that might otherwise have been omitted and not inducing further bias by controlling for either mediators or colliders), this further supports the position that ESC-DAGs satisfies the rationale for using DAGs. It also indicates that ESC-DAGs can be used to help overcome some of the common limitations of DAG usage – for example, Tennant et al. argued that DAGs are often oversimplified in applied research whereas DAGs produced from ESC-DAGs are relatively comprehensive (Tennant et al., 2017).

Even the most complex DAG possesses a degree of transparency insofar as each directed edge represents a clear causal hypothesis. However, the data I-DAG was complex enough that its intelligibility was somewhat inhibited, which could subsequently obfuscate the underpinning decision-making on the directed edges. This would be a severe limitation of any DAG building method, as a core advantage of DAGs is how they are explicit about the hypotheses encoded in the conceptual and statistical models. The approach taken in ESC-DAGs to overcome this problem was to leverage the record-keeping aspects of the process as tools to improve transparency. This included the Translation process (and decision log), the directed edge index, and the DAG that was produced for each of the reviewed studies. Indeed, a crucial part of the rationale behind the counterfactual thought experiment was to explicate the reviewer's reasoning on the directed edge in question. This emphasis on transparency is in keeping with DAG principles, making it a strength of the method. However, this highlights an innate tension between the desire to use DAGs to create efficient representations of causal phenomena, and the underlying complexity of these phenomena, especially in observational health and social science research. In other words, if the intention is to use a DAG to guide analysis such as that in this thesis, then a high degree of complexity is likely inevitable and steps such as those included in ESC-DAGs are necessary.

The overall approach of 'translating' research evidence into DAGs and combining these individual graphs has notable advantages. Firstly, and most importantly, it directly utilises the causal hypotheses

in the evidence base to inform the directed edges in the graphs. Positing and translating each directed edge also ensures critical appraisal of these hypotheses. This is the form that 'evidence synthesis' takes in the ESC-DAGs process. Secondly, this overall approach can detect undercontrol for confounders and overcontrol for mediators in individual studies. While ESC-DAGs is primarily a method for building DAGs, its ability to explicate modelling practices could be fruitfully leveraged. For example, it could help to systematise how certainty in specific directed edges is quantified or otherwise characterised for better informing sensitivity analysis using a data I-DAG. Alternatively, it could be used to systematically comment on modelling practice and thus certain forms of bias across a literature base.

The modularity of the approach was briefly discussed in Chapter 4. It allows other researchers to swap out stages for alternatives or to extend the method without the need for systematic reconfiguration. While this aspect of the design has not received much attention in this thesis, it was nevertheless an important consideration. Interested researchers can engage with ESC-DAGs without being beholden to aspects they consider to be weak. Thus the modularity may help encourage further development of ESC-DAGs or even position the method as a spring-board for competitor methods which might implement only part of the approach. For example, the Translation process currently emphasises a balance between rigour and speed. Other approaches might lean more towards rigour or speed - the counterfactual thought experiment could be dropped to emphasise speed or the Bradford Hill criteria as a whole could be applied beside the counterfactual thought experiment to emphasise rigour. Modularity is thus a notable strength of ESC-DAGs. So much so that it underpins the following discussion on limitations.

10.2.3 ESC-DAGs limitations and extensions

ESC-DAGs was, of course, not designed to become the authoritative method for building DAGs. The publication, for example, concluded by saying that the main hope for the method is that it will stimulate further debate on how researchers can use DAGs. There are several dimensions of the method which are clear candidates for this kind of attention: its efficiency; its reliance on theory and data rather than evidence; its reproducibility; and how it deals with uncertainty. Given the modular design, the following considers each as a key limitation and thus prospect for extension or alteration.

10.2.3.1 Efficiency

There is no way to guarantee or confirm that the DAGs produced from ESC-DAGs correspond to the real causal relationships under study. While this may appear a platitude that would apply to any

method for building DAGs, it is consequential for efficiency. Specifically, it is for this reason that ESC-DAGs adopted the conservative approach to deleting directed edges – that directed edges should only be deleted when researchers are certain. Indeed, as covered below, understanding how to employ the concept of certainty in this respect is challenging. Nevertheless, the main practical consequence of this is that, because it necessitates a higher volume of directed edges in the DAGs, more time must be spent applying the Translation processes. Despite how they were argued above to be among the strengths of the approach because of their ability to balance rigour and speed, the Translation processes are still time-consuming. Another consequence of taking a conservative approach to deleting directed edges was noted above in that the data I-DAG is likely to be very dense in terms of directed edges.

While the conservative approach to decision-making is likely to result in more directed edges than might be necessary, rapid ESC-DAGs also generated directed edges that were redundant regarding the data. Specifically, the process of producing a DAG for each study involved directed edges that were not possible with the ALSPAC data. This is not necessarily wasteful - it allows for a level of insight into modelling practices in the literature and encourages careful thought on conceptual relationships that indirectly pertain to the focal relationship (e.g. exposure-outcome). However, it does lengthen the DAG-building process.

As such, rapid ESC-DAGs could be further refined to remove redundancies and to limit the number of directed edges that have to be assessed. Firstly, the conservative approach to decision-making could be rejected in favour of an alternative. For example, a directed edge could be deleted if reviewers felt that, while some degree of causation may exist, it is not an important cause. Of course, this creates another difficulty in conceptualising what is 'important'. A more practical option would be to focus on altering the Mapping and Translation process, for example by changing the goal from producing a DAG for each study to simply differentiating the covariates into confounders and mediators. This could involve not saturating an IG and thus not translating directed edges between the non-focal covariates. Even though this would not allow researchers to observe modelling practices from a coherent DAG, it would still assist in identifying overcontrol for mediators and other modelling mistakes. Another change could be to omit the counterfactual thought experiment for some or all directed edges, for example to never perform it on socio-demographic confounders. However, this would negate the benefit of outlining reviewer assumptions in a systematic way.

10.2.3.2 Dealing with certainty

An important limitation of the data I-DAG is that there was no systematic method for determining the certainty or confidence in any directed edge. This has at least three notable implications. Firstly, it

meant that, while the data I-DAG can inform reviewers of how to conduct sensitivity analysis around a particular relationship, it cannot inform on which relationships are the best candidates for sensitivity analysis. The only exception to this was that cross-sectional variables were treated as candidates for sensitivity analysis before longitudinal relationships. Secondly, despite the difficulty in working with a high volume of covariates for conventional analysis such as the multiple regression models in Chapter 8 (e.g. parental monitoring had more than 30 potential confounders), the I-DAG was not able to suggest which covariates might be prioritised. A systematic ranking of certainty or confidence across the confounders could augment data-driven techniques thus allowing for more informed decision-making in this regard. Third, directed edges derived from studies with high risk of bias were treated in the same way as directed edges from studies with lower risk of bias. Integrating a bias appraisal stage is thus one of the most obvious extensions to ESC-DAGs. Due to the plethora of extant methods available, this need not be a new innovation. However, it is notable that ESC-DAGs creates information which could be useful in this regard, for example it can be used to count how many times a specific confounder is treated as such (assuming that more common usage equates to more confidence in the relationship).

10.2.3.3 Reliance on theory, rather than evidence

An ESC-DAGs review such as the rapid application here can certainly be viewed as an evidence synthesis protocol insofar as the relationship between each exposure and the outcome is concerned. However, the relationships across the rest of the DAG are not informed by evidence in the same way. The confounding structure as it pertains directly to the exposure and the outcome tended to be informed by the conclusions of the studies under review, while the relationships between the confounders themselves were rarely discussed in detail and thus relied on reviewer background knowledge. In terms of the ESC-DAGs processes, this was most notable with the recourse to theory step of the Translation process. Each primary study would generally cover theory and evidence extensively regarding the focal relationship(s) under investigation, particularly in the background and discussion. However, this was rarely the case for other relationships. For example, for relationships between variables treated as confounders in each study, the recourse to theory step would often be blank (especially for relationships for which the author did not have background knowledge). Thus, the recourse to theory step, for some inter-confounder relationships, could devolve to a reliance on comparatively uninformed theorising. There are at least three ways to avoid this. The first is to conduct some form of scoping review for these relationships. However, this would involve an exponentiation of an already time-consuming task. A second option is to use a broader team of reviewers in an attempt to maximise the breadth of reviewer background knowledge in terms of coverage of the DAG. Indeed, this is the default recommendation for ESC-DAGs and most other evidence synthesis methods. Third, participatory action research and further stakeholder engagement

could be used to help make decisions on these relationships. Indeed, as alluded to in Chapter 2, there is already a precedent for this when building DAGs in the form of consultation with experts.

10.2.3.4 Reliance on data, rather than evidence

Rapid ESC-DAGs rejects the principle that DAGs should be built independently of data. This was mainly for practical purposes – building a DAG with numerous nodes independently of data, when that DAG is intended for one-time application to a single data set, results in a high volume of directed edges that are not possible to realise in the data. As expanded on above, assessing every directed edge makes the process inefficient. Conversely, however, there is a risk that data I-DAGs can rely so heavily on data sources that they reflect the survey waves as much as an 'independent' theory of causation. This can be viewed as a limitation because it inhibits the DAG's capacity to represent the reality of the causal relationships under study. However, as the purpose of DAGs is to direct data analysis, a data I-DAG can, of course, only represent reality insofar as the data allows.

A second concern was how rapid ESC-DAGs used the timing of data collection to decide on the direction of some directed edges. This was justified on two counts. The first was purely pragmatic – to minimise redundancy as above. The second justification however, was more problematic, as it was based on the assumption that what is measured after cannot influence what is measured before. According to basic causal theory, this is technically correct (e.g. the Bradford Hill criterion of temporality). However, at least insofar as time-varying concepts are concerned, rapid ESC-DAGs risks conflating the *measurement* of phenomena with their occurrence. For example, if socio-economic position was measured after maternal drinking, the resulting data I-DAG would not have recommended controlling for socio-economic position when estimating the effect of maternal drinking on adolescent alcohol harm. This was not problematic for this study, as baseline confounders were all measured before or cross-sectionally to the exposure. Indeed, this is likely to be the case for many longitudinal data sources. However, it was somewhat problematic for the relationships between the intermediates and parental influences. For example, parental punishment at age 15.5 controlled for peer drinking at age 13.5, while peer drinking did not control for parental punishment. This approach had at least two limitations. Firstly, there is theoretical support for the effect of parental punishment on peer drinking via an indirect effect through the adolescent. Secondly, there is an argument to be made for controlling for parental punishment at age 15.5 while estimating the effect of peer drinking at age 13.5, in that parental punishment at age 13.5 would be expected to predict parental punishment at age 15.5. In other words, while parental punishment is time varying, the capacity of past behaviour to predict future behaviour is still very strong. As such, parental punishment at age 15.5 could be used as a proxy of parental punishment at age 13.5, and thus modelled as a confounder when estimating the effect of peer drinking. However, this argument was

not implemented into rapid ESC-DAGs for the simple reason that controlling for downstream measures can remove part of the indirect effect of the exposure (e.g. peer drinking). Indeed, accepting the bias that such a strategy might induce would contradict the emphasis on using DAGs to avoid overcontrol for mediators, which was also a central focus of the analyses in this thesis. Thus the timing of data collection was used to decide on the directionality of relationships to reduce redundancies in the Translation and Synthesis processes and to avoid overcontrolling for mediators, despite the risk of conflating measurement with the occurrence of the underlying construct. Note, however, that this does not preclude sensitivity analysis based on using future measures as proxies for missing past measures, although this might need to be decided on a case-by-case basis.

10.2.3.5 Reproducibility

The extent to which the data I-DAG might have been reproducible is unclear. This mainly pertains to the recombination of concepts and the decision-making during the Translation process. The recombination process (to recombine nodes that were deemed conceptually similar or that had similar profiles compared to the rest of the DAG) was the least systematic of ESC-DAGs processes. This was because the process of identifying and acting on conceptual similarity is inherently subjective. This is not to say that no method or tool could be developed for assisting in this task. However, such a task was beyond the scope of ESC-DAGs. One consequence of this is that the concepts in the conceptual I-DAG might have been different if a different reviewer had conducted the process. Similarly, the directed edges themselves may have been different. This is another reason for future research conducting ESC-DAGs with a team, particularly one which employs some mechanism to resolve inter-rater disagreement. For example, it is common in systemic reviews to use either an odd number of reviewers or for an additional reviewer to be employed specifically for resolving disagreements. Such a method would be appropriate for ESC-DAGs.

10.2.3.6 Reliance on evidence

Given its reliance on using existing research evidence for building DAGs, there is a risk that ESC-DAGs applications may perpetuate oversights or other mistakes from the pertaining literature. This is further exacerbated by time constraints and the resulting scope (e.g. in terms of how their confounders were identified from the literature, the mediators in this thesis were approached as a group rather than each being uniquely assessed). Consider, for example, any given confounder that is not present in the studies selected for ESC-DAGs, perhaps because it is difficult to measure (e.g. child abuse) or is not often treated as a confounder compared to more common examples, such as sex or ethnicity. The only way in which that confounder might be present in the final I-DAGs would be if the research team were to add it outside the ESC-DAGS protocol. There are numerous processes through which this

might happen, such as discussions among the research team; collaboration with external experts; as part of the peer review process; and more. This thesis has not been prescriptive on this count, other than to say that such variables can be added by any means. While clearly very difficult to define, a valuable extension to ESC-DAGs would be to introduce such processes more formally. For example, the recommendation to use at least 2 reviewers with 1 additional to resolve divergent decisions could be extended to suggest additional variables. In cases were researchers are not experts on particular relationships, external experts should be consulted. However, defining 'consultation' in a reproducible or otherwise systematic sense is challenging. Thus, while a concern beyond the scope of this research given the challenge involved, adding processes to formalise the identification of confounders which may have been missed in the literature is clear limitation and valid avenue for further developing for ESC-DAGs.

10.3 Discussion on analysis & results

This section pertains to the substantive research objectives in general, but with particular focus on the ACEs estimated in Chapter 8 and the NIEs from Chapter 9. The main findings were as follows.

- So-called average causal effects (ACEs) were found for each of the parental influences identified from the systematic review of reviews and ESC-DAGs review.
- The largest ACEs were for maternal drinking, parental monitoring, and parental permissiveness towards adolescent alcohol use. This was in keeping with expectations set by the review of systematic reviews
- However, peer and intra-personal influences, identified by the ESC-DAGs review, had larger effects, especially peer substance behaviours and the sensation-seeking personality of the adolescent.
- Only seven mediators of the effect of maternal drinking were detected, these were mainly substance-related factors pertaining to peers or the adolescent.
- Parenting did have an indirect effect, but it was small and only through parental punishment.

10.3.1 Implications of findings

As indicated by the research objectives, this thesis was primarily methodological. Generating recommendations for policy was never a focus, and as such is generally beyond the scope of this discussion. Nevertheless, this section does briefly consider wider implications of the main findings. Firstly, in general the ACEs triangulate with what is already known on parental influences, other substance behaviours, and peer effects. A common consideration across a research community is

triangulation of evidence. Prominent epidemiologists working in causal inference methods, Marcus Munafò and George Davey Smith argued in *Nature* in 2018 that, in order to be robust, research must not only be reproducible, but be based on "many lines of evidence" (Munafò and Davey Smith, 2018). For any research question then this means the research community must use a variety of methods on a variety of data. This thesis addressed the dearth of POF analysis of widely studied risk factors, and, because its findings converge with the wider literature, they can be argued to contribute to the robustness of this evidence base – especially the evidence base of parental influences on adolescent alcohol harm.

The ACEs in this study refer to average differences in adolescent AUDIT scores, with increasing scores indicating higher risk of adolescent alcohol harm. Risk refers to a wide variety of 'harms' from alcohol use such as alcohol poisoning, hangovers, anti-social behaviour, risky sexual behaviour, injurious accidents, and drunkenness more generally. As alluded in the Introduction, alcohol-related harm in any phase of the life course can be characteristic of that phase. Consequences that are more characteristic of adolescence include the formation of life-long habits and impaired neurological development, and others have clearer correlates with adulthood, even if they are arguably perceived differently (e.g. so-called 'teen pregnancy' and poor performance in school as a correlate of employment issues). Conceptions of alcohol harm in adolescence are not generally concerned with chronic alcohol harm (e.g. liver cirrhosis, cancer), due to the relative youth of the population.

Some individual ACEs are worth further discussion. Firstly, the finding that low parental monitoring has a harmful effect should be given careful attention due to its compatibility with normative parent-blaming narratives (Hansen and Ainsworth, 2007). It may instead reflect greater barriers to higher levels of parental monitoring. Indeed, this may also be the case for the parent-child relationship or other parenting aspects, such as parental involvement or parental punishment. Such barriers may reflect employment (e.g. long or unsociable work hours), family structure (e.g. single parenthood or other siblings requiring more attention) and the mental or physical health of the parents, including their own alcohol and substance use. Rather than simply concluding that parents are to blame, there are numerous interventions and policies which could be relevant considerations, such as flexible working hours, or more accessible childcare.

As mentioned in Chapter 8, the high ACE for parental permissiveness towards adolescent alcohol use was in keeping with a recent *Lancet* article by Mattick et al that found parental supply of alcohol was an important risk factor for adolescent alcohol use (Mattick et al., 2018). Thus, the ACE comments on the debate on whether parental supply of alcohol is a valid harm reduction approach. In both this thesis and the Mattick paper, the findings suggest the opposite – parental supply of alcohol increases harm. The Mattick paper recommends zero-tolerance policies instead, and the ACE supports that

recommendation. However, it must be noted that parental permissiveness towards adolescent alcohol use is not the same as parental supply, and that the measure used had limitations. Firstly, the binary response meant that those answering 'yes' to the question of whether they had tried alcohol with their parents' permission, could vary from adolescents who were once allowed a sip of their parent's drink at a special occasion, to adolescents who have been allowed whole drinks regularly or even been supplied alcohol to consume in an unsupervised setting. Thus, a large degree of variation may be obfuscated. This was compounded by a second limitation - there was very low variation in the measure in that around 93% of respondents had tried alcohol with their parents' permission. This has several implications. First, and unsurprisingly given the positivity assumption that there must always be a positive probability that participants can be exposed or unexposed for each stratum of the confounder set, this was the variable which required the most attention in terms of finding a balanced propensity score. Moreover, this low variation may reflect a further substantive difference between the exposed and unexposed – the unexposed may be adolescents living in teetotal households. Thus, even given how a balanced propensity score was used, it may have been the case that the ACE may be inflated by using an inappropriate comparator (i.e. it may not compare parental permissiveness so much as the effect of teetotal vs. non-teetotal parents). This was an underlying concern with all analyses in this thesis, but was most pertinent to this variable, given its low variation and the relative rarity of teetotal households. Nevertheless, the corresponding ACE still triangulates with the Mattick et al findings.

The non-parental ACEs were more difficult to interpret simply as they were not the focus of the review of systematic reviews. However, it was notable that peer effects had higher ACEs than the parental influences. This is an important consideration when weighing resource allocation for interventions, and it begs numerous questions – are peer exposures more prevalent than parental exposures, are there interventions that work for either, to what extent, and at what cost? Again, these considerations are largely beyond the scope of this thesis but, the following paragraph briefly considers the evidence on the efficacy of relevant interventions.

The evidence on parental interventions is slightly mixed and peer interventions are relatively uncommon, possibly due to the increased ethical requirements of working with children. Perhaps unsurprisingly, a scoping search of interventions on parental drinking as a cause for offspring drinking failed to return any results. Parental interventions instead tend to focus on parenting specifically or take a more holistic approach to parenting, parental behaviour, and the parent-child relationship together. A meta-analysis of RCTs published in 2008 by Smit et al looked at the effect of interventions focusing on either the parent themselves or the parent and child together, and found that all studies (nine) saw reduced initiation and frequency of alcohol use before age 16 (Smit et al., 2008). A systematic review published the year before by Petrie et al on the specific effect of parenting

programmes on alcohol use in under 18s found that only six of 14 studies reduced alcohol use (Petrie et al., 2007). Indeed, two of the primary studies used to build the DAGs focused on the same intervention (Örebro Prevention Program, Sweden), but produced conflicting results. Koutakis et al concluded that "working via parents proved to be an effective way to reduce underage drinking" (Koutakis et al., 2008), but Bodin and Strandberg, conversely found that the programme "does not appear to reduce or delay youth drunkenness" (Bodin and Strandberg, 2011).

In terms of interventions focused on peers, in 2016 MacArthur et al published a systematic review on the effect of peer-led interventions on alcohol use in young people aged 11-21, and while they found that all of the studies that were included reduced alcohol use, there were only six such studies (MacArthur et al., 2016). As such, it is difficult to draw conclusions on this evidence base in terms of whether either peer or parental influences should be targeted before the other. More research is needed. However, the greater number of parental interventions contrasts with how peer effects had higher ACEs in Chapter 8. Indeed, the finding in Chapter 9 of this thesis, that peer effects were the most important mediators of the harmful effect of maternal drinking, may have implications in that it suggests that successful intervention on peer effects (drinking and smoking especially) may also reduce some of the harm caused by maternal drinking. This is based on several assumptions, including: non-replacement by competing risks (i.e. that the indirect effect through other mediators wouldn't increase in the absence of the indirect effect through peer effects); that the indirect effect does not simply represent past drinking; that adolescents do not overestimate or overstate peer drinking; etc. If these assumptions hold, then directing resources towards interventions on peer effects may be a more efficient option for reducing adolescent alcohol harm, as successful intervention on a mediator has the added benefit of acting on the indirect effect of the exposure, and peer effects appeared more harmful in these analyses. It is possible that this benefit may be further augmented by the comparative success of peer-led interventions compared to parental interventions.

As noted in Chapter 7, ALSPAC participants were as likely to have harmful AUDIT scores as to score 0, indicating a high degree of underage alcohol consumption. This directs attention on the legal age of purchase, and whether raising it could reduce adolescent alcohol harm, especially given the importance of peer effects. However, this finding must be given a more cautious interpretation in light of the widely reported decline in youth drinking in the UK over the last decade and more (Fat et al., 2018, Oldham et al., 2018). The ALSPAC cohort were age 16.5 in 2007-8. In terms of generalisability, it is unclear how their alcohol outcomes and wider experiences would compare to current UK adolescents (e.g. exogeneous influences that are different include access to internet technology, experience of recession, availability of e-cigarettes, etc.). As such, repeating these analyses in a more recent cohort is a clear option for further research. For example, age 14 data for the Millennium Cohort Study (MCS) has been available since 2017 and the anticipated release for age

17 data is early 2020. MCS participants were age 17 in 2018. As such, the conceptual I-DAG from Chapter 6 could be applied to the MCS to develop a data I-DAG which could then direct similar analyses to those in this thesis. This is covered further below.

10.3.2 Limitations

10.3.2.1 Data and Generalisation

Many of the key limitations of these results derive from the decision to use ALSPAC as the sole data source. As alluded to in Chapter 7, this was despite ALSPAC being the best available option. Nevertheless, these limitations were across several dimensions. First, the external validity of ALSPAC is not only inhibited by attrition bias, but the cohort was of relatively high SES compared to the UK average even at baseline. Thus, while the multiple imputation process used herein appeared to reduce the effect of attrition bias to an extent, the ability to generalise these findings to the UK population is reduced by a degree of collider bias. Further, as noted above, even if ALSPAC had been representative of adolescents prior to 2008, it may be less so for UK adolescents in 2019, given trends in alcohol consumption, and a political climate still resonating to the 2008 financial crash. Secondly, as with any secondary data source, what is measured and how it is measured was problematic for some concepts. For example, there was a strong reliance on self-report data; there were several unmeasured confounders and parental exposures; low participation rates from fathers limited the extent to which their data could be used; the only measure of peer drinking potentially conflates peer drinking with the adolescent's own; and the only available option for the outcome was a measure that has not been fully validated for use on adolescents. The timing of the data was also a notable limitation and fundamentally shaped the analyses. For example, it was the key reason for using maternal drinking as the exposure for the mediation models; it caused some indirect effects to be estimated with fewer EIMOCs than others; and the ACEs for the later variables were estimated with far more confounders in their IPW. Finally, the lack of repeated measures necessitated mediation analysis rather than more in-depth models such as time-varying confounding analysis. However, it must be noted that, despite these limitations, the data I-DAG was very comprehensive and covered 80.1% of concepts from the conceptual I-DAG.

10.3.2.2 Pragmatic decisions

As noted in Chapter 3, binary measures were preferred for each exposure because non-binary variables are more unbalanced in weighted analyses. However, there were also pragmatic reasons for this decision. Firstly, the analysis plan placed an emphasis on breadth – many variables were

investigated as exposures and then as mediators. Specifically, 26 ACEs and 20 NIEs were estimated, and maternal drinking, maternal smoking, and father's drinking were all investigated using regression modelling. Thus, preferring binary measures for all exposures had the advantages of easier comparison between results and easier operationalisation in software. A second reason was simply the relative lack of normally distributed continuous variables. Most skewed heavily towards 0. While these might have been transformed, doing so would have further complicated interpretation and operationalisation. Another set of pragmatic decisions that had limitations were the steps taken to make the causal mediation analysis more efficient. As described in Chapter 9, this was an extensive strategy. The decision that was most problematic was how the total number of EIMOCs was reduced to a maximum of six. While these variables were selected systematically, this reduction arguably reduces conditional exchangeability in the mediation models. Another decision was to include further inclusion criteria when building the DAGs. Studies from the U.S.A. were largely excluded, although the data I-DAG might be similarly useful using American data. Note that part of the justification for excluding American studies was the different legal age of purchase. Relatedly, the decision to exclude genetic causes from the data I-DAG was also made for pragmatic reasons.

10.3.2.3 Missing data

This thesis used multiple imputation of chained equations (MICE) in order to reduce the attrition bias in the ALSPAC sample. However, this process makes at least two assumptions that are largely untestable. Firstly, in order for MICE to work, the missingness in the data must be 'missing at random' (MAR). In short, this means that the missingness of the data must be able to be explained by the non-missing data. Relatedly, the multiple imputation models must be correctly specified (i.e. correct variables, correct measures of variables, correct relationships between variables, etc.). There is no empirical way to test whether these assumptions are met. However, as noted in Chapter 7, the imputed data were much more similar to the age 12 ALSPAC data, than the age 16.5 data, and there appeared to be no problems with the imputation models themselves (Appendix B).

10.3.2.4 Scope for valid causal inference

Chapter 3 outlined several assumptions for valid causal inference: conditional exchangeability; consistency; positivity; and non-interference. As noted in Chapters 8 and 9, there were instances wherein statistical models were in breach of one or more of these assumptions. Positivity was generally not a problem due to how it can be directly assessed using the `teffects` programme. Similarly, while there was clear unmeasured confounding, one of the great strengths of the data I-DAG was how it systematically allocates confounders, and thus makes great strides towards conditional exchangeability as possible. Arguably, while conditional exchangeability was not

achieved, the estimates were much more 'conditionally exchangeable' than in most statistical analysis because of ESC-DAGs. However, non-interference and consistency were problematic. Non-interference cannot be assumed given how social and geographical clustering were not accounted for and consistency is generally problematic when using subjective measures. Further, problems with consistency were exacerbated by the reliance on binary measures. Thus, it cannot be argued that the estimates in this thesis are 'causal estimates' in the truest sense.

10.3.3 Further research

This thesis indicated several avenues for further investigation, especially for using ESC-DAGs to study adolescent alcohol harm. Besides more general ideas such as applying ESC-DAGs to other outcomes or further development the method, a more specific programme of work was developed that can be divided into two groups – planned future research; and other ideas that may be pursued by the author in future, or that would be valuable for other researchers to consider. The planned research will focus on performing multiple mediation analysis. Specifically, a multiple mediation analysis comparing the group of peer mediators to the group of substance-related intra-personal mediators. This would allow for comment on which group was responsible for more of the indirect effect. This was deemed as beyond the scope of this thesis for two reasons. Firstly, the largely negative results for parental mediators would have required a great deal of resource allocation towards concepts and variables that were not the substantive focus of this thesis (i.e. that were not the focus of the review of systematic reviews). In other words, multiple mediation models that do not focus on parental mediators were not within the scope of the research objectives. Secondly, the technical challenge is a notable increase on what has already been performed in a methodologically broad project. For example, while modelling multiple mediators collectively can give an indication of the total indirect effect though a groups of mediators, methods for decomposing the pathways between mediators (e.g. comparing path M1 → M2 → Y to M1 → M3 → Y) in settings with EIMOCs are still being developed such that no statistical packages are yet available (Steen et al., 2017). Thus, multiple mediation models of non-parental mediators were set for post-thesis work.

The other research can potentially be conceptualised as a broad project with three 'pillars'. The first pillar would involve comparisons between different methods for mediation analysis. For example, while there are high-profile studies comparing the various mediation methods (De Stavola et al., 2015), this literature could be added to by using the data I-DAG to compare results from SEM and the Baron & Kenny approach to the results of this thesis. This would allow for systematic comment on how much the EIMOCs affect the indirect effect in an applied setting. Further, the data I-DAG identified a much greater degree of DAG-based sensitivity analysis than could be performed here. The second pillar could involve engaging other researchers to act as ESC-DAGs reviewers to produce

a conceptual I-DAG on the same topic, or on a related topic if necessary (e.g. alcohol-related deaths or hospitalisations). This would test the reproducibility of ESC-DAGs as well as contribute to the robustness of DAGs in alcohol studies in general. This feeds directly into the third pillar, alluded to above, which would involve applying the conceptual I-DAG(s) to the MCS to produce a data I-DAG which could be compared to the ASLPAC data I-DAG. At least in terms of how recent it is, results from the MCS would be more immediately generalisable to the UK adolescent population.

Additionally, further testing of the ESC-DAGs is scheduled for mid-2020, involving collaboration with researchers who have not used DAGs before, specifically to attempt to create DAGs pertaining to the effect of changes in minimum legal drinking age on alcohol harm in adolescents. Besides acting as a testbed for how ESC-DAGs can be used more broadly in practice, this will also provide an avenue for further dissemination. Further, besides teaching in universities (MPH at University of Glasgow and University of Edinburgh), knowledge sharing events have been scheduled on ESC-DAGs at NHS Health Scotland in mid-2020.

Finally, ESC-DAGs may help to bridge the gap between increasingly popular complex systems methods (such as agent-based modelling) and wider epidemiology. For example, agent-based models focus on programming simulated individuals with 'parameters' (i.e. characteristics under pre-specified rules such as statistical associations). The DAGs produced by ESC-DAGs may be helpful in determining these parameters in much the same ways as for variables selection for statistical models. Indeed, ESC-DAGs may be able to inform the causal theory behind other complex systems methods such as social network analysis. Investigating this potential is a clear and topical avenue for further research.

10.4 Concluding remarks

In summary, this thesis identified a gap in the literature on how directed acyclic graphs (DAGs) are used – specifically there are no systematic methods or guidelines for building DAGs. ESC-DAGs was developed in response to this gap and was applied to the literature on parental influences of adolescent alcohol harm. A data I-DAG was then produced by further applying ESC-DAGs to data from the Avon Longitudinal Study of Parents and Children (ALSPAC). This was a wide-ranging and comprehensive causal diagram that usefully directed several forms of analysis, including DAG-based sensitivity analysis using multiple regression; estimation of average causal effects (ACEs) using inverse probability weighting (IPW); and estimation of natural in/direct effects using causal mediation analysis in the form of mediational g-computation. Several parental influences were found to be harmful, especially maternal drinking, parental permissiveness towards adolescent alcohol use, and

parental monitoring. Subsequently, indirect effects of the effect of maternal drinking on adolescent alcohol harm were found for peer effects and the adolescent's substance use and attitudes, whereas parenting and the parent-child relationship were not found to be important mediators. This Discussion chapter has argued that ESC-DAGs as a process succeeded in its goal but requires further testing of its reproducibility and improvement of its efficiency. The findings had notable strengths, triangulated with the wider research literature, and had novel implications.

This thesis argued in Chapter 1 that the potential of DAGs for improving statistical modelling in Public Health is great, but that further methodological developments are required before this potential can be realised. One crucial area in which developments are required is methods for building DAGs. The I-DAGs produced from the ESC-DAGs application in this thesis were extensively demonstrated to be useful for directing complex causal analysis. As far as the author is aware, no other method has been successful in this regard. As such, as well as contributing novel findings on alcohol harm in adolescents, this thesis has contributed a method which may help improve statistical practices in health and social sciences.

Appendices

A: Multiple imputation process

Imputed Variables

As described in Chapter 3, Multiple Imputation of Chained Equations (MICE) works by iteratively modelling each incomplete variable on all others in the imputation model, and then using the information from these models to replace the missing data. The imputation literature recommends that any variable that is used in the analytical models should also be included in the imputations (Royston and White, 2011, White et al., 2011). Thus, all variables from the data I-DAG were included in the multiple imputations, including all baseline confounders, parental influences, intermediates, and the outcome. The process resulted in an analytical sample of N=7,959. Table A1 below lists each imputed variable and the number of cases that were imputed.

Table A1: Imputed Variables

Variable	Age	N imputed	% Imputed (of 7,959)
AUDIT	16.5	3,207	40.3
Relationships with teachers	16	3,904	49.1
GCSEs	16	1,268	15.9
IQ	15.5	3,492	43.9
Sexuality	15.5	3,301	41.5
Peer anti-social behaviour	15.5	3,152	39.6
Peer substance use	15.5	3,161	39.7
Parental monitoring	15.5	3,134	39.4
Parental punishment	15.5	3,165	39.8
Parent-child tension	15.5	3,148	39.6
Parental involvement	15.5	3,144	39.5
Secret-keeping	15.5	3,804	47.8
Anti-social behaviour	14	2,372	29.8
Attitudes towards alcohol	14	2,452	30.8
Smoking	14	2,595	32.6
Substance use	14	2,596	32.6
School engagement	14	2,913	36.6
Neighbourhood safety	14	2,656	33.4
Sensation seeking personality	13.5	2,588	32.5
Depression	13.5	2,559	32.2

Peer drinking	13.5	3,084	38.8
Peer smoking	13.5	2,645	33.2
Number of friends	13.5	2,536	31.9
Early alcohol initiation	12.5	4,664	58.6
Parental permissiveness	12.5	4,286	53.9
Pubertal timing	12.5	3,064	38.5
Maternal drinking	12	1,355	17.0
Maternal smoking	12	1,251	15.7
Maternal depression	12	1,155	14.5
BMI	11	1,810	22.7
IMD	10	1,911	24.0
Urbanity	10	1,682	21.1
Maternal religiosity	8	1,245	15.6
Older siblings	6	960	12.1
Maternal education	0	783	9.8
Maternal stress	0	948	11.9
Family history of alcohol problems	0	518	6.5
Social class	0	1,538	19.3
Birthweight	0	407	5.1
Maternal age at birth	0	310	3.9
Maternal smoking during pregnancy	0	1124	14.1
Maternal marital status	0	470	5.9
Ethnicity	0	638	8.0
Sex	0	0	0.0

There were three ways in which these variables could have been employed – as imputed, regular, or passive. Imputed variables are those for which missing data is replaced – they can be thought of as independent variables in all models except for the one in which they are the dependent variable. Regular variables can only be independent variables. They have no missing data that is imputed, and as such are usually variables that are fully observed. Passive variables are generated from the imputations and are commonly interaction terms. However, asides from sex which was modelled as a regular variable, the imputations in this thesis only used imputed variables. There were no passive variables because interaction terms were modelled using the Just Another Variable (JAV) approach, as discussed in Chapter 3. The comparative lack of regular variables was because few baseline confounders were fully observed and the missingness pattern across them is non-monotone. For

example, Table A2 conveys how around 10% of the sample had a non-monotone missingness pattern between four of the baseline confounders measured during pregnancy: maternal marital status, family history of alcohol problems, maternal education, and class.

Table A2: Non-monotone missing pattern in baseline confounders in ALSPAC

Marital status	Class	Family history	Maternal education	Percent
1	1	1	0	4
1	1	0	1	1
1	1	0	0	<1
1	0	0	1	3
0	1	1	1	<1
0	1	1	0	<1
0	1	0	1	<1
0	1	0	0	<1
0	0	1	1	<1
0	0	1	0	<1
0	0	0	1	1

Sample

Treating almost all variables as imputed presented two distinct problems. First, the sample size for the analytical sample would be the same as for the variable with the most observations (14,841 for sex). Second, as discussed briefly in Chapter 3, there would be participants for whom most variables were imputed. In other words, information from individuals who dropped out within the first year, would be used to inform imputed values for the rest of the sample, including variables measured 15 years later. To avoid this, the sample was restricted *before* performing the imputations. Participants who were observed for the outcome and for one of any of the maternal questionnaires from ages eight to 11 were retained. This resulted in an analytical sample of 7,959 participants who had data for both the outcome and late childhood.

Missingness Predictors

Chapter 3 also noted that variables that predict the probability of missingness should be included in the imputations also. For this analysis, the same 'missingness predictors' were used that have been used in other studies of ALSPAC focusing on adolescents (Mahedy et al., 2018). They were: social class measured by NS-SEC; whether or not there was a car at home; household conditions measured by asking if there was damp in the home; parity between mother and partner; a binary measure of

maternal smoking during pregnancy; maternal education; and maternal marital status. Each was measured during pregnancy. However, it is not their correlation with observed or imputed values that is important, but rather their ability to predict the presence or absence of data. Table A3 represents this with odd ratios from unadjusted logistic regressions on the probability of having missing data in the AUDIT score at age 16.5. Each predicted missing data on this variable except for household conditions, which was subsequently not included as a missingness predictor.

Unadjusted odds ratios on indicator of missingness for AUDIT score

Variable	OR	95% C.I.
NS-SEC social class		
i	1	
ii	1.51	1.21, 1.87
iii (non-manual)	1.87	1.50, 2.32
iv (manual)	2.25	1.7, 2.97
v	1.83	1.27, 2.39
Car ownership	1.56	1.27, 1.91
Household Conditions	1.03	0.94, 1.13
Parity	1.06	1.04. 1.38
Smoking during pregnancy	1.37	1.20, 1.57
Maternal education		
Degree	1	
A-Level	1.57	1.35, 1.84
O-Level	2.04	1.76, 2.36
Vocational	2.55	2.01, 3.12
CSE	2.90	2.40, 3.51

Interactions

Given the high volume of variables being imputed, a relatively low number of interactions were modelled. This is because a higher degree of imputed variables (especially binary variables) reduces the chance that the imputations will 'converge'. Imputed variables and missingness predictors were prioritised over interaction terms such that the interaction terms were the last ones that were added to the model and were added one by one until the imputation failed to converge. As noted in Chapter 9, the only variables that interacted with maternal drinking when modelled on AUDIT score were parental monitoring and parent-child tension. These interaction terms were both included, as were the interactions for both with the outcome. Finally, an interaction between the exposure and the outcome was also added. Each was modelled using the JAV approach – the interaction term was generated

beforehand. Imputing categorical interaction terms can be problematic as it is possible that non-sensical values can be imputed. For example, an interaction term between sex and parental monitoring could impute a value for a male that can only exist for females. To overcome this, the JAV interaction terms were derived from the numerical versions of the variables in question. Specifically, number of units drank last week by the mothers, AUDIT score of the adolescent, the compiled monitoring score (up to 12) and the compiled parent-child tension score (up to 6). Each of the interaction terms was then transformed into normal distributions and modelled using linear regression in their respective imputations. This took the total number of variables used in the imputations to 50.

B: Diagnostic tests of imputation process

There should be no correlation between the current imputation and the previous or subsequent imputation. For example, take the mean of X in imputed dataset m1 - it should be no more successful at predicting the mean of X in m2 than it is at predicting the value of the mean of X in m3, and so on. Thus, there should be no detectable trend between the means (allowing for how they will be within a certain range of the observed mean). For both numerical and categorical variables, checks are commonly performed by plotting a line graph in which the x axis corresponds to the count of iterations (i.e. *m*) and the y axis is the mean for the variable in question. Problematic imputations would thus be represented by a clear trend across the iterations. Line graphs for the mean and standard deviation of the imputed outcome (AUDIT) are included in Figure B1 below, and the mean of each parental variable is presented in Figure B2. As can clearly be seen from these graphs, there was no convergence problem with the imputations as the datapoint for any iteration is as likely to be above the previous iteration as below it.

Figure B1: Convergence of means and SDs for AUDIT score across 50 imputed datasets

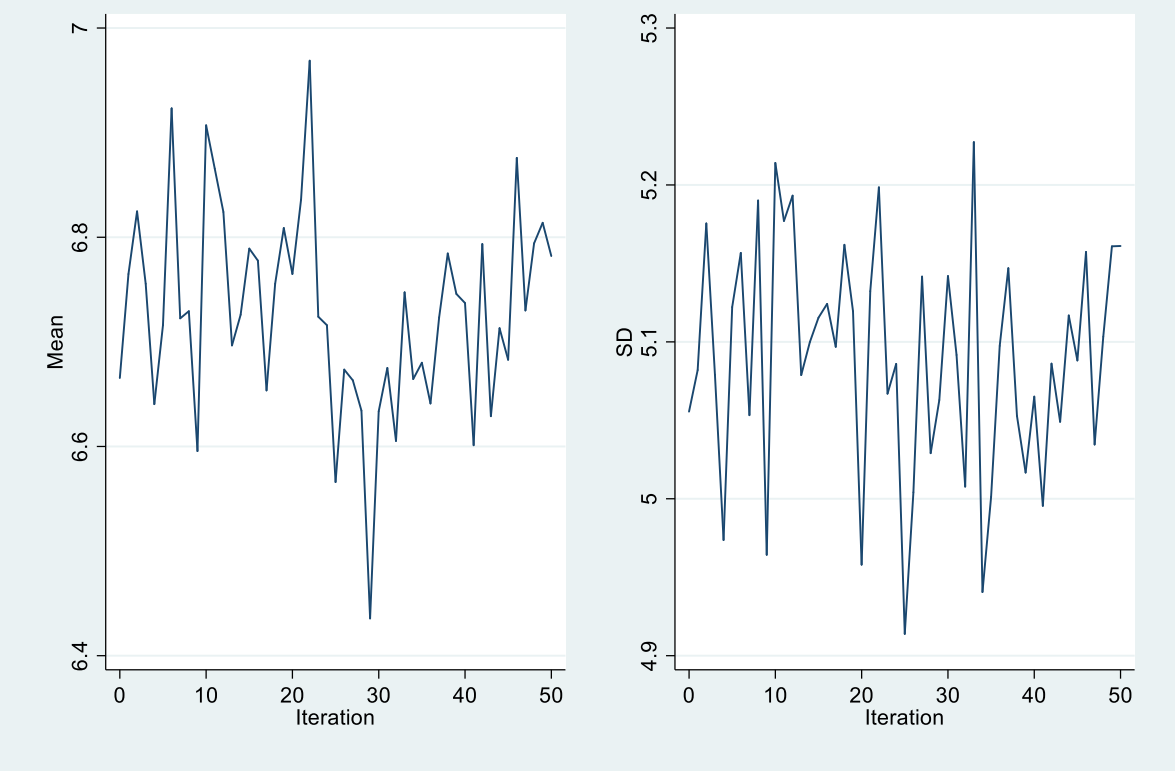
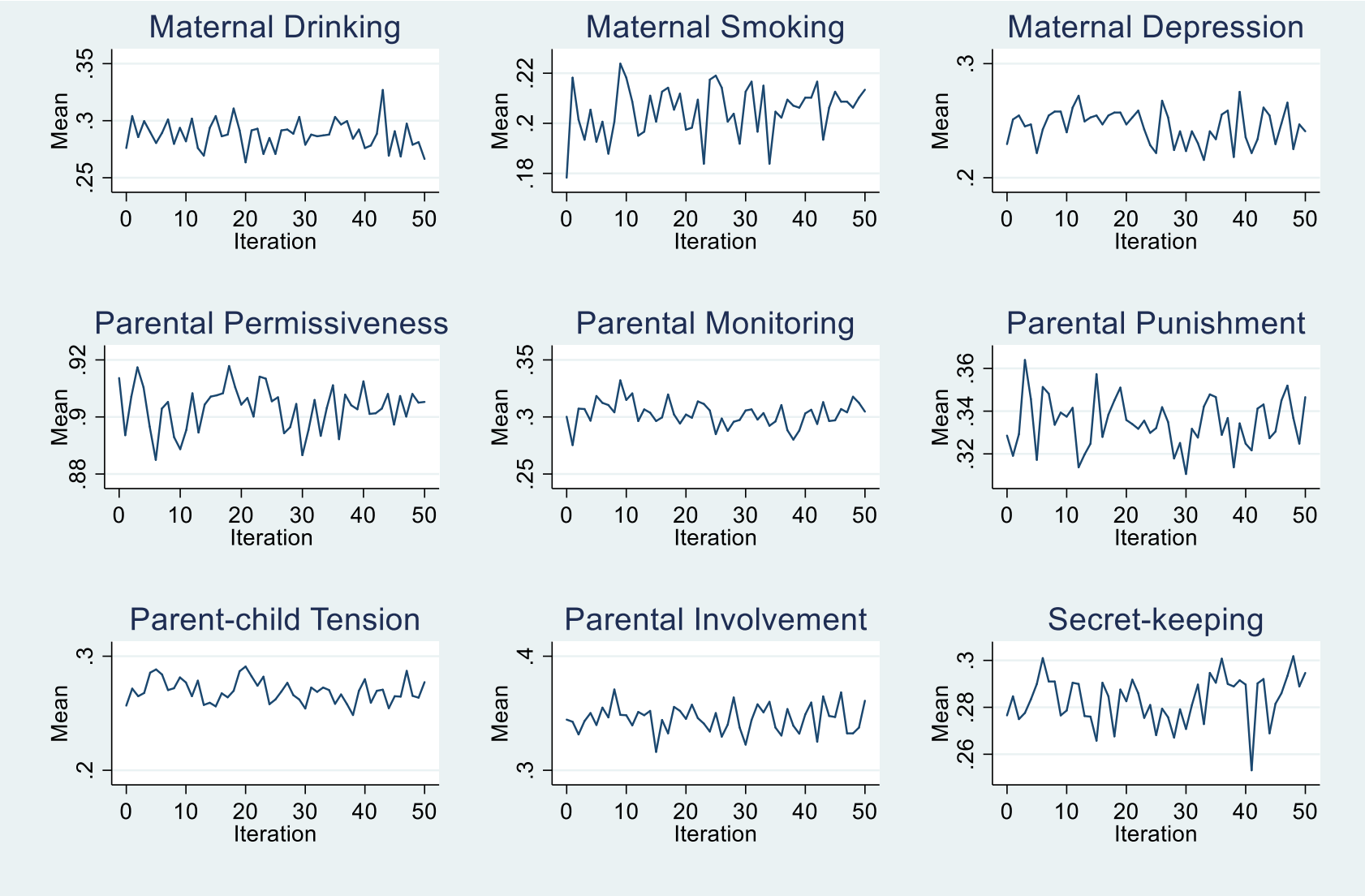


Figure B2: Convergence of means for parental influences across 50 imputed datasets



C: Descriptives of imputed data

This appendix compares the distribution of the observed values to the imputed values and focuses on the outcome and the parental influences. Socio-demographic confounders were presented in Chapter 7 in Table 7-3 (page 124).

Outcome: AUDIT score at adolescent age 16.5 years

A total of 4,752 adolescents completed the AUDIT questionnaire at age 16.5. Of the 7,959 cases in the analytic sample, 3,207 were imputed (32.5%). Predictive mean matching was used in the multiple imputation model for this variable. Table C1 below compares the percentage of observations with AUDIT scores of 0; between 1 and 7 (low risk); between 8 and 15 (hazardous drinking); and 16+ (harmful drinking). It compares the observed data, the average proportion across each *m* for the analytical sample of N=7,959, and the average proportion in the imputed values for the missing observations (N=3,207). Note that these averages only depict the average count of observations in each category across the 50 imputed datasets. The table shows that there is very little difference between the observed and the analytical sample – the proportion in the observed sample with hazardous or harmful AUDIT scores only increased from 36.6% to 36.8%. However, the proportion of the imputed sample with AUDIT scores over 8 was higher at 37.5%, compared to 36.6% in the observed sample.

Table C1: Differences in AUDIT score between observed, analytical, and imputed samples

AUDIT score	Sample		
	% Observed (N=4,752)	% Analytical (N=7,959)	% Imputed (N=3,207)
0	5.2	5.2	5.0
Low risk	58.2	58.1	57.4
Hazardous	31.0	31.0	31.5
Harmful	5.6	5.8	6.0
Hazardous or harmful	36.6	36.8	37.5

Comparisons between imputed datasets found only slight variation – the medians were the same at 6 and the interquartile ranges were either 6 or 7. Overall, the imputed AUDIT scores estimated that the missing data was for adolescents of slightly higher AUDIT score than the observed sample, although the effect was small.

Exposures: Parental influences between adolescent age 12-15.5 years.

Table C2 below shows the proportion imputed and the proportion exposed in the observed, imputed and analytical samples (i.e. N=7,959) for the parental influences. Each was imputed using logistic regression. Maternal drinking saw relatively little variation, but the proportion exposed for maternal smoking and maternal depression was larger in the imputed sample. This was in keeping with expectations from literature on attrition bias (Munafò et al., 2017), in that the imputations estimated that the missing data would be from adolescents whose mothers were more likely to smoke and/or be depressed. Each parental influence measured at age 15.5 also had a higher proportion of exposed in the imputed sample than the observed sample. This suggests that the missing data were for adolescents who had worse relationships with their parents, and whose parents subjected them to less discipline and parental monitoring. It is possible that this reflects a relationship between attrition rates and the nature of the parent-child relationship. Parental permissiveness, however had a smaller proportion exposed.

Table C2: Average proportions in imputed vs observed

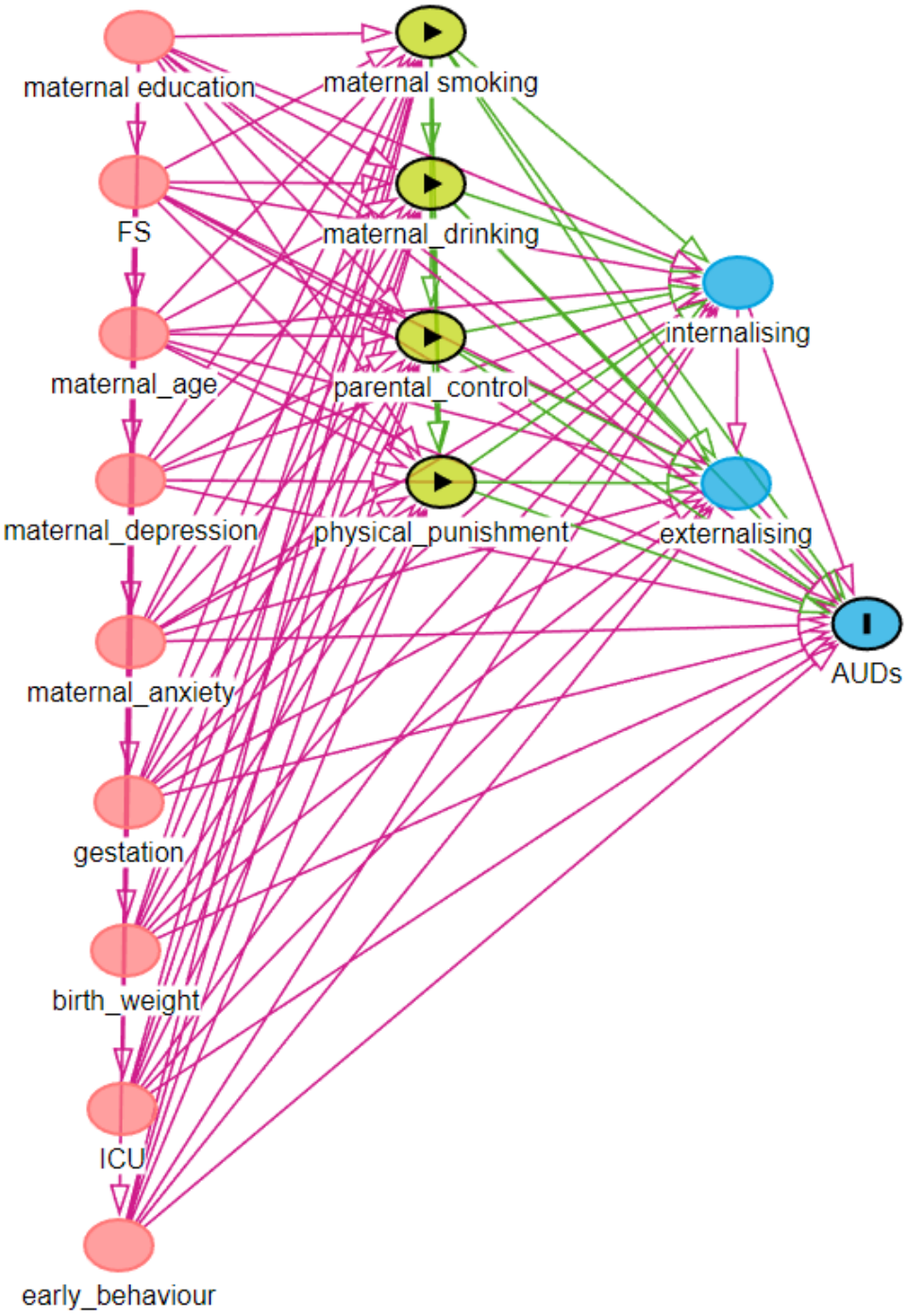
Variable	Age	N observed	N imputed (%)	% Exposed		
				Observed	Analytical	Imputed
Maternal drinking	12	6,604	1,355 (17.0)	30.50	30.24	29.0
Maternal smoking	12	6,708	1,251 (15.7)	16.99	17.6	20.6
Maternal depression	12	6,804	1,155 (14.5)	22.27	22.5	24.0
Permissiveness	12.5	3,673	4,286 (53.8)	92.84	91.3	89.9
Monitoring	15.5	4,825	3,134 (39.3)	28.25	29.1	30.4
Tension	15.5	4,811	3,148 (39.5)	24.67	25.5	29.8
Punishment	15.5	4,794	3,165 (39.7)	33.00	33.3	33.8
Secret-keeping	15.5	4,155	3,804 (47.7)	26.16	27.2	28.5
Shared time	15.5	4,815	3,144 (39.5)	30.30	31.9	34.4

D: DAGs for primary studies used in ESC-DAGs review

This appendix presents the DAGs produced by the ESC-DAGs Translation process, as applied to 50 primary studies. Several points are of note;

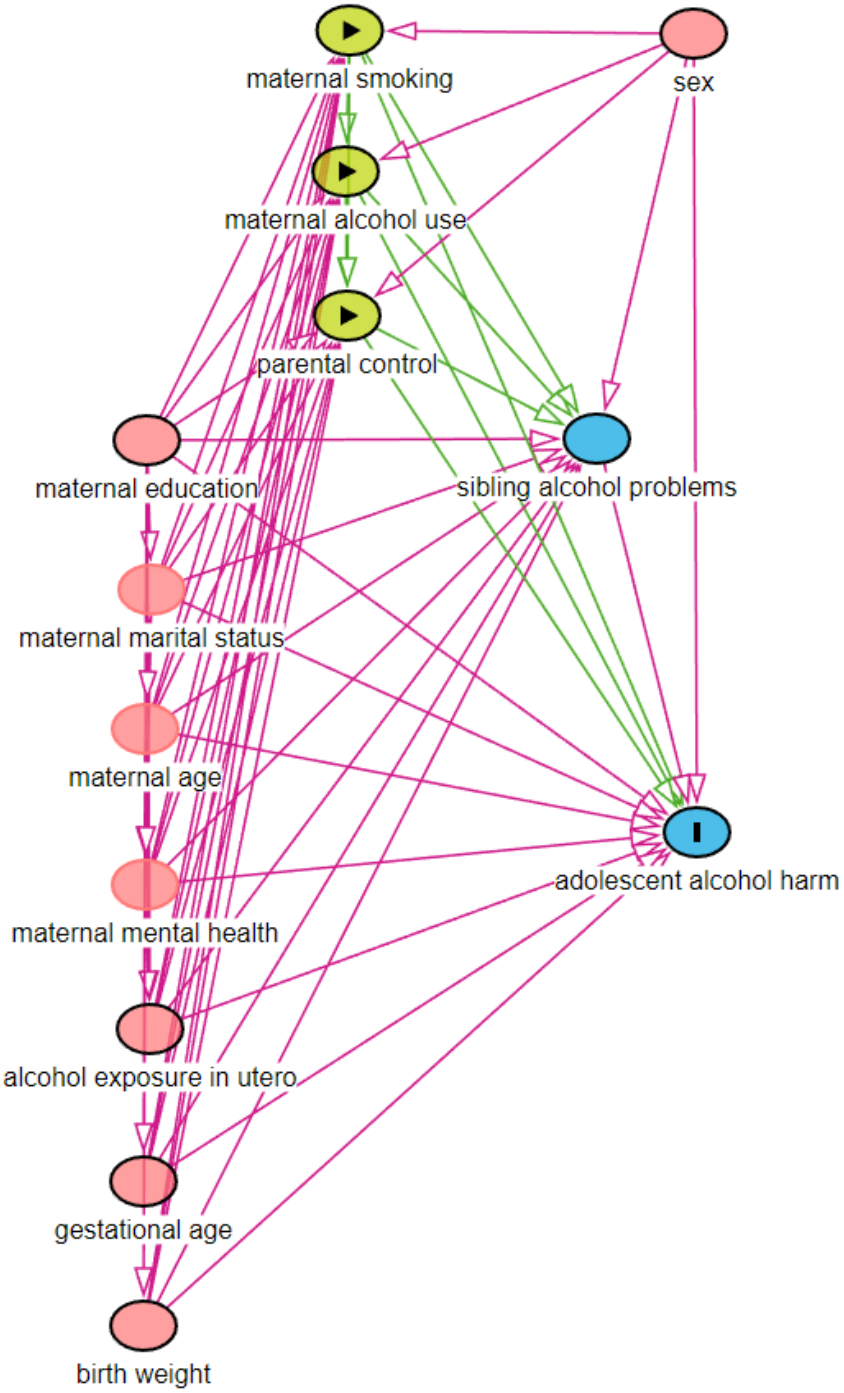
- The key information provided from these DAGs for the rapid ESC-DAGs application in this thesis was the differentiation between confounders, exposures and intermediates. In other words, the directed edge between each exposure and the confounder/intermediates.
- Every effort was made to make the more complex examples as intelligible as possible. Similar principles were applied as in Chapter 7, in that causation generally flows from left to right, and from top to bottom (when variables are arranged in a column). Confounders, exposures, and intermediates were grouped together as much as was possible.
- The labels for some nodes were also abbreviated for presentation. These are explained beneath the DAG for the first instance.
- As discussed in Chapter 10, given that the ESC-DAGs translation process assumes that all concepts/variables are measured at the same time, bi-directional relationships were ubiquitous. Indeed, this would be case for any application of ESC-DAGs to an inherently social exposure-outcome relationship, such as parental influences on adolescent alcohol harm. Thus, except for directed edges associated with minority of sociodemographic variables (e.g. sex), directed edges in the following DAGs do not represent the assumption that causation is uni-directional, but rather than the posited cause is 'more' of a cause than an effect. As noted in Chapter 7, what determines this will often be affected by the research question. This was acceptable given that these DAGs do not represent data.
- There was one DAG (Van der Vorst et al, 2007; number 43, page 266) for which deciding the direction of certain edges prohibited by the degree of similarity between variables and the author's lack of familiarity with the concepts as 'causal variables'. This was not problematic given that the later Synthesis process only focused on relationships between exposures and the outcome.

1. Alati et al. 2005. Early predictors of adult drinking: a birth cohort study.

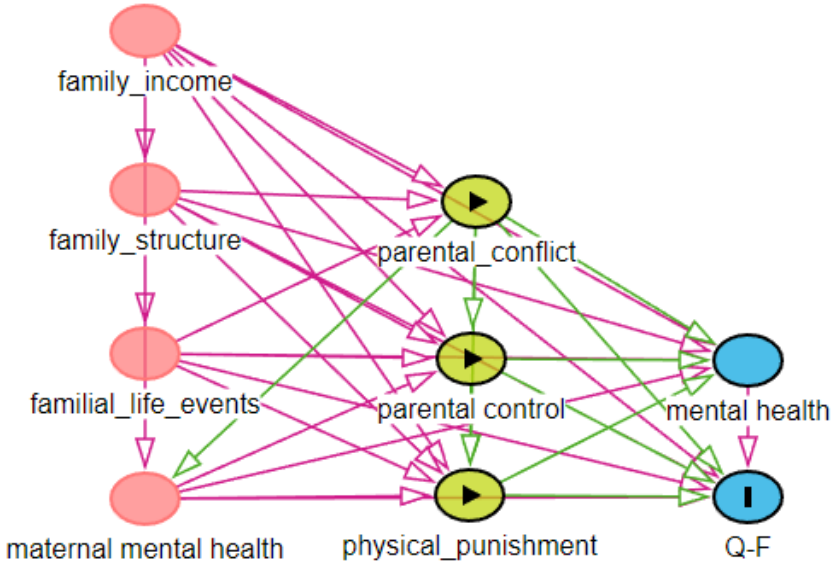


FS = family structure; ICU = postnatal intensive care; AUD = Alcohol use disorders

2. Alati et al. 2008. The developmental origin of adolescent alcohol use: Findings from the Mater University Study of Pregnancy and its outcomes.

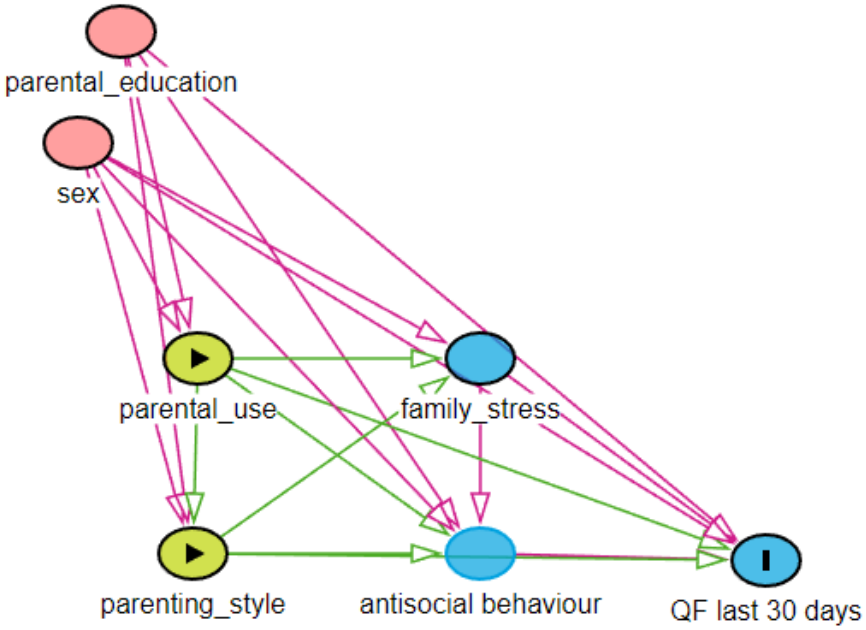


3. Alati et al. 2010. Do maternal parenting practices predict problematic patterns of adolescent alcohol consumption?

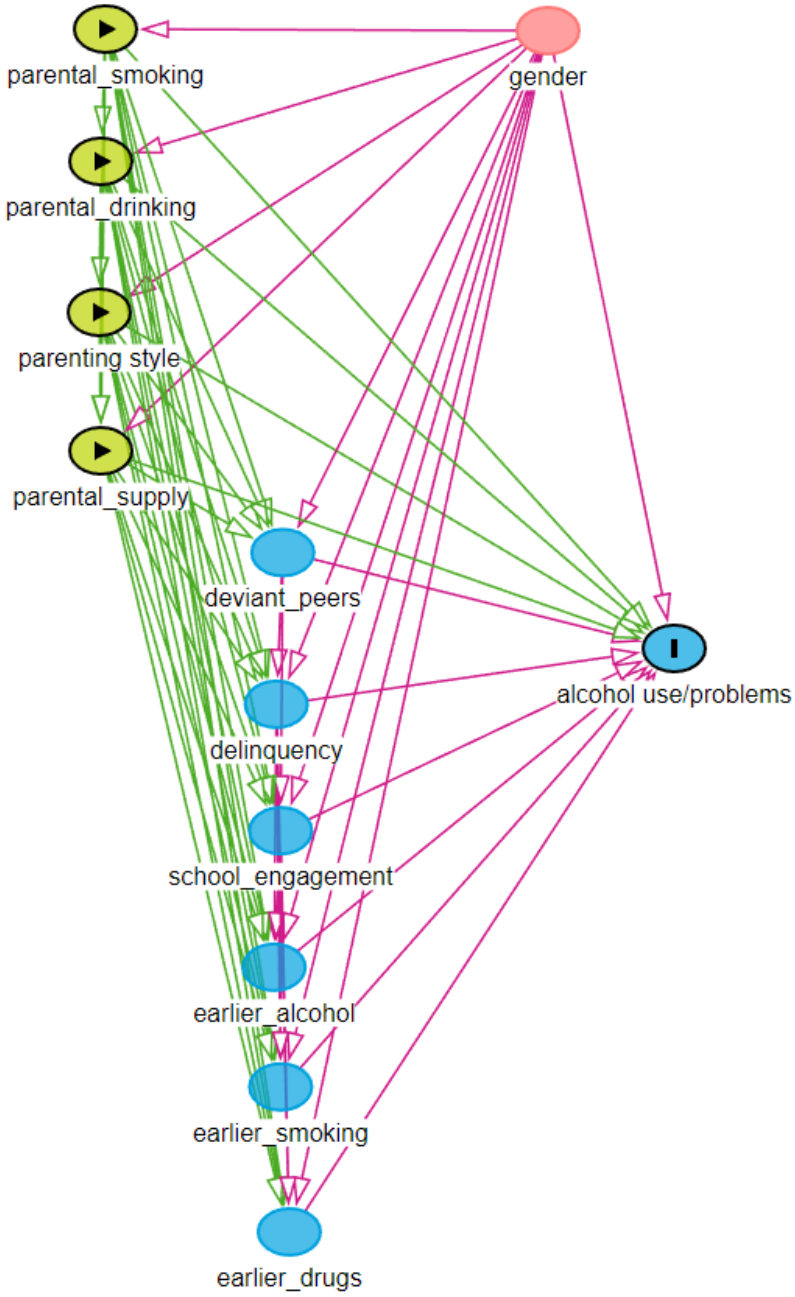


Q-F = quantity frequency measure

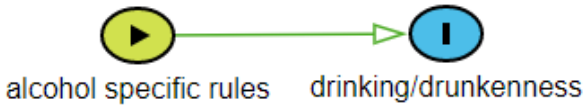
4. Alati et al. 2013. The role of parental alcohol use, parental discipline and antisocial behaviour on adolescent drinking trajectories.



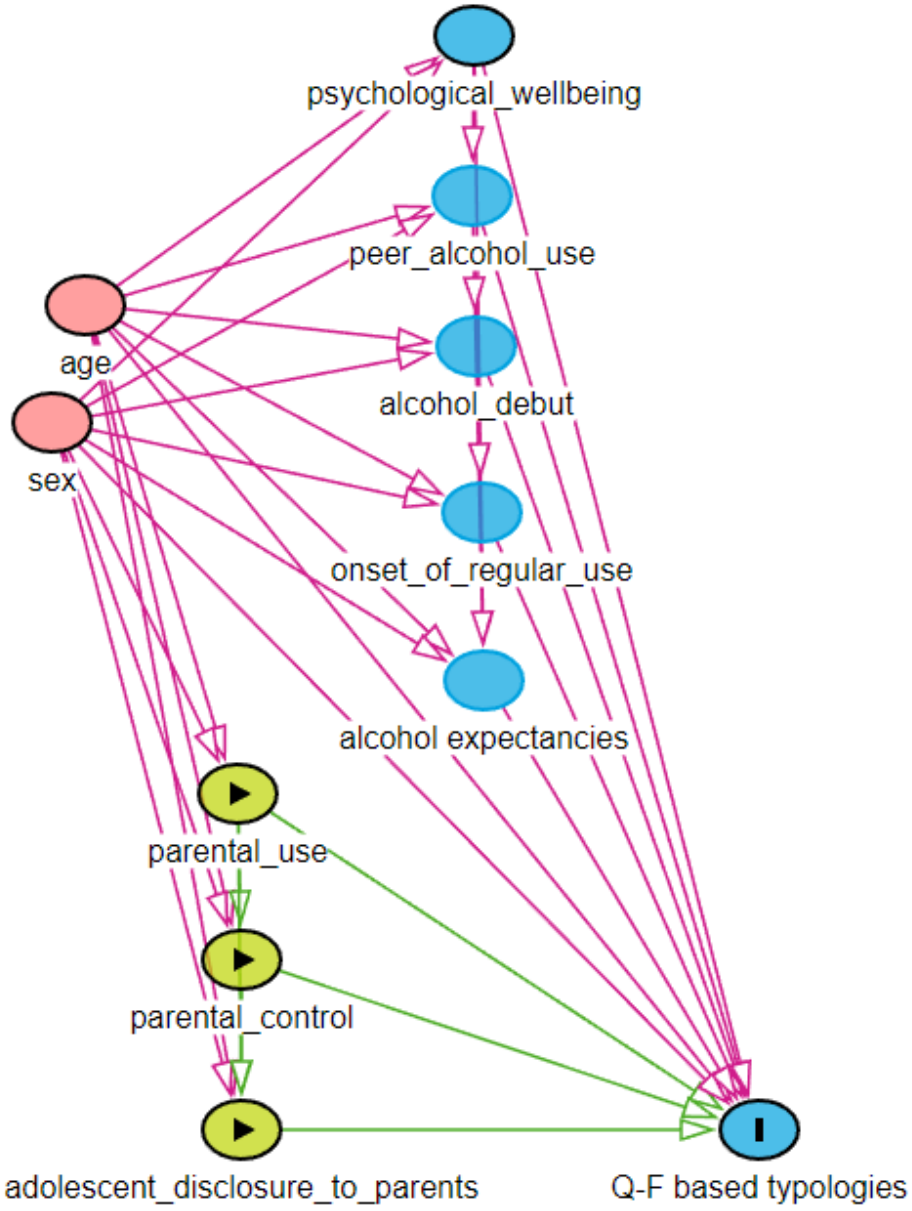
5. Berge et al. 2016. Role of parenting styles in adolescent substance use: results from a Swedish longitudinal cohort study.



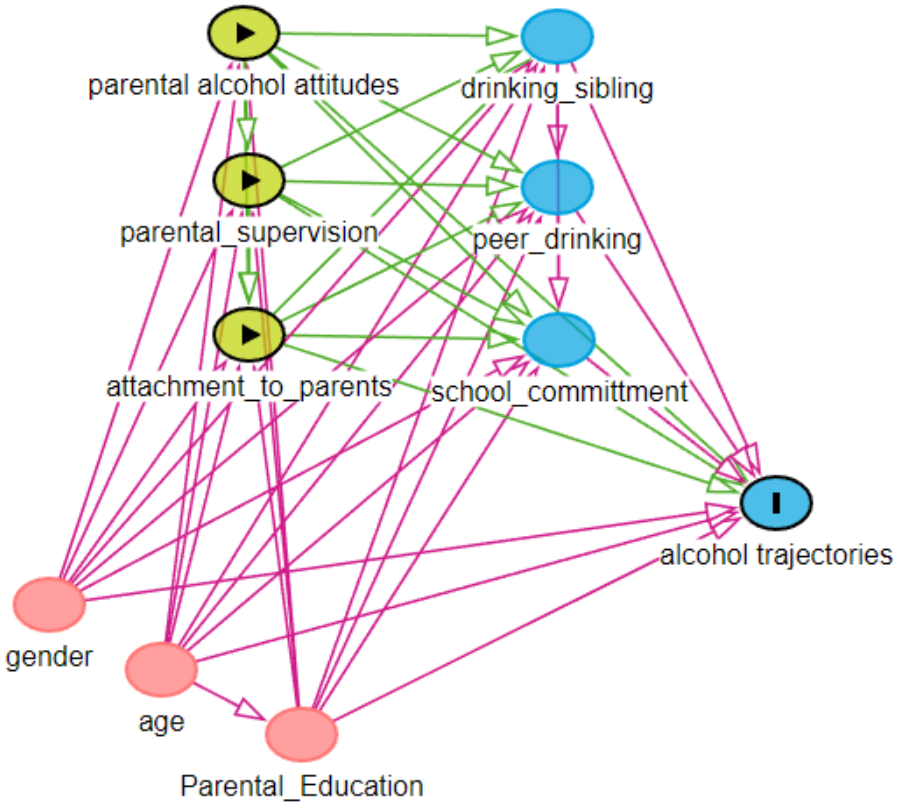
6. Bodin & Strandberg. 2011. The Örebro prevention programme revisited: A cluster-randomized effectiveness trial of programme effects on youth drinking.



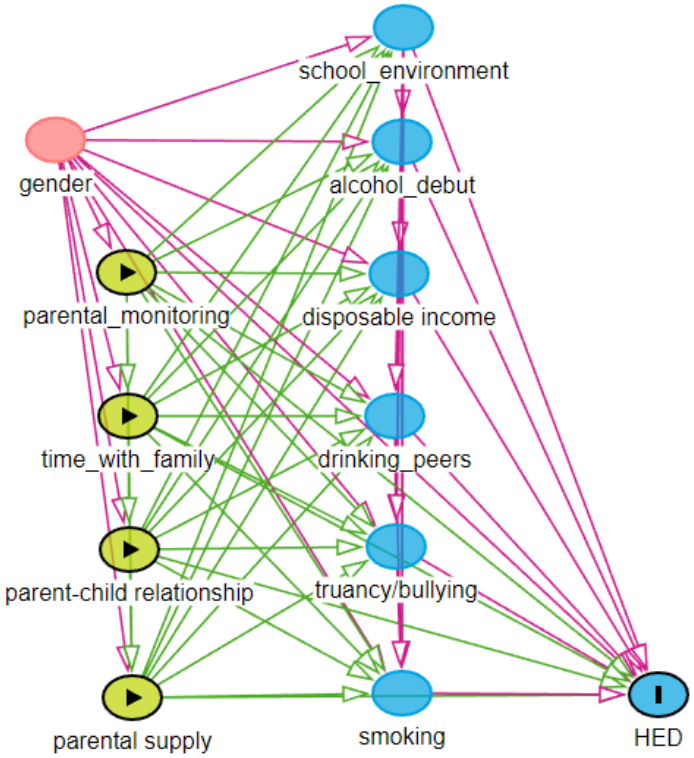
7. Cable et al. 2007. Typologies of alcohol consumption in adolescence: Predictors and adult outcomes.



8. Chan et al. 2013. Predicting steep escalations in alcohol use over the teenage years: Age-related variations in key social influences.

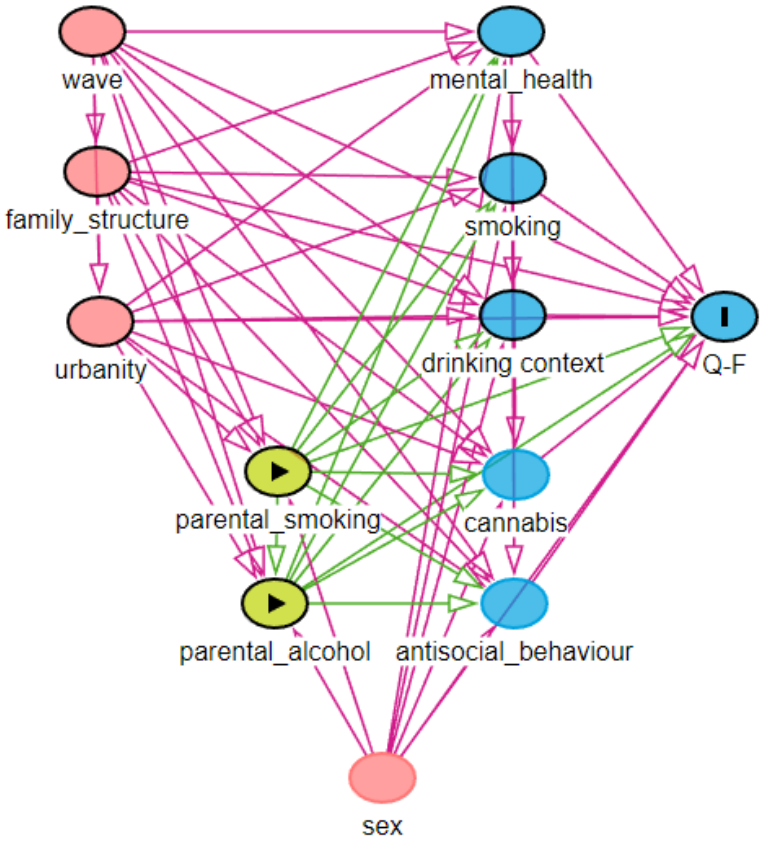


9. Danielsson et al. 2011. Heavy episodic drinking in early adolescence: Gender-specific risk and protective factors.

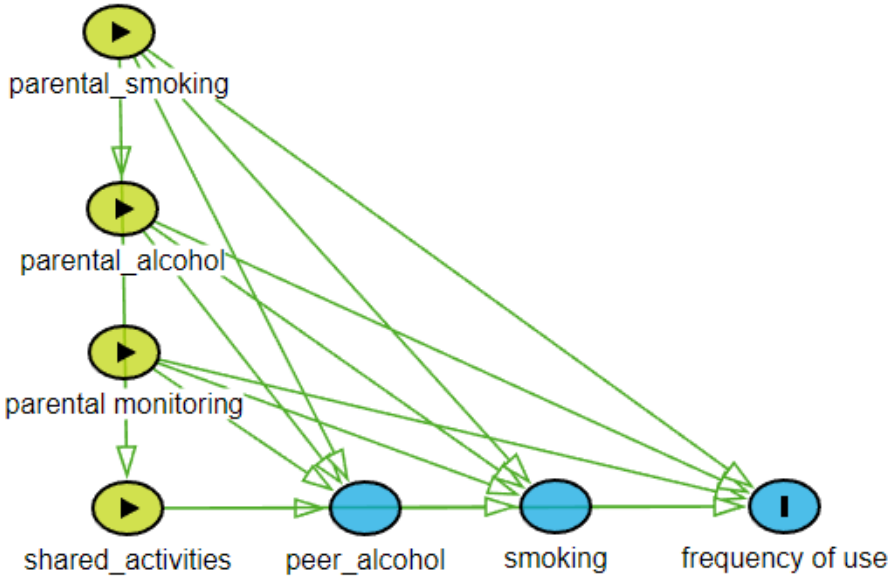


HED = heavy episodic drinking

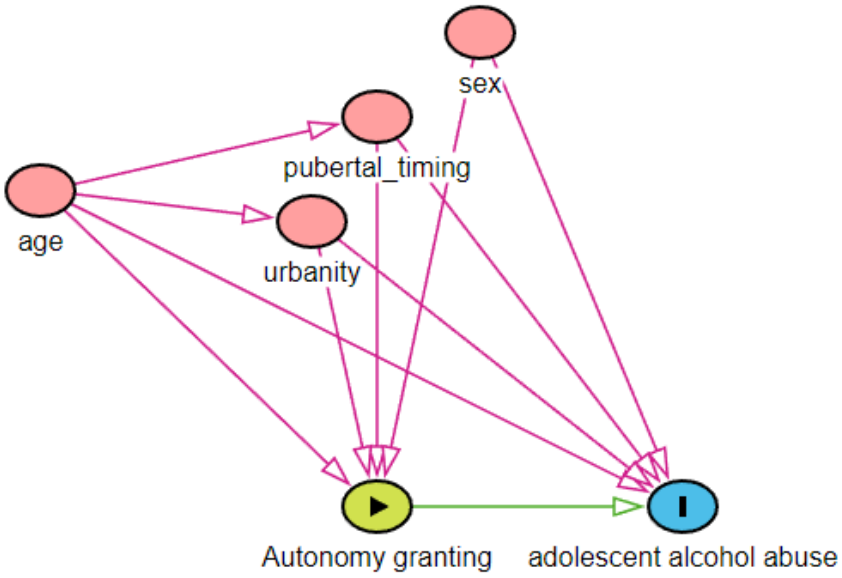
10. Degenhardt et al. 2015. Does the social context of early alcohol use affect risky drinking in adolescents? A prospective cohort study.



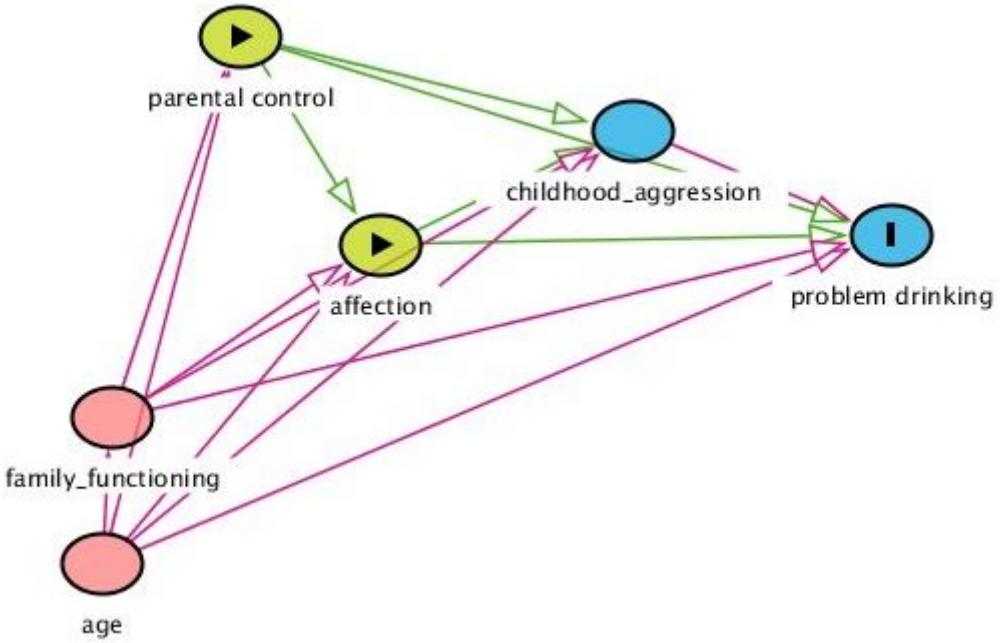
11. Dick et al. 2007. Changing environmental influences on substance use across development.



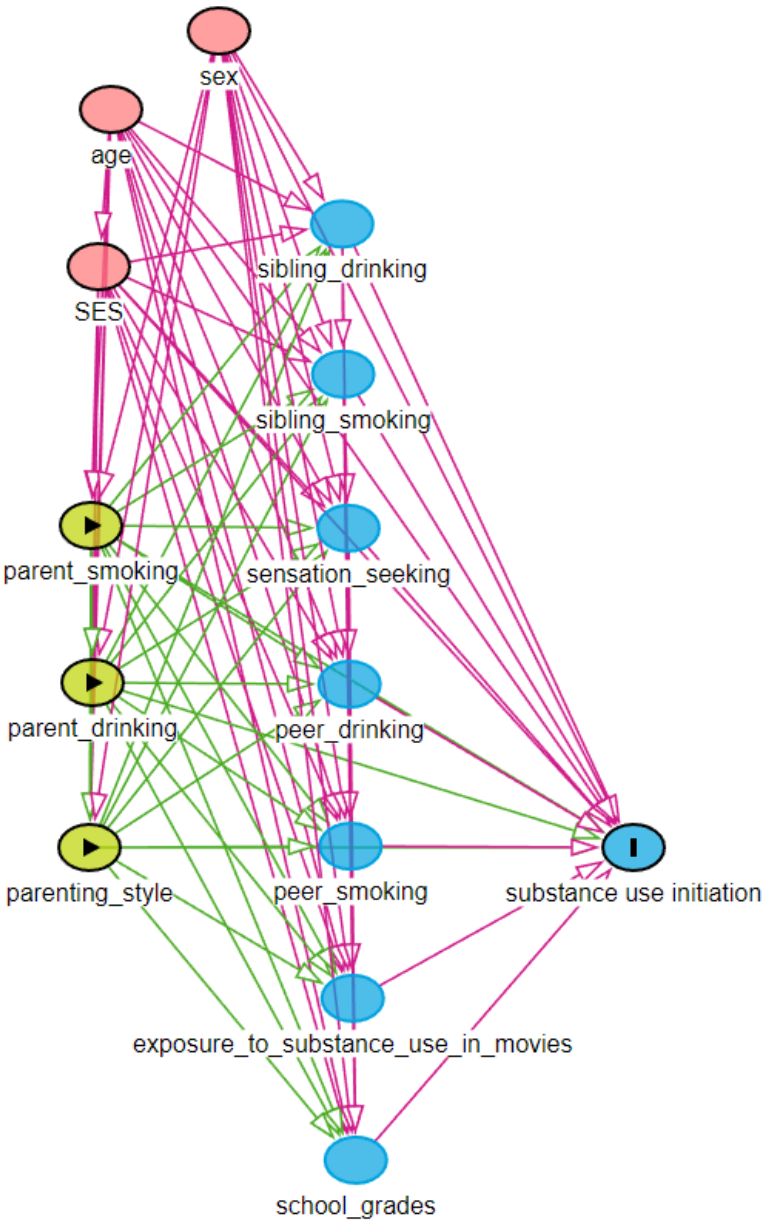
12. Dickson et al. 2015. Parental supervision and alcohol abuse among adolescent girls.



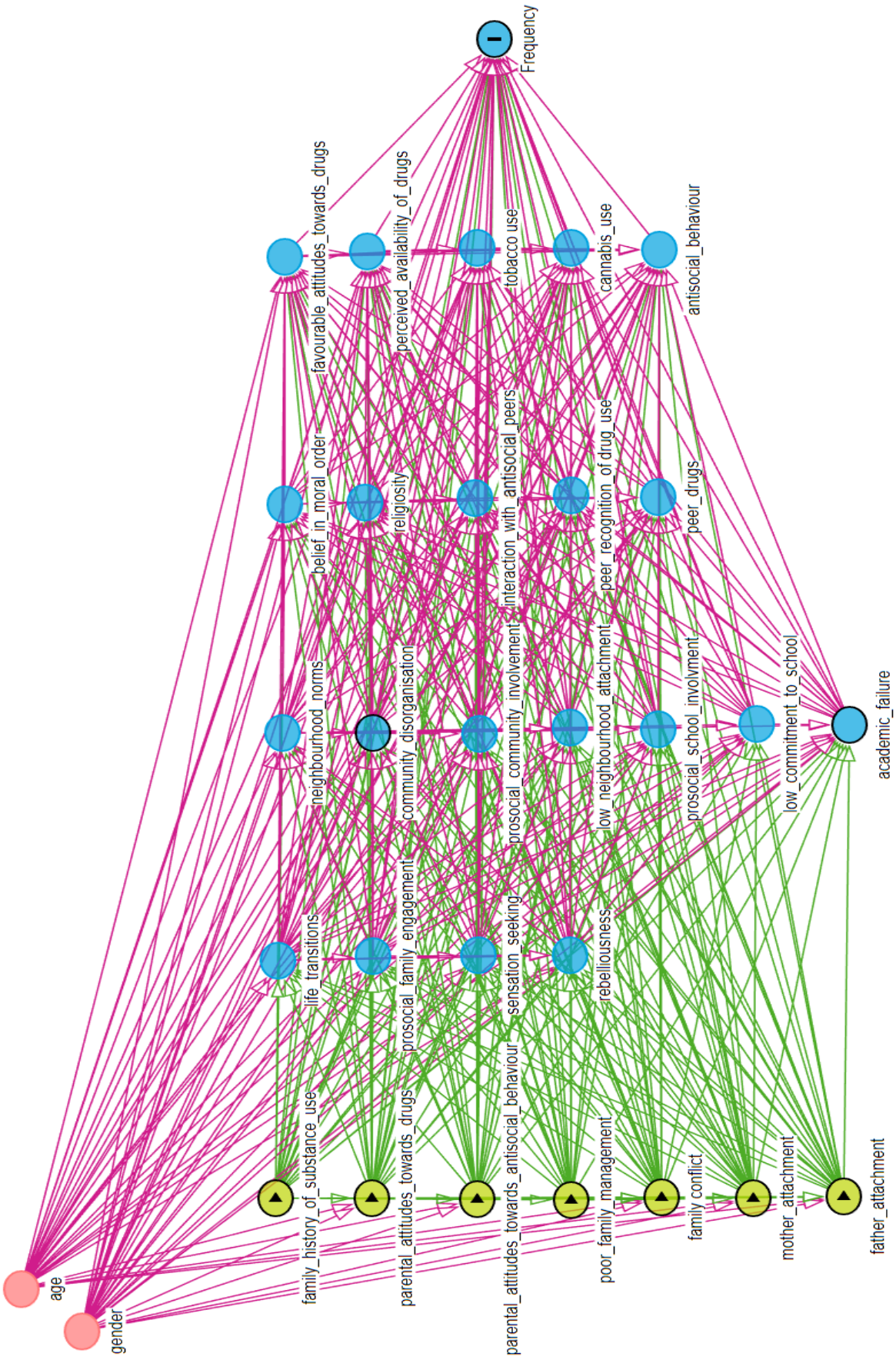
13. Engels et al. 2005. Long-term effects of family functioning and child characteristics on problem drinking in young adulthood.



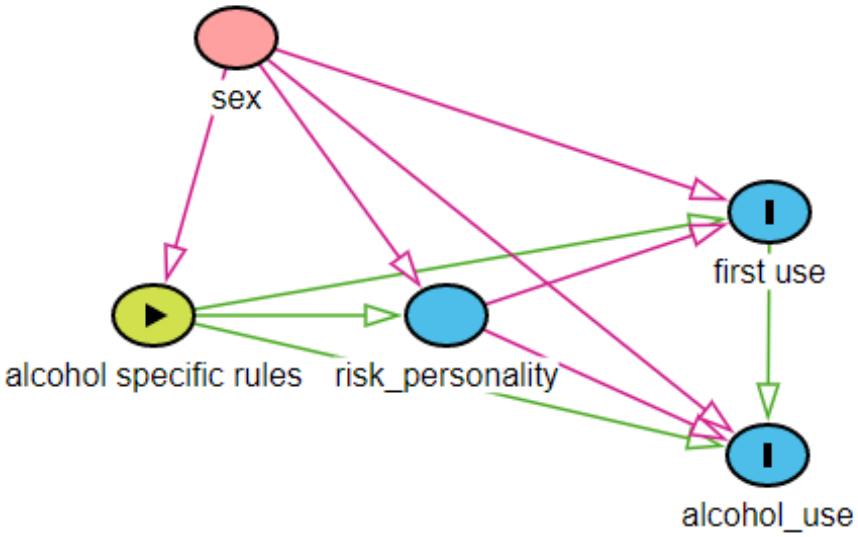
14. Hanewinkel. 2008. Longitudinal study of parental movie restriction on teen smoking and drinking in Germany.



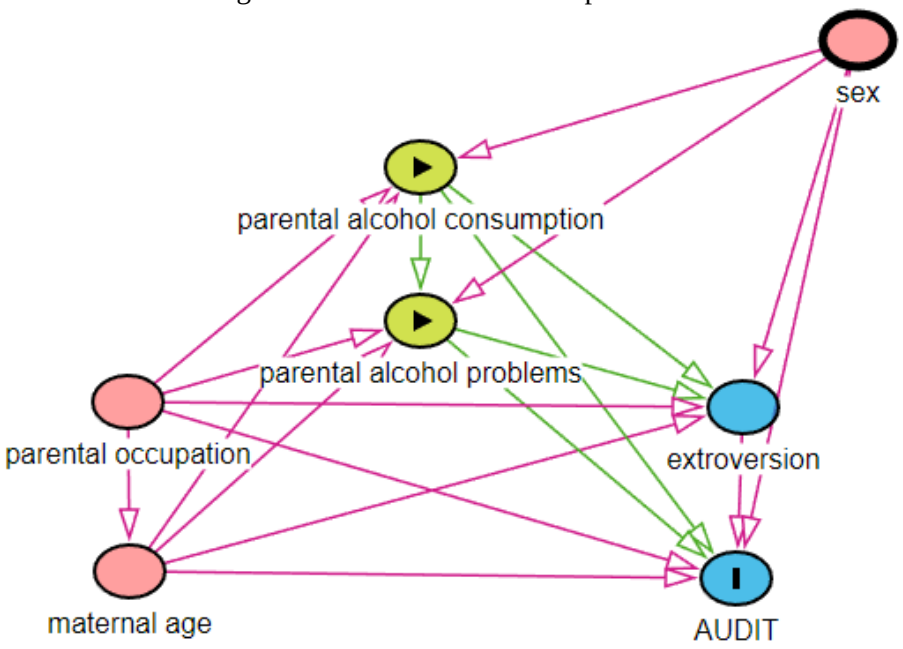
15. Hemphill et al. 2011. Risk and protective factors for adolescent substance use in Washington State, United States and Victoria, Australia: A longitudinal study.



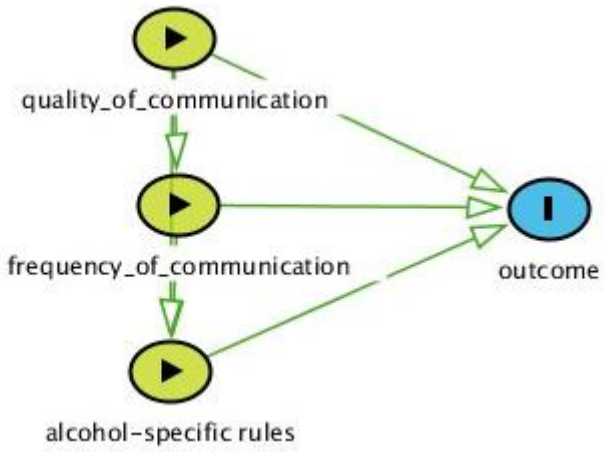
16. Janssen et al. 2014. Interactions between parental alcohol-specific rules and risk personalities in the prediction of adolescent alcohol use.



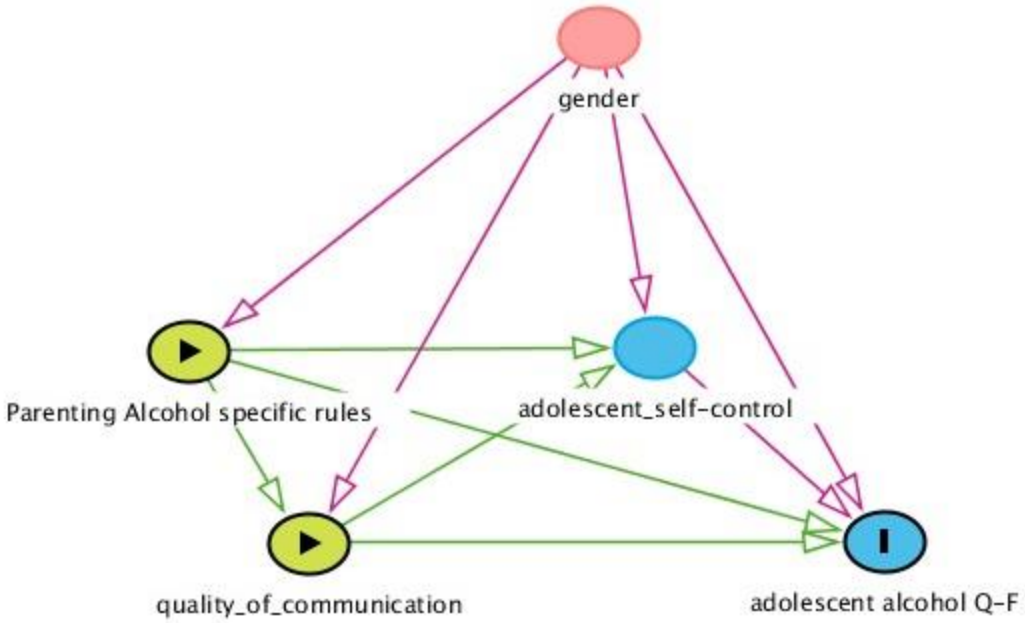
17. Kendler et al. 2013. Dimensions of parental alcohol use/problems and offspring temperament, externalizing behaviours and alcohol use/problems.



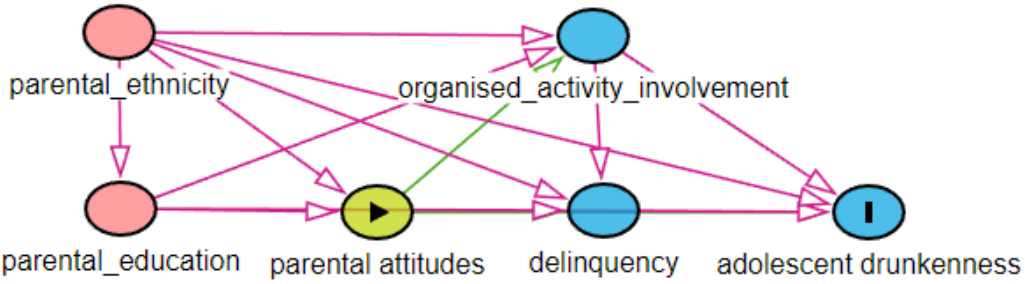
18. Koning et al. 2011. Developmental alcohol-specific parenting profiles in adolescence and their relationships with adolescents' alcohol use.



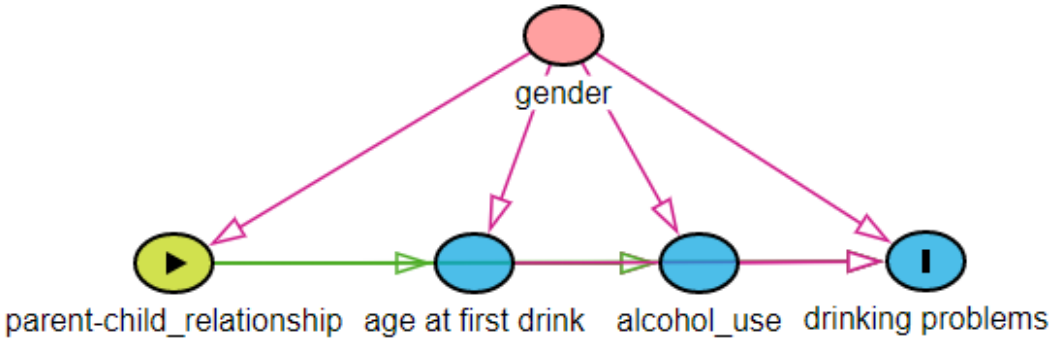
19. Koning et al. 2013. Alcohol-specific parenting, adolescents' self-control, and alcohol use: A moderated mediation model.



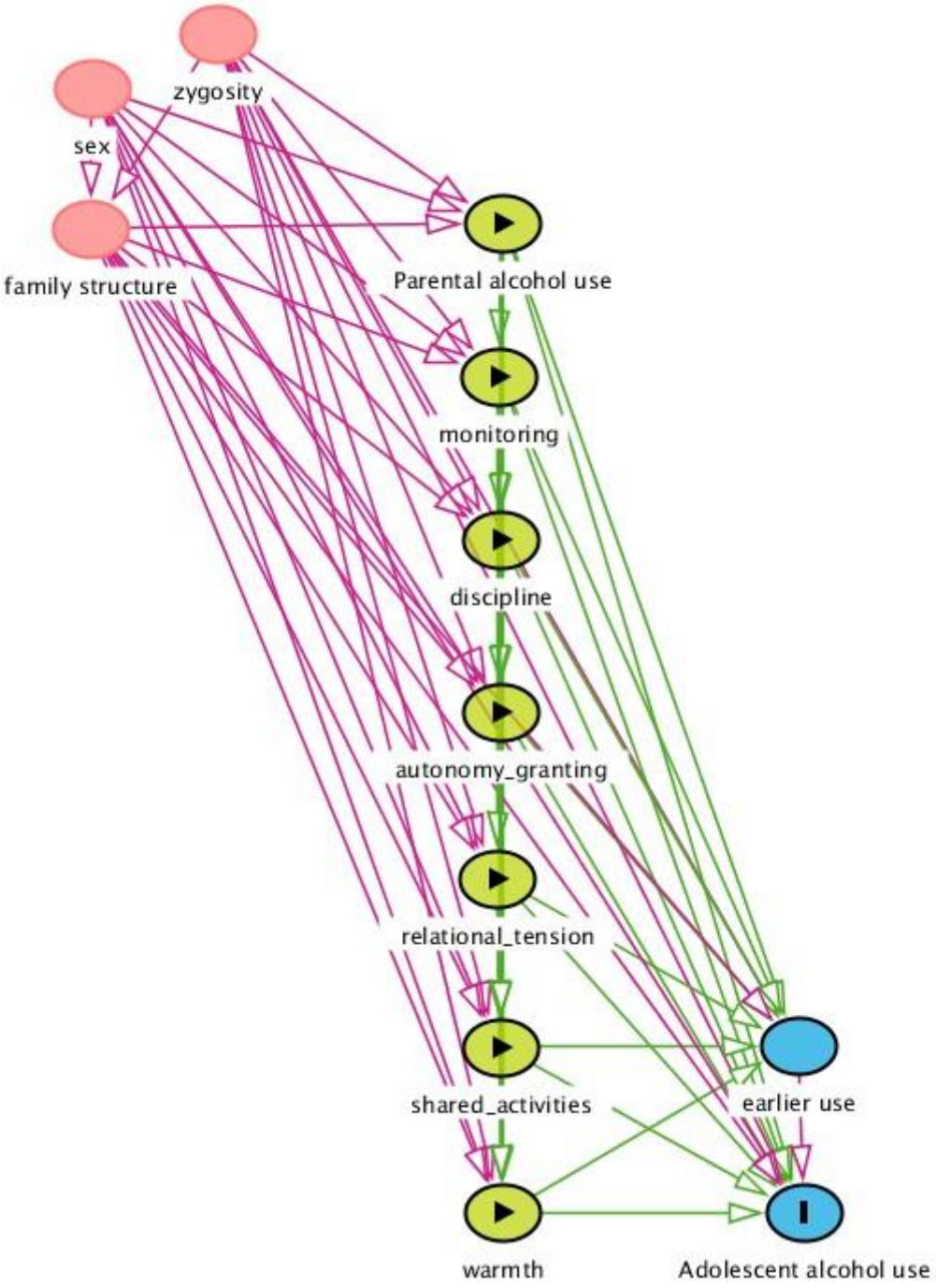
20. Koutakis et al. 2008. Reducing youth alcohol drinking through a parent-targeted intervention: the Örebro Prevention Program.



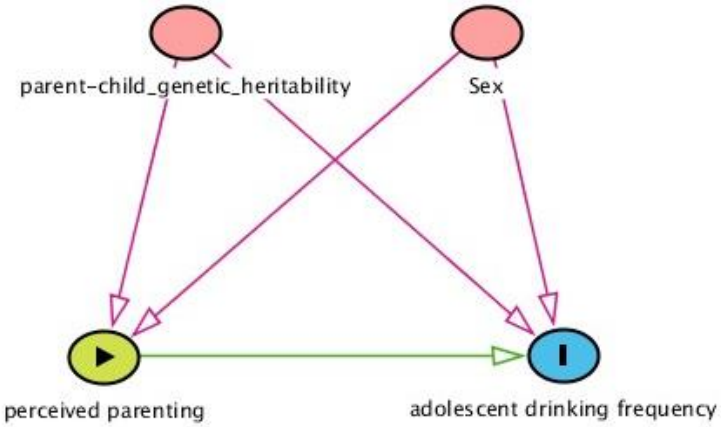
21. Kuntsche et al. 2009. The earlier the more? Differences in the links between age at first drink and adolescent alcohol use and related problems according to quality of parent–child relationships.



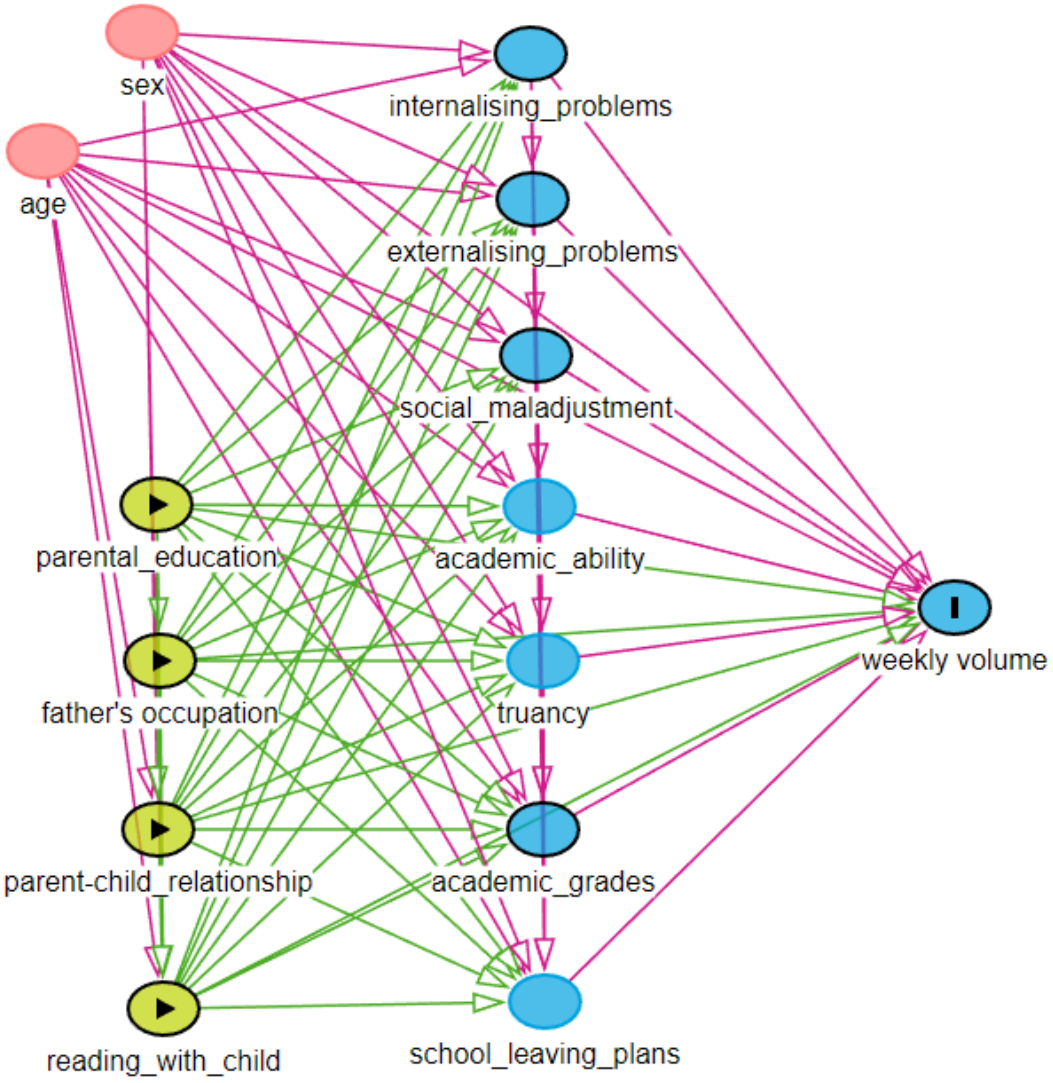
22. Latendresse et al. 2008. Parenting mechanisms in links between parents' and adolescents' alcohol use behaviours.



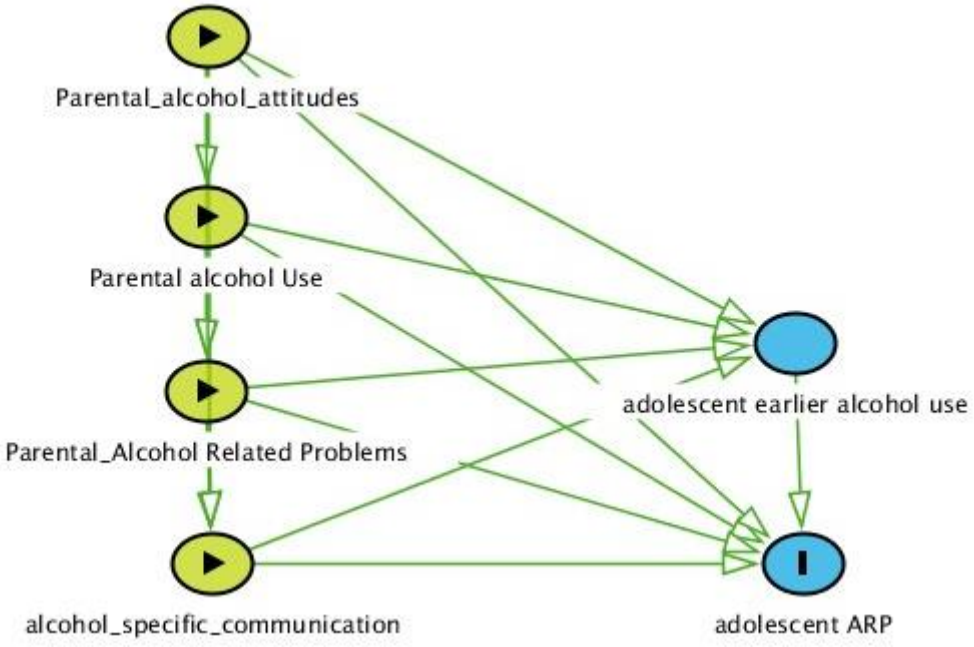
23. Latendresse et al. 2010. Examining the etiology of associations between perceived parenting and adolescents' alcohol use: common genetic and/or environmental liabilities?



24. Maggs et al. 2007. Childhood and adolescent predictors of alcohol use and problems in adolescence and adulthood in the National Child Development Study.

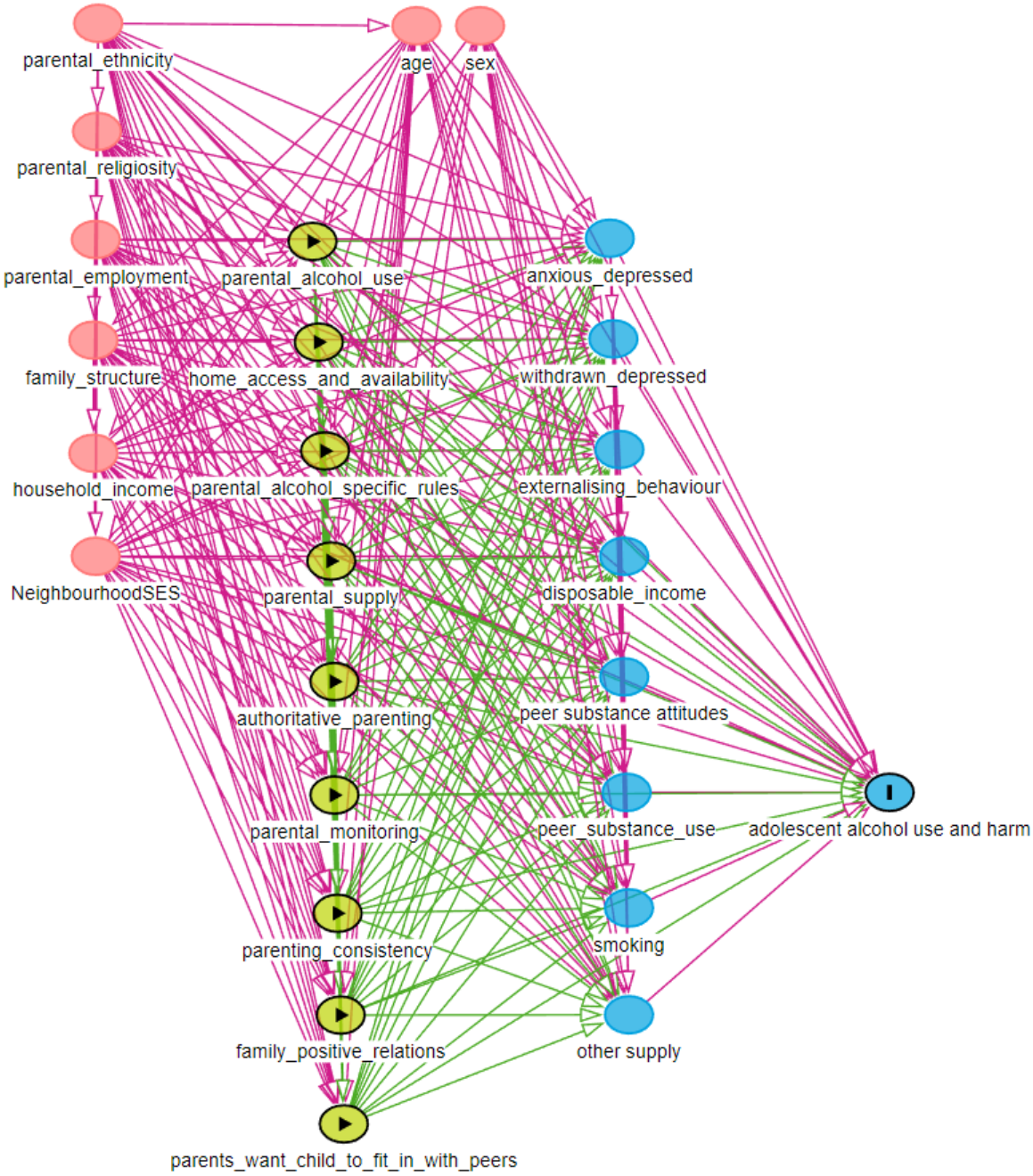


25. Mares et al. 2011. Parental alcohol use, alcohol-related problems, and alcohol-specific attitudes, alcohol-specific communication, and adolescent excessive alcohol use and alcohol-related problems: An indirect path model.

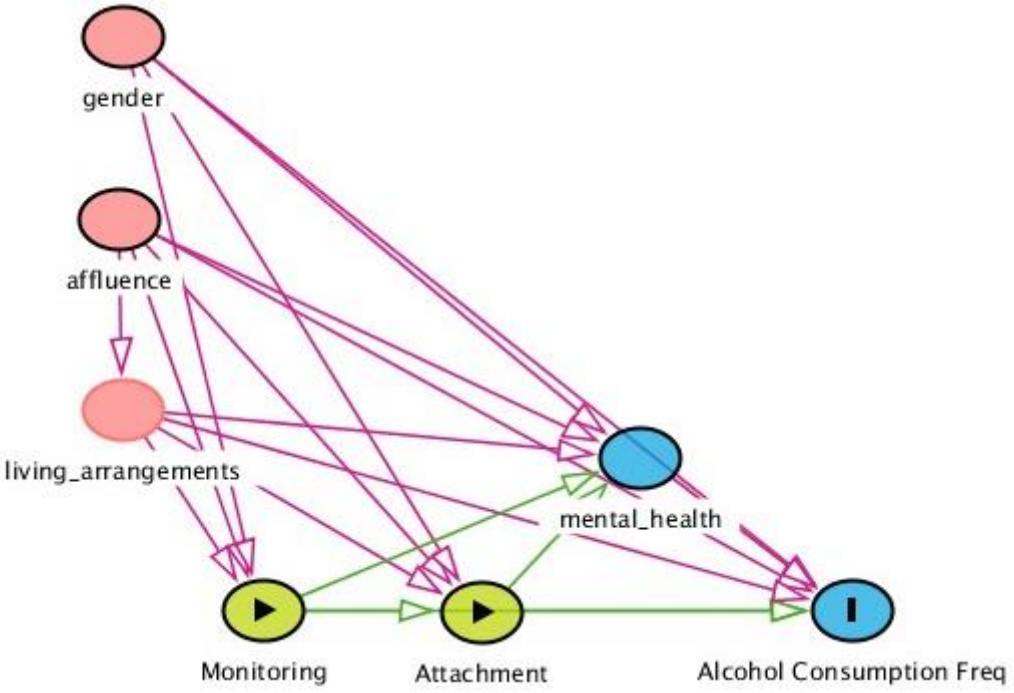


ARP = alcohol related problems

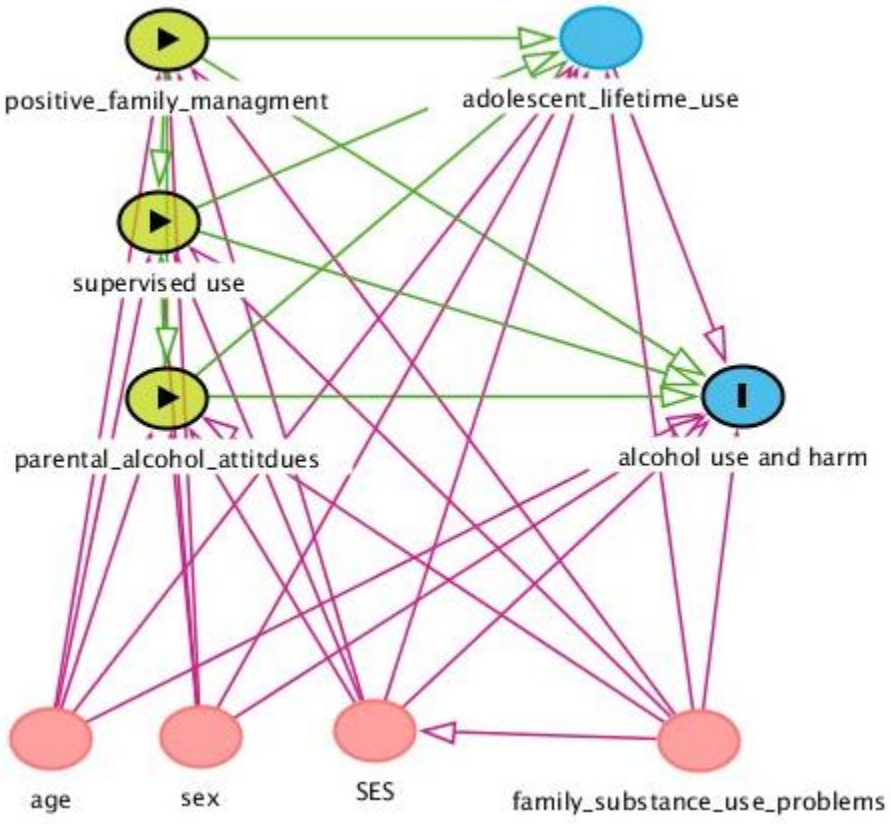
26. Mattick et al. 2018. Association of parental supply of alcohol with adolescent drinking, alcohol-related harms, and alcohol use disorder symptoms: a prospective cohort study.



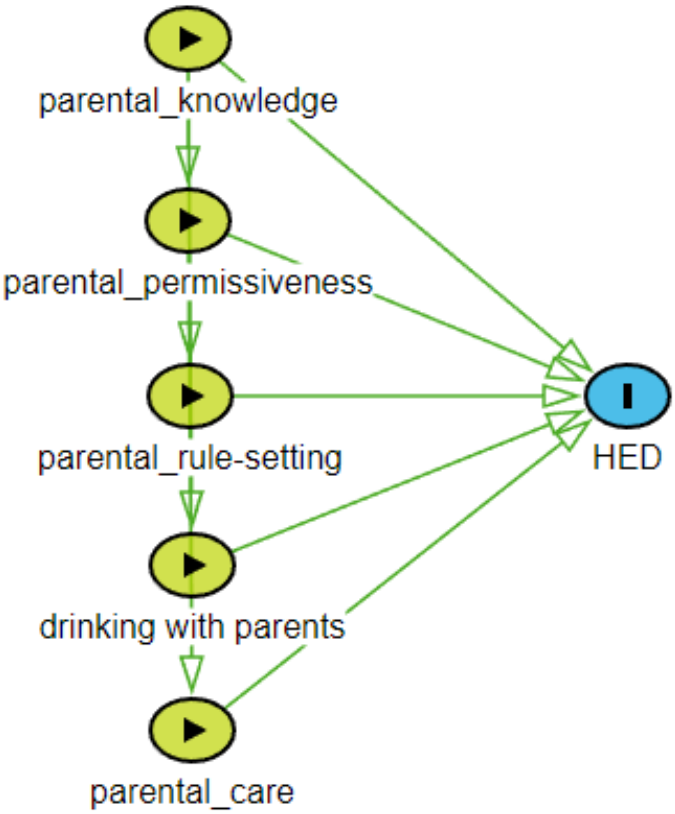
27. McCann et al. 2016. Assessing elements of a family approach to reduce adolescent drinking frequency: parent–adolescent relationship, knowledge management and keeping secrets.



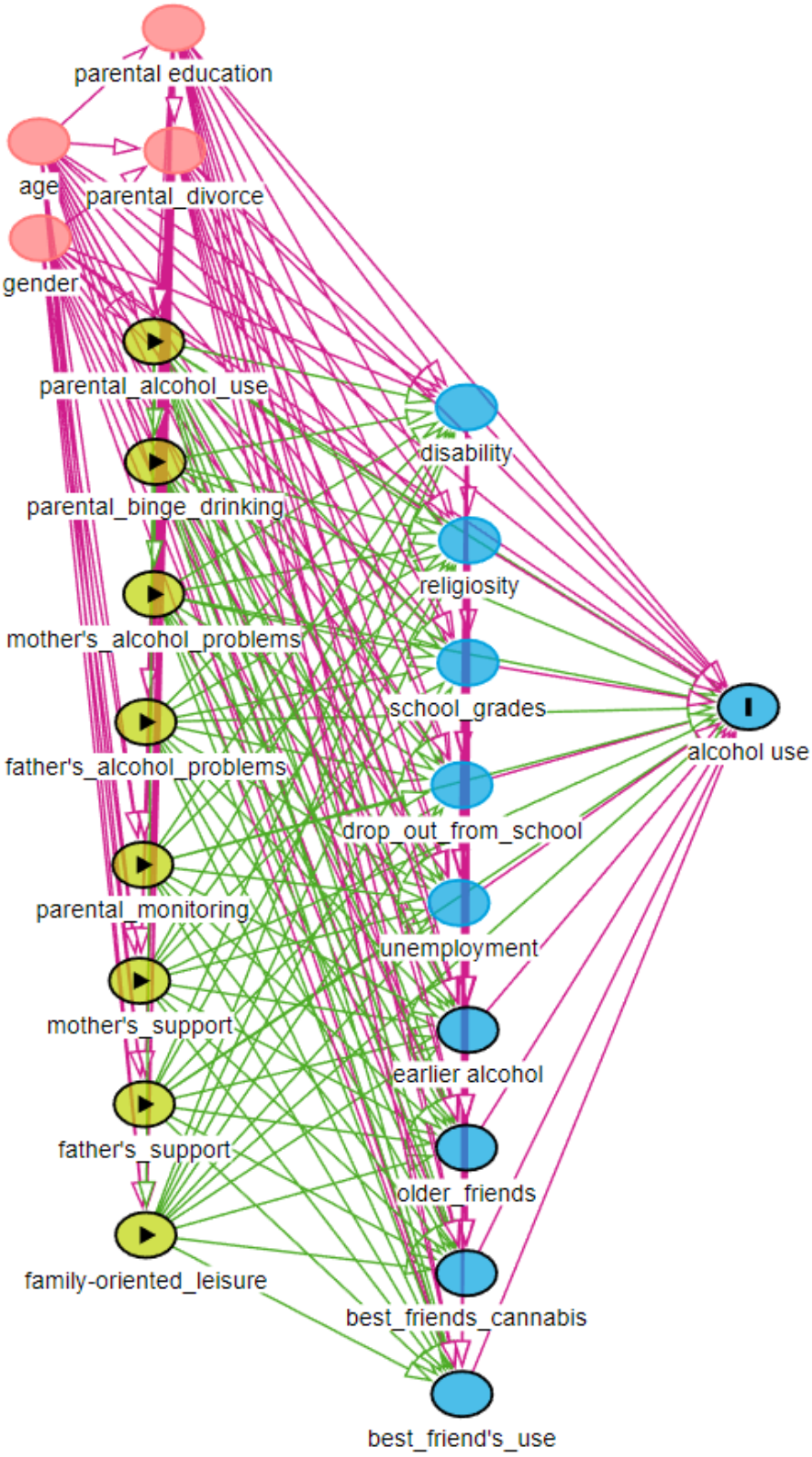
28. McMorris et al. 2011. Influence of family factors and supervised alcohol use on adolescent alcohol use and harms: similarities between youth in different alcohol policy contexts.



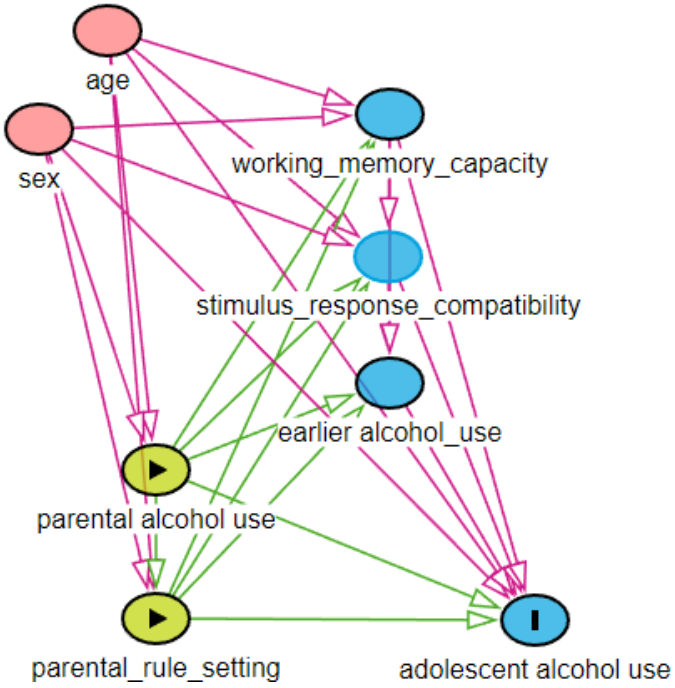
29. Pape et al. 2017. Drinking with parents: Different measures, different associations with underage heavy drinking?



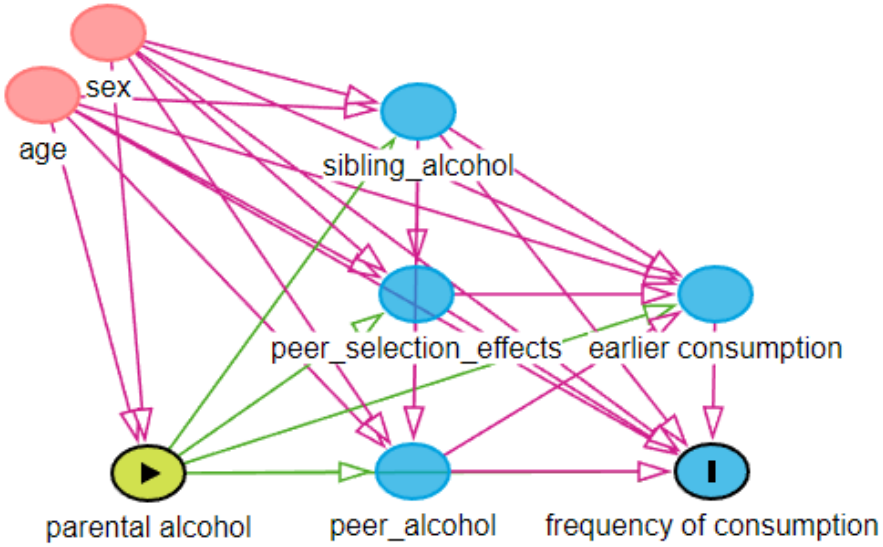
30. Pedersen. 2013. Socialization to binge drinking: A population-based, longitudinal study with emphasis on parental influences.



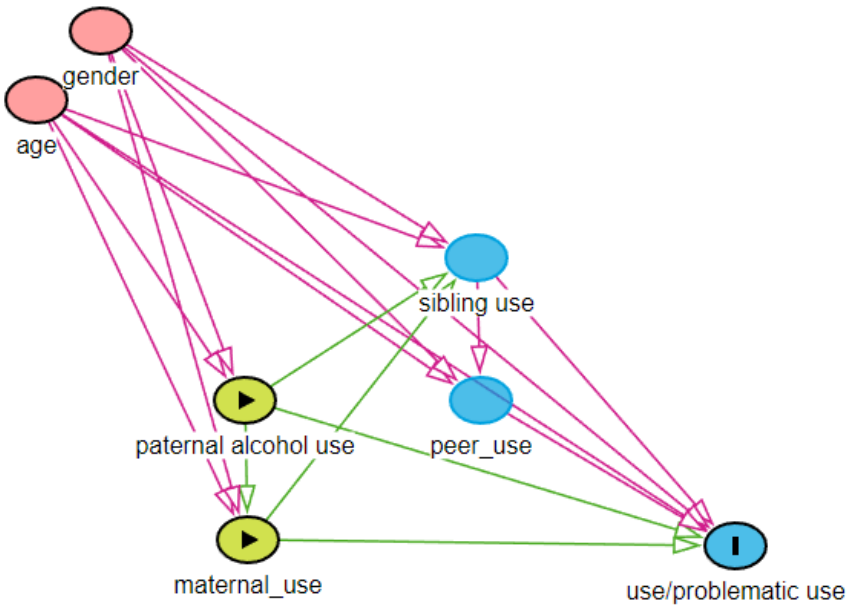
31. Pieters et al. 2012. The moderating role of working memory capacity and alcohol-specific rule-setting on the relation between approach tendencies and alcohol use in young adolescents.



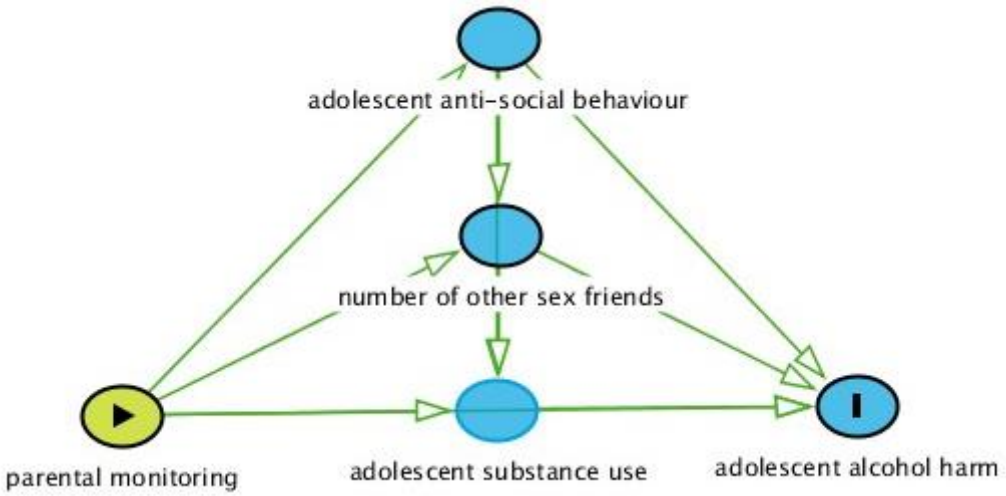
32. Poelen et al. 2007. Drinking by parents, siblings, and friends as predictors of regular alcohol use in adolescents and young adults: A longitudinal twin-family study.



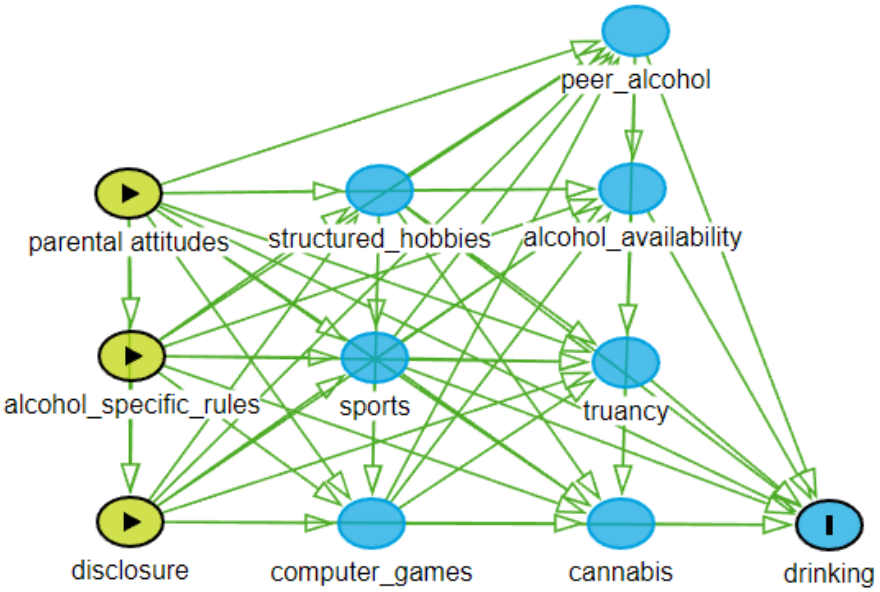
33. Poelen et al. 2009. Predictors of problem drinking in adolescence and young adulthood: A longitudinal twin-family study.



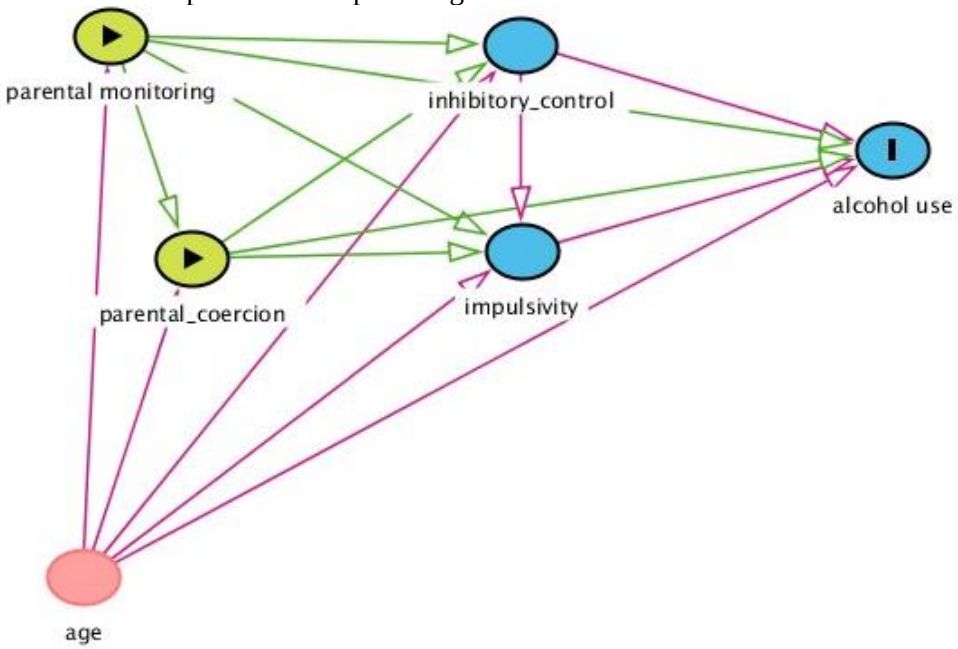
34. Poulin and Denault. 2012. Other-sex friendships as a mediator between parental monitoring and substance use in girls and boys.



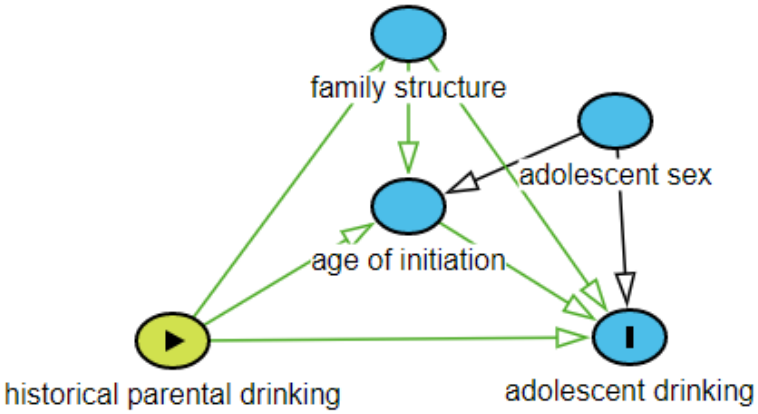
35. Raninen et al. 2018. One explanation to rule them all? Identifying sub-groups of non-drinking Swedish ninth graders.



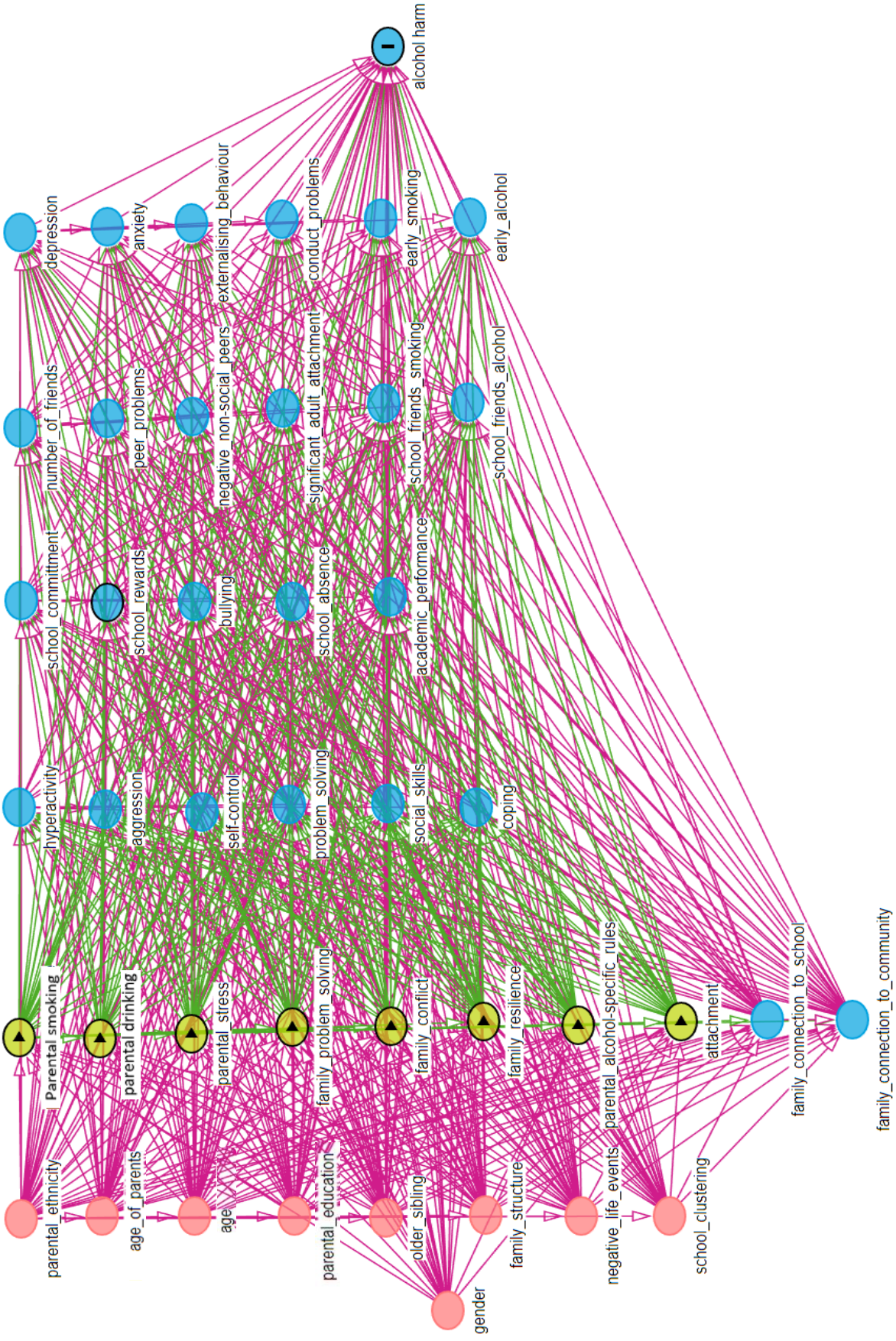
36. Rioux et al. 2016. Differential susceptibility to environmental influences: interactions between child temperament and parenting in adolescent alcohol use.



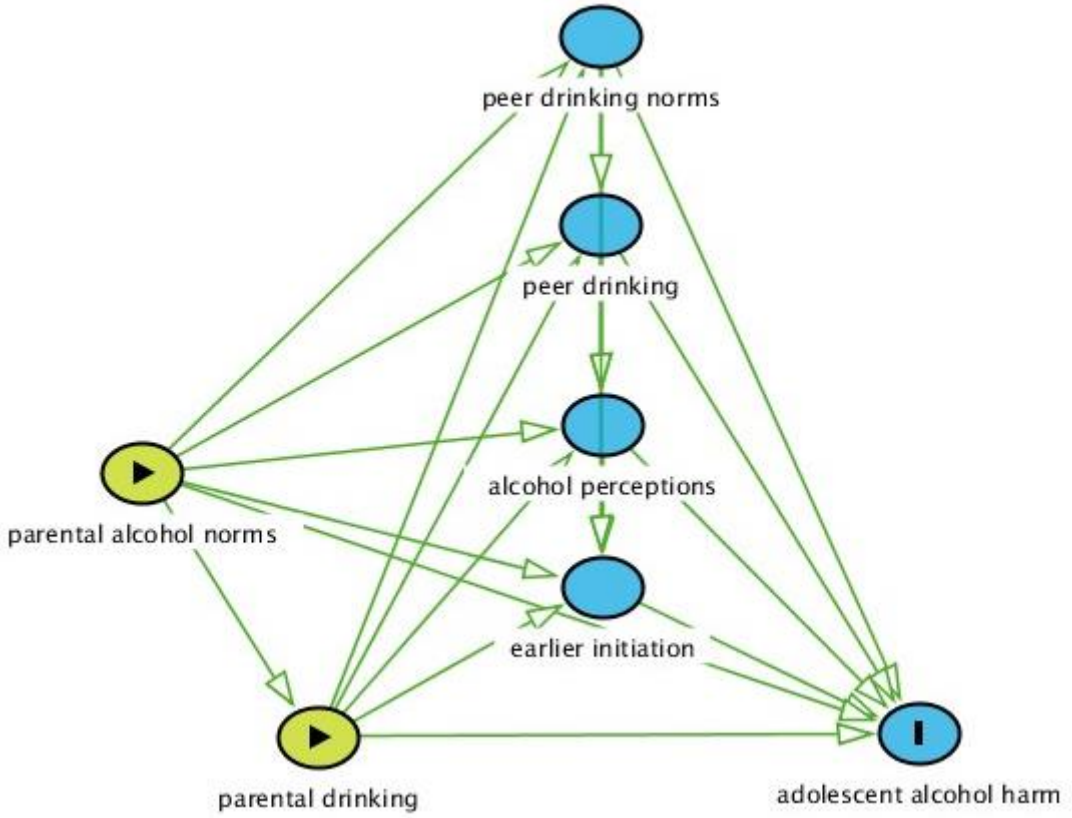
37. Seljamo et al. 2006. Alcohol use in families: a 15-year prospective follow-up study.



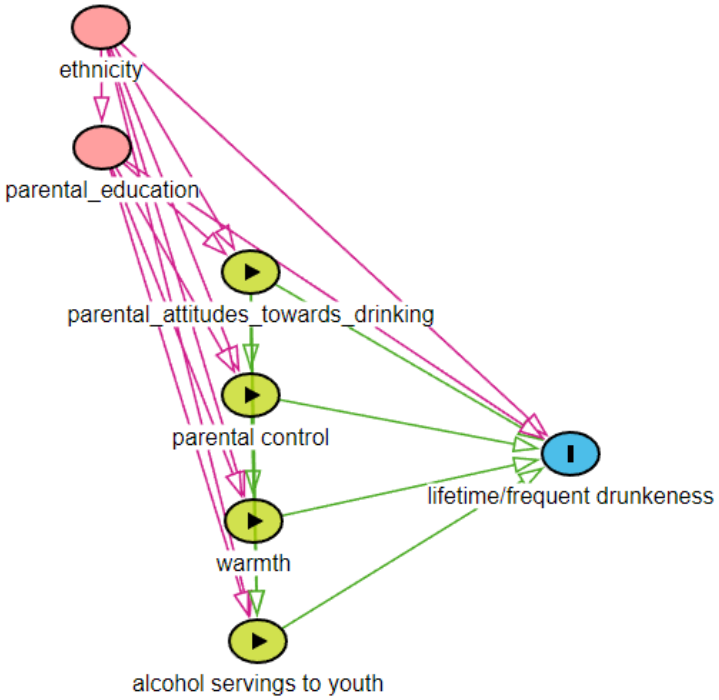
38. Shortt et al. 2007. Family, school, peer and individual influences on early adolescent alcohol use: first-year impact of the Resilient Families programme.



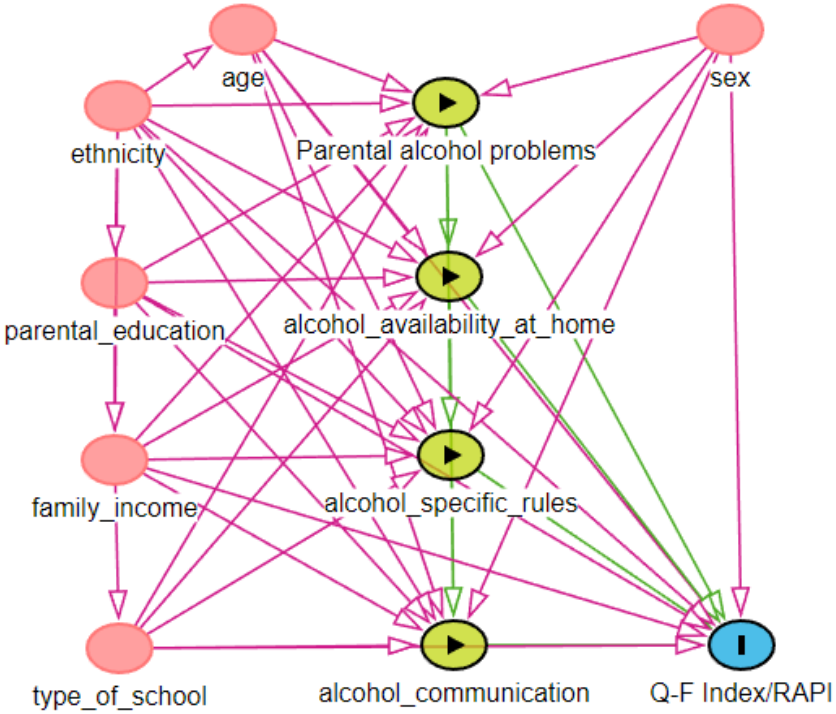
39. Spijkerman et al. 2007. The impact of peer and parental norms and behaviour on adolescent drinking: The role of drinker prototypes.



40. Strandberg. 2014. Gender differences in the prediction of parental servings of alcohol to adolescents and youth drunkenness.

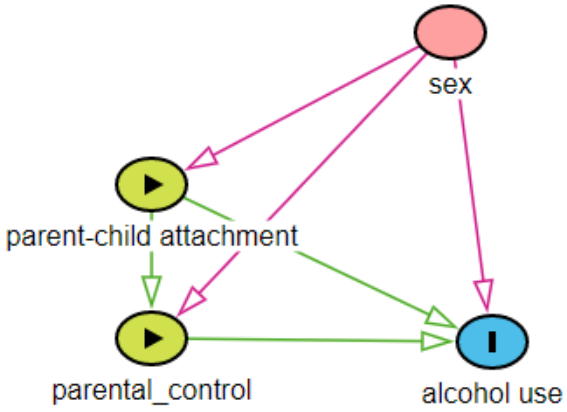


41. Van Eijnden et al. 2010. Alcohol-specific parenting and adolescents' alcohol-related problems: the interacting role of alcohol availability at home and parental rules.

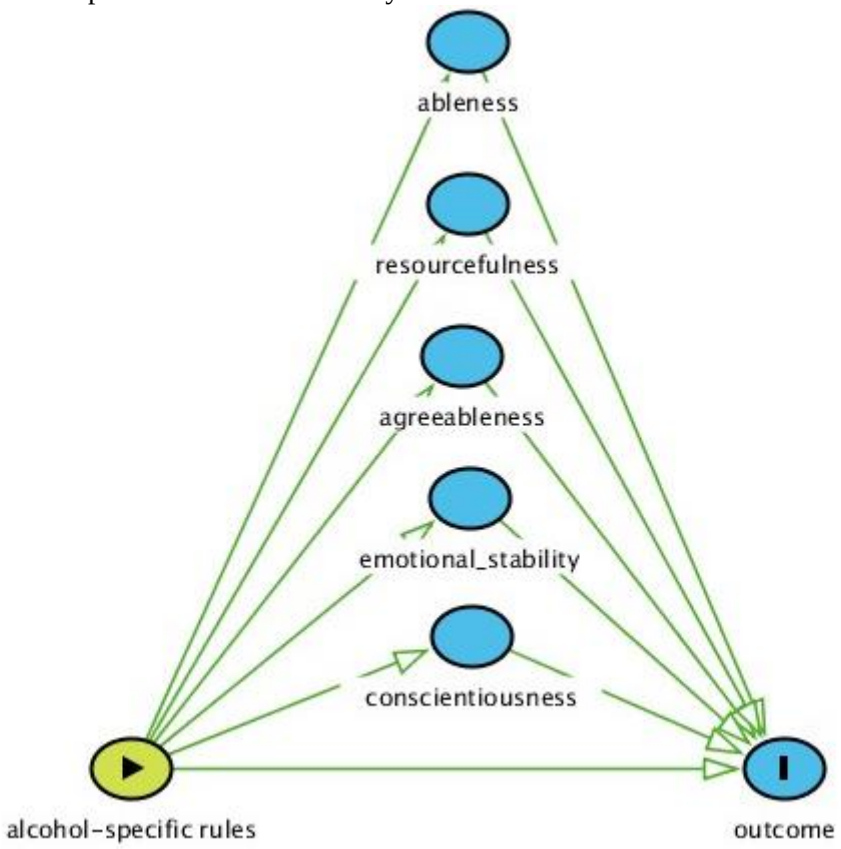


RAPI = Rutgers alcohol problems index

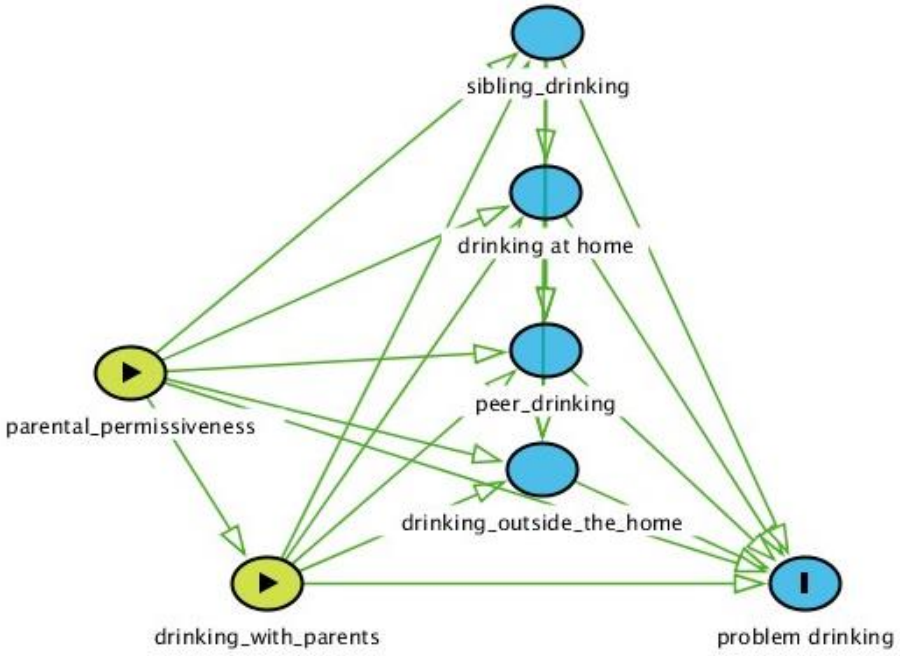
42. Van der Vorst. 2006. Parental attachment, parental control, and early development of alcohol use: a longitudinal study.



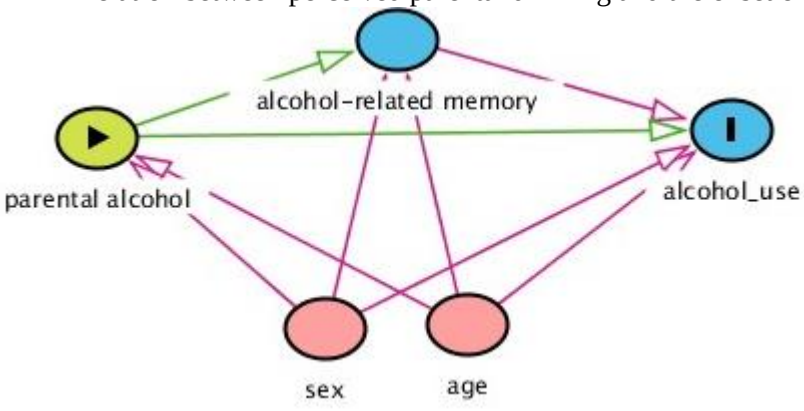
43. Van der Vorst et al. 2007. Alcohol-specific rules, personality and adolescents' alcohol use: a longitudinal person–environment study.



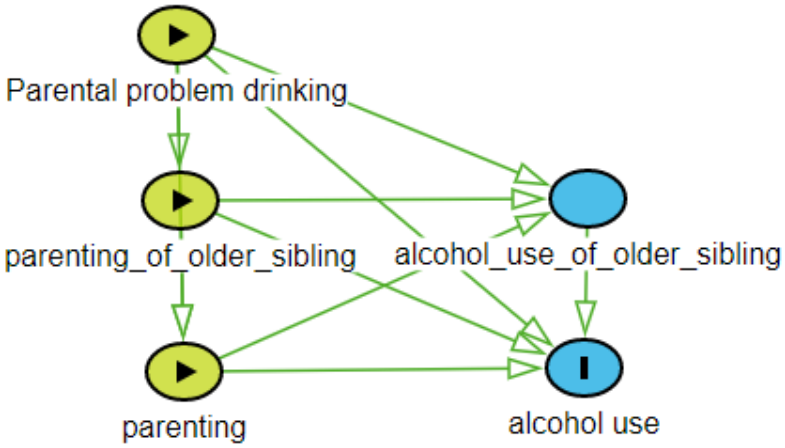
44. Van der Vorst et al. 2008. Do parents and best friends influence the normative increase in adolescents' alcohol use at home and outside the home?



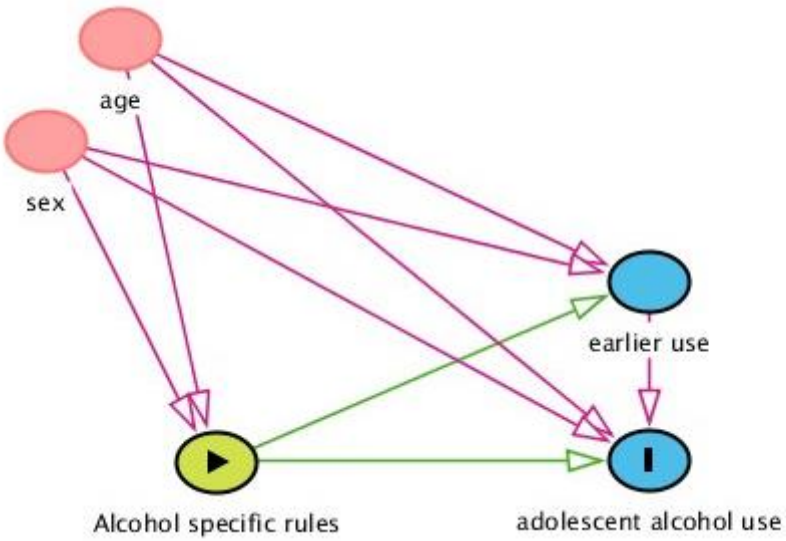
45. Van der Vorst et al. 2012. The mediating role of alcohol-related memory associations on the relation between perceived parental drinking and the onset of adolescent alcohol use.



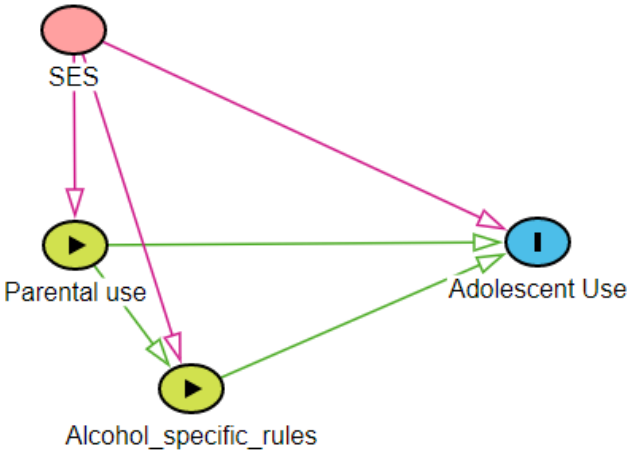
46. Van der Zwaluw. 2008. Parental problem drinking, parenting, and adolescent alcohol use.



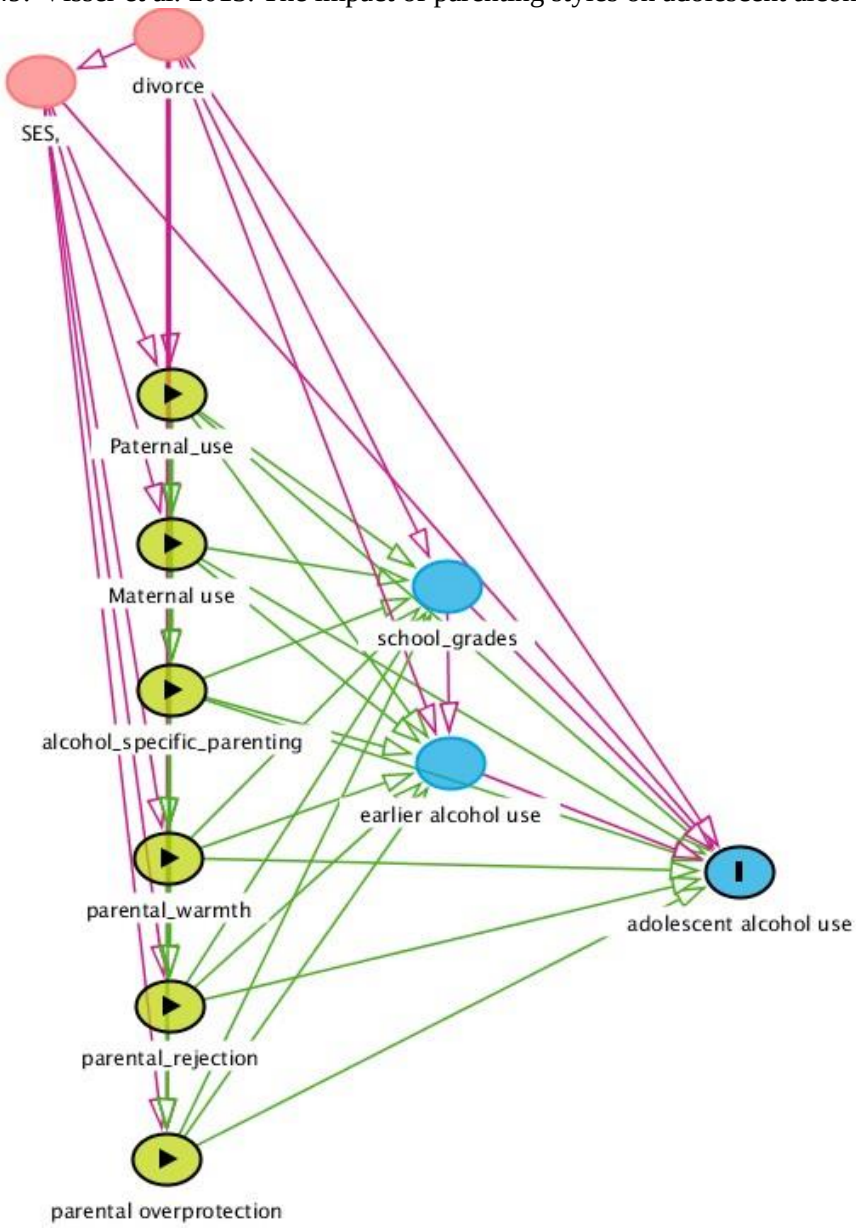
47. Van der Zwaluw. 2013. Different trajectories of adolescent alcohol use: testing gene–environment interactions.



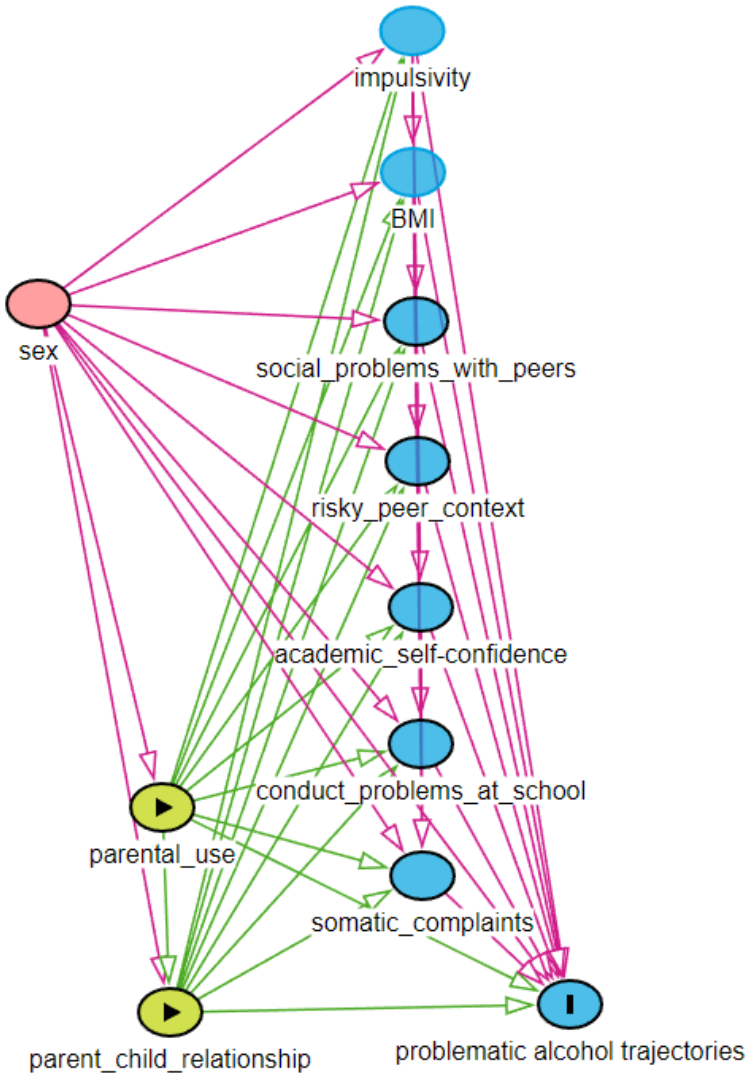
48. Vermeulen-Smit et al. 2012. The influence of paternal and maternal drinking patterns within two-partner families on the initiation and development of adolescent drinking.



49. Visser et al. 2013. The impact of parenting styles on adolescent alcohol use: The TRAILS Study.



50. Weichold et al. 2014. Childhood predictors and mid-adolescent correlates of developmental trajectories of alcohol use among male and female youth.



E: ESC-DAGs development log

This appendix briefly describes how ESC-DAGs was developed. The decision to develop a method for building DAGs was conceived early on in the project. The initial interest in DAGs was presumably similar to most studies – they were to be used to inform original analysis of secondary data. However, given the relative lack of guidance mentioned above, it quickly became clear that the process of building a DAG is itself an interesting topic of research.

The core of the process – using causal criteria to translate individual studies into DAGs that are then synthesised into one – was conceived almost immediately (circa 2016). The subsequent detail of the ESC-DAGs processes, especially Translation, were developed in reference to the DAG principles discussed in Chapter 4. There are, of course, numerous other 'principles' pertaining to DAGs. These are divisible into two groups. The first are evidence synthesis principles, already discussed in Chapter 2 as too unwieldly for building DAGs with dozens or hundreds or relationships under investigation. The second are the mechanistic principles of DAGs such as d-separation, the backdoor criterion, conditional independencies, etc. This group of principles are, in general, not directly concerned with how DAGs are built. Rather, they pertain to how DAGs function *after* variables and directed edges have been determined. Note, however, that some principles concerning DAG mechanisms can be read as contradicting the principle that DAGs should be built independently of data. For example, conditional independency implies that there is no partial correlation between variables (i.e. in the data) without directed edges between them. It is possible to test this using the DAGitty-R software. However, this preference for data-driven decision-making was rejected from the ESC-DAGs protocol, which emphasises wider evidence over available data in the DAG-building process. It is also worth noting that a potential compromise and avenue for further investigation could be to contrast ESC-DAGs data I-DAGS with the corresponding DAGitty-R DAG. There are currently no plans for doing so, however.

ESC-DAGs was originally piloted on a systematic review of neighbourhood influences on adolescent alcohol harm by Jackson et al (2014). Figure E1 below was a poster presented at the 2017 UK Causal Inference Meeting at the University of Exeter when ESC-DAGs was still being piloted on this review. The piloting process was valuable in demonstrating inefficiencies in ESC-DAGs itself, with implications for feasibility of conducting an ESC-DAGs review and original data analysis. In other words, the piloting process highlighted opportunities to make ESC-DAGs more efficient by either changing inefficient processes or removing those that may have been beyond the scope of the PhD research. Several moderations are worth highlighting. First, the differentiation from the full version (i.e. full saturation) to the rapid version (i.e. only saturating when more relevant) as discussed above.

Second, the original process used the DAGs produced from each study as the unit of synthesis by sequentially superimposing them, rather than simply adding the individual directed edges (the latter was found to be much faster and less cumbersome). Third, a bespoke risk of bias tool was used to evaluate the 'strength' of each study. This was dropped as it was not only inferior to other extant tools, but it also made the overall process too unwieldy to be conducted within the scope of this PhD research. Fourth, ESC-DAGs originally placed a much greater emphasis on commenting on modelling practices in the literature and how they may influence published effect estimates, in that instances of variables being assigned incorrectly in statistical models were to be quantified and interpreted. This was left out of the version of ESC-DAGs used in this thesis for similar reasons pertaining to scope (for example, to allow for greater emphasis on original analysis of an original DAG rather than critical evaluation of modelling practices in the literature).

Formal criteria for success were not used as a development tool, but certain milestones were identified as important indicators of the protocol's merit. First, opportunities for dissemination to the research community (conferences and publication) were utilised both to engage with feedback processes (such as reviewer comments), and to informally assess overall appetite for a DAG-building process. Second, the conceptual and data integrated DAGs themselves were important milestones in that, if they were inadequate (itself an uncertain concept), this would have indicated important flaws. As it was, the only issue identified at this stage was that such complex DAGs may be off-putting. Chapter 7 demonstrated a solution. Finally, to borrow the counterfactual heuristic, if the DAGs had not improved statistical modelling over a situation in which ESC-DAGs *was not used* then this would have indicated the project was not making a worthwhile contribution beyond self-falsification (which is not the same as failure). As noted in Chapter 10, while it is not possible to make this comparison, the DAGs produced were much more comprehensive than any models used in the primary studies, which was taken as an indicator that ESC-DAGs may improve modelling practices.

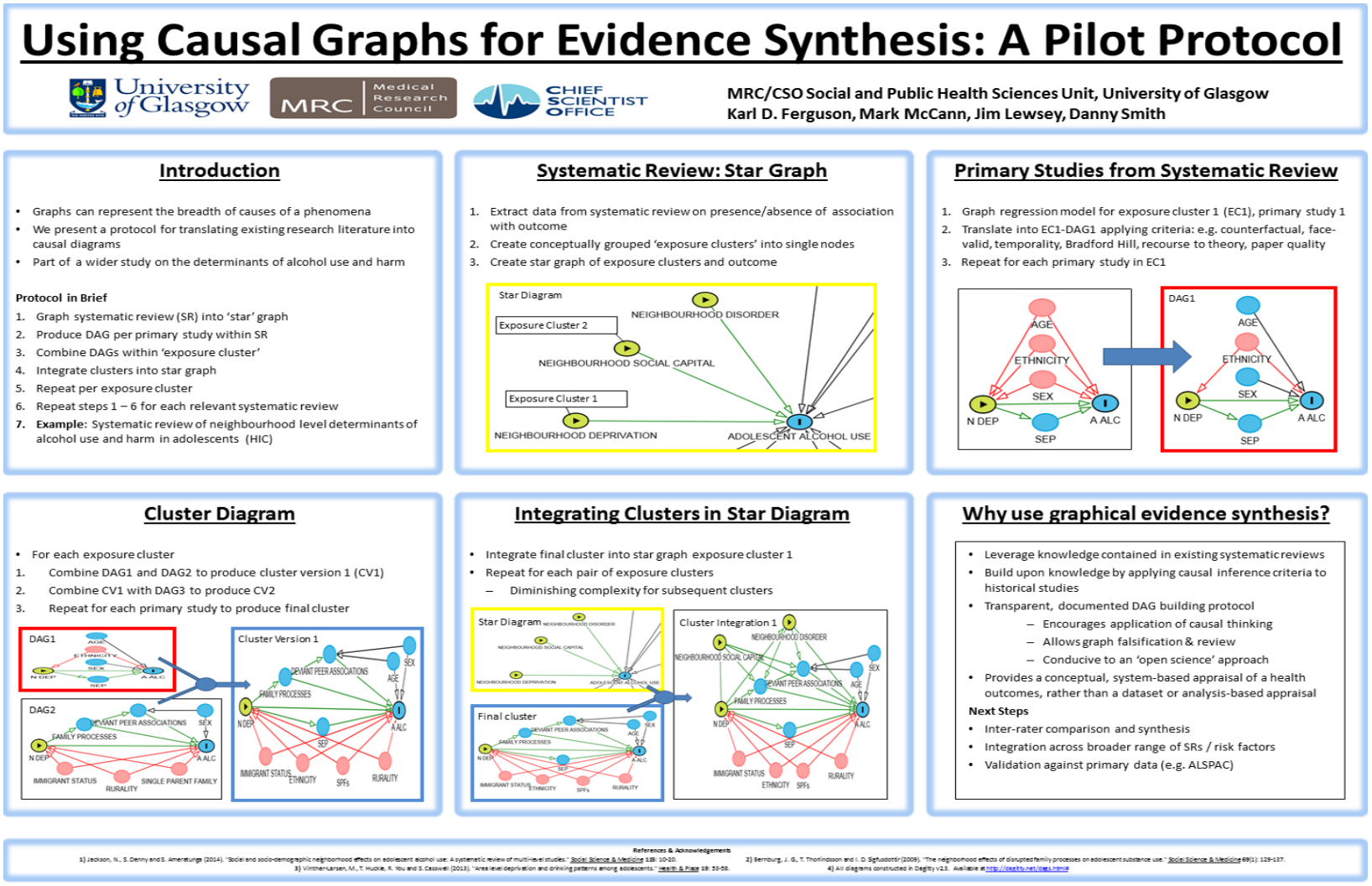


Figure E1: Early version of ESC-DAGs working with neighbourhood exposures

F: Example search strategy for review of systematic reviews

This appendix demonstrates the search strategy from the review of systematic reviews in Chapter 5 as applied to the PubMed database on the 6th April 2018. The search terms at the highest level were:

- Population: Adolescent
- Exposure: Parental influences
- Outcome: Adolescent drinking
- Design 1: Longitudinal
- Design 2: Systematic review

The search strategy used in PubMed (emphasis added) was:

> ((((((adolescent OR youth OR student OR teen* OR underage OR minor)) **AND** (parent-child relations OR parental influences OR parenting OR parental provision OR social hosting OR parental source of alcohol OR social supply)) **AND** (alcoholism OR underage drinking OR risky drinking OR excessive drinking OR binge drinking)) **AND** (longitudinal studies OR cohort OR prospective)) **AND** (systematic reviews OR review OR meta-analysis))

This search found 18 records (21 as of February 2020). The restriction to systematic reviews of longitudinal studies was narrow enough that the number of returns was low across all databases searched (i.e. Web of Science returned 25 and PsychInfo 28). For this reason, no further exclusion criteria were used (e.g. geography, language, date, etc.). The below table summarises the screening process applied to the results from PubMed. Five of the six systematic reviews included in the review of systematic reviews were present in the PubMed search results. The other 13 records were excluded for the reasons stipulated below.

Table F1: Screening process as applied to PubMed search results

Review	**Exclusion stage**	**Reason**	**Detail**
Sharmin et al, 2017	-	-	-
Sharmin et al, 2017	-	-	-
Yap et al, 2017	-	-	-
Rossow et al, 2016	-	-	-
Ryan et al, 2010	-	-	-
Wlodarczyk et al, 2017	Title	Wrong outcome	Mental health factors
Edwards et al, 2017	Title	Not a systematic review	Original study using survey data

Short et al, 1995	Title	Not a systematic review	Original evaluation of an intervention
Zucker & Gomberg, 1986	Title	Not a systematic review	Theoretical paper
Göktepe et al, 1975	Title	Wrong outcome	Sterotactic subcaudate tractotomy
Kaynak et al, 2014	Abstract	Not a systematic review	Review
Gilligan et al, 2012	Abstract	Not a systematic review	Narrative review
Enoch et al, 2006	Abstract	Not a systematic review	Theoretical paper
Jacob et al, 2001	Abstract	Not a systematic review	Original study using survey and other data
Flanagan and Kokotailo, 1999	Abstract	Wrong outcome	Adolescent pregnancy
Johnson and Leff, 1999	Abstract	Not a systematic review	Overview
Velleman et al, 1992	Abstract	Not a systematic review	Review
Hingson & White, 2014	Full text	Not a systematic review	Review

G: AUDIT questionnaire

Alcohol screening questionnaire (AUDIT)

One drink equals:  12 oz. beer  5 oz. wine  1.5 oz. liquor (one shot)

Scoring:	0 points per question	1 point per question	2 points per question	3 points per question	4 points per question
1. How often do you have a drink containing alcohol?	Never	Monthly or less	2 - 4 times a month	2 - 3 times a week	4 or more times a week
2. How many drinks containing alcohol do you have on a typical day when you are drinking?	0 - 2	3 or 4	5 or 6	7 - 9	10 or more
3. How often do you have four or more drinks on one occasion?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
4. How often during the last year have you found that you were not able to stop drinking once you had started?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
5. How often during the last year have you failed to do what was normally expected of you because of drinking?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
7. How often during the last year have you had a feeling of guilt or remorse after drinking?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
8. How often during the last year have you been unable to remember what happened the night before because of your drinking?	Never	Less than monthly	Monthly	Weekly	Daily or almost daily
9. Have you or someone else been injured because of your drinking?	No		Yes, but not in the last year		Yes, in the last year
10. Has a relative, friend, doctor, or other health care worker been concerned about your drinking or suggested you cut down?	No		Yes, but not in the last year		Yes, in the last year
Add the score for each column:		+	+	+	+

Total Score (add column scores) =

Image taken from https://www.omao.noaa.gov/find/media/images/alcohol-screening-questionnaire-audit. 23rd July 2019.

H: Monte Carlo error across imputed datasets

Monte Carlo error is defined as the "standard deviation across repeated runs of the same imputation procedure with the same data" (White et al., 2011). This standard deviation applies to any estimate. Table F1 below replicates Table 8-4 (page 172) from Chapter 8's analysis of the average causal effect (ACE) of parental influences on adolescent AUDIT score. An additional column is featured here with the Monte Carlo error for the ACEs and their 95% confidence intervals. The Monte Carlo error column thus represents the standard deviation for the estimate in question across the 50 imputed datasets. While not zero, Table H1 indicates that repeat analysis of data using the same imputation procedure would generate comparable results, as the Monte Carlo error was generally small in comparison to the estimates.

Table H1: Average causal effects of parental influences, including Monte Carlo simulation error

Parental variable	Age	X-sec control?	ACE (95% C.I.)	Monte Carlo error
Maternal drinking	12	No	1.08 (0.73, 1.42)	0.017 (0.021, 0.020)
		Yes	1.04 (0.70, 1.39)	0.017 (0.021, 0.021)
Maternal smoking	12	No	0.79 (0.28, 1.29)	0.026 (0.040, 0.035)
		Yes	0.71 (0.20, 1.22)	0.027 (0.041, 0.036)
Maternal depression	12	No	0.43 (0.05, 0.81)	0.019 (0.024, 0.024)
		Yes	0.56 (-0.0, 0.75)	0.019 (0.024, 0.024)
Parental permissiveness	12.5	No	1.67 (0.98, 2.36)	0.041 (0.060, 0.065)
		Yes	1.58 (0.85, 2.32)	0.041 (0.060, 0.064)
Secret-keeping	15.5	No	0.78 (0.36, 1.20)	0.024 (0.036, 0.048)
		Yes	0.57 (0.16, 0.98)	0.023 (0.035, 0.041)
Parental monitoring	15.5	No	1.41 (1.02, 1.79)	0.020 (0.02, 0.035)
		Yes	1.26 (0.87, 1.65)	0.021 (0.027, 0.036)
Time spent together	15.5	No	0.52 (0.17, 0.88)	0.019 (0.024, 0.032)
		Yes	0.30 (-0.06, 0.67)	0.025 (0.036, 0.049)
Parental punishment	15.5	No	0.68 (0.34, 1.02)	0.018 (0.029, 0.030)
		Yes	0.46 (0.09, 0.82)	0.020 (0.033, 0.030)
Parent-child tension	15.5	No	0.65 (0.26, 1.04)	0.021 (0.028, 0.027)
		Yes	0.31 (-0.15, 0.76)	0.025 (0.034, 0.031)

I: Expanded mediation results

In Chapter 9, the presentation of the wave 1 mediation results was restricted to the NIE and the number of models for which the confidence interval crossed the null. This appendix expands those results by including the TCE and NDE, as well as their 95% confidence intervals. It also includes the CDE (controlled direct effect) which was not included in Chapter 9 as it is not technically identified in the presence of EIMOCs, and nor does the difference between the TCE and CDE correspond to any indirect effect. It is included here for completeness as it is the final form of causal mediation estimate calculated using nested counterfactuals (as discussed in Chapter 3). Table I1 presents all parental influences that were tested as mediators of the effect of maternal drinking on adolescent AUDIT score, and Table I2 does the same for all non-parental intermediates.

Table I1: Parental mediators from wave 1 mediation models (20 datasets, 250 reps)

Parental Mediator		**Effect size**	**Lower C.I.**	**Upper C.I.**	**Count of models for which confidence intervals crossed null**
Maternal smoking	TCE	1.087	0.837	1.336	3/20
	NDE	1.048	0.799	1.297	
	NIE	0.039	0.016	0.061	
	% of TCE	3.6			
	CDE	1.050	0.800	1.299	
Maternal depression	TCE	1.086	0.838	1.355	20/20
	NDE	1.079	0.831	1.327	
	NIE	0.007	-0.004	0.019	
	% of TCE	0.6			
	CDE	1.079	0.832	1.327	
Parental permissiveness	TCE	1.047	0.799	1.295	13/20
	NDE	1.029	0.782	1.275	
	NIE	0.018	-0.007	0.044	
	% of TCE	1.7			
	CDE	1.027	0.781	1.273	
Parental monitoring	TCE	1.011	0.764	1.259	13/20

	NDE	1.076	0.776	1.273	
	NIE	-0.064	-0.147	0.018	
	% of TCE	5.9			
	CDE	1.035	0.796	1.274	
Parental monitoring with interaction term with maternal drinking	TCE	0.928	0.595	1.261	12/20
	NDE	0.988	0.658	1.318	
	NIE	-0.060	-0.138	0.018	
	% of TCE	5.4			
	CDE	1.107	0.841	1.372	
Parental punishment	TCE	1.068	0.822	1.315	0/20
	NDE	1.141	0.893	1.390	
	NIE	-0.073	-0.115	-0.031	
	% of TCE	6.8			
	CDE	1.132	0.886	1.378	
Relational tension	TCE	1.061	0.813	1.308	14/20
	NDE	1.100	0.853	1.346	
	NIE	-0.039	-0.084	0.007	
	% of TCE	3.7			
	CDE	1.094	0.851	1.338	
Parental tension with interaction term with maternal drinking	TCE	1.111	0.839	1.383	15/20
	NDE	1.156	0.881	1.432	
	NIE	-.045	-0.100	0.009	
	% of TCE	3.8			
	CDE	1.019	0.741	1.296	
Secret-keeping	TCE	1.043	0.795	1.290	12/20
	NDE	1.097	0.0853	1.341	
	NIE	-0.054	-0.114	0.006	
	% of TCE	4.9			
	CDE	1.071	0.831	1.311	

Parental involvement	TCE	1.063	0.816	1.309	15/20
	NDE	1.088	0.842	1.333	
	NIE	-0.025	-0.070	0.020	
	% of TCE	2.4			
	CDE	1.066	0.823	1.310	

Table I2: Non-parental mediators from wave 1 mediation models (20 datasets, 250 reps)

Non-Parental Mediator		**Effect size**	**Lower C.I.**	**Upper C.I.**	**Count of models for which confidence intervals crossed null**
Early alcohol initiation	TCE	1.047	0.799	1.294	16/20
	NDE	1.037	0.790	1.284	
	NIE	0.010	-0.013	0.033	
	% of TCE	1.0			
	CDE	1.038	0.791	1.285	
Sensation seeking	TCE	1.059	0.822	1.297	20/20
	NDE	1.053	0.814	1.291	
	NIE	0.007	-0.020	0.033	
	% of TCE	0.7			
	CDE	1.056	0.819	1.293	
Depression	TCE	1.068	0.830	1.306	20/20
	NDE	1.067	0.829	1.306	
	NIE	0.001	-0.011	0.013	
	% of TCE	0.1			
	CDE	1.071	0.833	1.309	
Anti-social behaviour	TCE	1.000	0.764	1.236	17/20
	NDE	0.954	0.725	1.183	
	NIE	0.046	-0.019	0.112	
	% of TCE	4.3			

	CDE	0.982	0.756	1.208	
Positive alcohol attitudes	TCE	1.090	0.838	1.343	0/20
	NDE	0.999	0.747	1.251	
	NIE	0.091	0.048	0.135	
	% of TCE	8.3			
	CDE	1.008	0.758	1.258	
Smoking	TCE	1.024	0.781	1.268	2/20
	NDE	0.846	0.612	1.080	
	NIE	0.178	0.077	0.279	
	% of TCE	17.4			
	CDE	0.869	0.642	1.096	
Substance use	TCE	1.061	0.815	1.306	3/20
	NDE	0.913	0.676	1.149	
	NIE	0.148	0.049	0.247	
	% of TCE	13.9			
	CDE	0.920	0.690	1.150	
Number of friends	TCE	1.084	0.837	1.311	15/20
	NDE	1.068	0.820	1.307	
	NIE	0.015	-0.003	0.034	
	% of TCE	1.4			
	CDE	1.069	0.821	1.317	
Peer smoking	TCE	1.040	0.791	1.288	1/20
	NDE	0.952	0.709	1.195	
	NIE	0.088	0.027	0.149	
	% of TCE	8.5			
	CDE	0.959	0.718	1.200	
Peer alcohol	TCE	1.054	0.800	1.308	0/20
	NDE	0.861	0.610	1.112	
	NIE	0.193	0.122	0.264	

	% of TCE	18.3			
	CDE	0.879	0.633	1.124	
Peer anti-social behaviour	TCE	0.993	0.757	1.228	10/20
	NDE	1.042	0.805	1.280	
	NIE	-0.050	-0.096	-0.003	
	% of TCE	4.8			
	CDE	1.022	0.787	1.257	
Peer substance use	TCE	0.916	0.681	1.152	20/20
	NDE	0.906	0.669	1.144	
	NIE	0.010	-0.028	0.048	
	% of TCE				
	CDE	0.899	0.663	1.135	
Positive relationships with teachers	TCE	1.070	0.827	1.314	10/20
	NDE	1.101	0.857	1.34	
	NIE	-0.030	-0.058	-0.002	
	% of TCE	2.8			
	CDE	1.091	0.848	1.334	

References

AALEN, O. O., ROYSLAND, K., GRAN, J. M., KOUYOS, R. & LANGE, T. 2016. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Stat Methods Med Res,* 25**,** 2294-2314.

AHRENS, K. A. & SCHISTERMAN, E. F. 2013. A time and place for causal inference methods in perinatal and paediatric epidemiology. *Paediatric and perinatal epidemiology,* 27**,** 258.

ALATI, R., BAKER, P., BETTS, K., CONNOR, J., LITTLE, K., SANSON, A. & OLSSON, C. 2014. The role of parental alcohol use, parental discipline and antisocial behaviour on adolescent drinking trajectories. *Drug and alcohol dependence,* 134**,** 178-184.

ALATI, R., CLAVARINO, A., NAJMAN, J. M., O'CALLAGHAN, M., BOR, W., AL MAMUN, A. & WILLIAMS, G. M. 2008. The developmental origin of adolescent alcohol use: Findings from the Mater University Study of Pregnancy and its outcomes. *Drug and Alcohol Dependence,* 98**,** 136-143.

ALATI, R., MALONEY, E., HUTCHINSON, D. M., NAJMAN, J. M., MATTICK, R. P., BOR, W. & WILLIAMS, G. M. 2010. Do maternal parenting practices predict problematic patterns of adolescent alcohol consumption? *Addiction,* 105**,** 872-880.

ALATI, R., NAJMAN, J. M., KINNER, S. A., MAMUN, A. A., WILLIAMS, G. M., O'CALLAGHAN, M. & BOR, W. 2005. Early Predictors of Adult Drinking: A Birth Cohort Study. *American Journal of Epidemiology,* 162**,** 1098-1107.

ANGUS, C., HOLMES, J., PRYCE, R., MEIER, P. & BRENNAN, A. 2016. Model-based appraisal of the comparative impact of Minimum Unit Pricing and taxation policies in Scotland: An adaptation of the Sheffield Alcohol Policy Model version 3. ScHARR, University of Sheffield.

AUSTIN, P. C. 2008. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in medicine,* 27**,** 2037-2049.

AUSTIN, P. C. 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine,* 28**,** 3083-3107.

AUSTIN, P. C. 2011. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res,* 46**,** 399-424.

BAGNARDI, V., ROTA, M., BOTTERI, E., TRAMACERE, I., ISLAMI, F., FEDIRKO, V., SCOTTI, L., JENAB, M., TURATI, F., PASQUALI, E., PELUCCHI, C., GALEONE, C., BELLOCCO, R., NEGRI, E., CORRAO, G., BOFFETTA, P. & LA VECCHIA, C. 2014. Alcohol consumption and site-specific cancer risk: a comprehensive dose–response meta-analysis. *British Journal Of Cancer,* 112**,** 580.

BANG, H. & ROBINS, J. M. 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics,* 61**,** 962-973.

BARON, R. M. & KENNY, D. A. 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol,* 51**,** 1173-82.

BERGE, J., SUNDELL, K., ÖJEHAGEN, A. & HÅKANSSON, A. 2016. Role of parenting styles in adolescent substance use: results from a Swedish longitudinal cohort study. *BMJ open,* 6**,** e008979.

BERO, L., CHARTRES, N., DIONG, J., FABBRI, A., GHERSI, D., LAM, J., LAU, A., MCDONALD, S., MINTZES, B. & SUTTON, P. 2018. The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures. *Systematic reviews,* 7**,** 242.

BIONDI-ZOCCAI, G., ROMAGNOLI, E., AGOSTONI, P., CAPODANNO, D., CASTAGNO, D., D'ASCENZO, F., SANGIORGI, G. & MODENA, M. G. 2011. Are propensity scores really superior to standard multivariable analysis? *Contemporary clinical trials,* 32**,** 731-740.

BISCHOF, G., REINHARDT, S., GROTHUES, J., DYBEK, I., MEYER, C., HAPKE, U., JOHN, U. & RUMPF, H. J. 2005. Effects of item sequence on the performance of the AUDIT in general practices. *Drug and Alcohol dependence,* 79**,** 373-377.

BLAKELY, T., LYNCH, J. & BENTLEY, R. 2016. Commentary: DAGs and the restricted potential outcomes approach are tools, not theories of causation. *Int J Epidemiol,* 45**,** 1835-1837.

BODIN, M. C. & STRANDBERG, A. K. 2011. The Örebro prevention programme revisited: a cluster-randomized effectiveness trial of programme effects on youth drinking. *Addiction,* 106**,** 2134-2143.

BRADLEY, K. A., BUSH, K. R., EPLER, A. J., DOBIE, D. J., DAVIS, T. M., SPORLEDER, J. L., MAYNARD, C., BURMAN, M. L. & KIVLAHAN, D. R. 2003. Two brief alcohol-screening tests From the Alcohol Use Disorders Identification Test (AUDIT): validation in a female Veterans Affairs patient population. *Archives of internal medicine,* 163**,** 821-829.

BRIERLEY-JONES, L., LING, J., MCCABE, K. E., WILSON, G. B., CROSLAND, A., KANER, E. F. & HAIGHTON, C. A. 2014. Habitus of home and traditional drinking: a qualitative analysis of reported middle-class alcohol use. *Sociology of health & illness,* 36**,** 1054-1076.

BRONFENBRENNER, U. 1979. *The Ecology of Human Development,* Cambridge, Harvard University Press.

BROOKHART, M. A., SCHNEEWEISS, S., ROTHMAN, K. J., GLYNN, R. J., AVORN, J. & STÜRMER, T. 2006. Variable selection for propensity score models. *American journal of epidemiology,* 163**,** 1149-1156.

BRYDEN, A., ROBERTS, B., MCKEE, M. & PETTICREW, M. 2012. A systematic review of the influence on alcohol use of community level availability and marketing of alcohol. *Health & Place,* 18**,** 349-357.

CABLE, N. & SACKER, A. 2008. Typologies of alcohol consumption in adolescence: predictors and adult outcomes. *Alcohol Alcohol,* 43**,** 81-90.

CALIENDO, M. & KOPEINIG, S. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys,* 22**,** 31-72.

CANADIAN CENTRE ON SUBSTANCE ABUSE. 2017. *The Impact and Effectiveness of Minimum Legal Drinking Age Legislation in Canada* [Online]. https://www.ccsa.ca/impact-and-effectiveness-minimum-legal-drinking-age-legislation-canada. [Accessed 29 July 2019].

CHAN, G. C., KELLY, A. B., TOUMBOUROU, J. W., HEMPHILL, S. A., YOUNG, R. M., HAYNES, M. A. & CATALANO, R. F. 2013. Predicting steep escalations in alcohol use over the teenage years: age-related variations in key social influences. *Addiction,* 108**,** 1924-1932.

COKER, J. K. & BORDERS, L. D. 2001. An analysis of environmental and social factors affecting adolescent problem drinking. *Journal of Counseling & Development,* 79**,** 200-208.

COLE, S. R. & HERNÁN, M. A. 2008. Constructing inverse probability weights for marginal structural models. *American journal of epidemiology,* 168**,** 656-664.

CULLEN, B. 2018. *Cognitive function in people with psychiatric and neurological disorders in UK Biobank.* University of Glasgow.

DANIEL, R. M., DE STAVOLA, B. L. & COUSENS, S. N. 2011. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *Stata Journal,* 11**,** 479-517.

DANIEL, R. M., DE STAVOLA, B. L. & VANSTEELANDT, S. 2016. Commentary: The formal approach to quantitative causal inference in epidemiology: misguided or misrepresented? *Int J Epidemiol,* 45**,** 1817-1829.

DANIELSSON, A. K., ROMELSJO, A. & TENGSTROM, A. 2011. Heavy episodic drinking in early adolescence: gender-specific risk and protective factors. *Subst Use Misuse,* 46**,** 633-43.

DAVEY SMITH, G. 2002. Commentary: Behind the Broad Street pump: aetiology, epidemiology and prevention of cholera in mid-19th century Britain. *International Journal of Epidemiology,* 31**,** 920-932.

DE MENESES-GAYA, C., ZUARDI, A. W., LOUREIRO, S. R. & CRIPPA, J. A. S. 2009. Alcohol Use Disorders Identification Test (AUDIT): An updated systematic review of psychometric properties. *Psychology & Neuroscience,* 2**,** 83-97.

DE STAVOLA, B. L., DANIEL, R. M., PLOUBIDIS, G. B. & MICALI, N. 2015. Mediation analysis with intermediate confounding: structural equation modeling viewed through the causal inference lens. *Am J Epidemiol,* 181**,** 64-80.

DEGENHARDT, L., ROMANIUK, H., COFFEY, C., HALL, W. D., SWIFT, W., CARLIN, J. B., O'LOUGHLIN, C. & PATTON, G. C. 2015. Does the social context of early alcohol use affect risky drinking in adolescents? Prospective cohort study. *BMC public health,* 15**,** 1137.

DEGENHARDT, L., STOCKINGS, E., PATTON, G., HALL, W. D. & LYNSKEY, M. 2016. The increasing global health priority of substance use in young people. *The Lancet Psychiatry,* 3**,** 251-264.

DEPARTMENT OF HEALTH 2016. UK Chief Medical Officers' Alcohol Guidelines Review. Summary of the Proposed New Guidelines. Department of Health and Social Care London.

DICK, D. M., PAGAN, J. L., VIKEN, R., PURCELL, S., KAPRIO, J., PULKKINEN, L. & ROSE, R. J. 2007. Changing environmental influences on substance use across development. *Twin Res Hum Genet,* 10**,** 315-26.

DICKSON, D. J., LAURSEN, B., STATTIN, H. & KERR, M. 2015. Parental Supervision and Alcohol Abuse Among Adolescent Girls. *Pediatrics,* 136**,** 617-624.

DONALDSON, L. 2009. *Guidance on the consumption of alcohol by children and young people*, Department of health.

DUBE, S. R., ANDA, R. F., FELITTI, V. J., EDWARDS, V. J. & CROFT, J. B. 2002. Adverse childhood experiences and personal alcohol abuse as an adult. *Addictive behaviors,* 27**,** 713-725.

DUBE, S. R., MILLER, J. W., BROWN, D. W., GILES, W. H., FELITTI, V. J., DONG, M. & ANDA, R. F. 2006. Adverse childhood experiences and the association with ever using alcohol and initiating alcohol use during adolescence. *Journal of Adolescent Health,* 38**,** 444. e1-444. e10.

DUNN, G., EMSLEY, R., LIU, H., LANDAU, S., GREEN, J., WHITE, I. & PICKLES, A. 2015. Treatment effect mediation. *Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme.* NIHR Journals Library.

ELZE, M. C., GREGSON, J., BABER, U., WILLIAMSON, E., SARTORI, S., MEHRAN, R., NICHOLS, M., STONE, G. W. & POCOCK, S. J. 2017. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. *Journal of the American College of Cardiology,* 69**,** 345-357.

ENGELS, R. C., VERMULST, A. A., DUBAS, J. S., BOT, S. M. & GERRIS, J. 2005. Long-term effects of family functioning and child characteristics on problem drinking in young adulthood. *European Addiction Research,* 11**,** 32-37.

FAGGION, C. M. 2015. Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor. *BMC Medical Research Methodology,* 15**,** 63.

FASANELLI, F., BAGLIETTO, L., PONZI, E., GUIDA, F., CAMPANELLA, G., JOHANSSON, M., GRANKVIST, K., JOHANSSON, M., ASSUMMA, M. B., NACCARATI, A., CHADEAU-HYAM, M., ALA, U., FALTUS, C., KAAKS, R., RISCH, A., DE STAVOLA, B., HODGE, A., GILES, G. G., SOUTHEY, M. C., RELTON, C. L., HAYCOCK, P. C., LUND, E., POLIDORO, S., SANDANGER, T. M., SEVERI, G. & VINEIS, P. 2015. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nature Communications,* 6**,** 10192.

FAT, L. N., SHELTON, N. & CABLE, N. 2018. Investigating the growing trend of non-drinking among young people; analysis of repeated cross-sectional surveys in England 2005–2015. *BMC public health,* 18**,** 1090.

FERGUSON, K. D., LEWSEY, J., MCCANN, M. & SMITH, D. 2018. P12 How to build the 'right' directed acyclic graph (DAG): a systematic, transparent and accessible method for evidence synthesis. *Journal of Epidemiology and Community Health,* 72**,** A66.2-A66.

FERGUSON, K. D., MCCANN, M., KATIKIREDDI, S. V., THOMSON, H., GREEN, M. J., SMITH, D. J. & LEWSEY, J. D. 2019. Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. *International Journal of Epidemiology*.

FRASER, A., MACDONALD-WALLIS, C., TILLING, K., BOYD, A., GOLDING, J., DAVEY SMITH, G., HENDERSON, J., MACLEOD, J., MOLLOY, L., NESS, A., RING, S., NELSON, S. M. & LAWLOR, D. A. 2013. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol,* 42**,** 97-110.

GAO, M., ALLEBECK, P., MISHRA, G. D. & KOUPIL, I. 2019. Developmental origins of endometriosis: a Swedish cohort study. *J Epidemiol Community Health***,** jech-2018-211811.

GILETTA, M., SCHOLTE, R. H., PRINSTEIN, M. J., ENGELS, R. C., RABAGLIETTI, E. & BURK, W. J. 2012. Friendship context matters: examining the domain specificity of alcohol and depression socialization among adolescents. *J Abnorm Child Psychol,* 40**,** 1027-43.

GLASER, B., SHELTON, K. H. & VAN DEN BREE, M. B. 2010. The moderating role of close friends in the relationship between conduct problems and adolescent substance use. *J Adolesc Health,* 47**,** 35-42.

GLASS, T. A., GOODMAN, S. N., HERNAN, M. A. & SAMET, J. M. 2013. Causal inference in public health. *Annu Rev Public Health,* 34**,** 61-75.

GOLDING, PEMBREY, JONES & THE ALSPAC STUDY, T. 2001. ALSPAC-The Avon Longitudinal Study of Parents and Children. *Paediatric and Perinatal Epidemiology,* 15**,** 74-87.

GOLDING, J. 1989. Research Protocol. European longitudinal study of pregnancy and childhood (ELSPAC). *Paediatr Perinat Epidemiol,* 3**,** 460-469.

GOLDING, J. 1990. Children of the nineties. A longitudinal study of pregnancy and childhood based on the population of Avon (ALSPAC). *West Engl Med J,* 105**,** 80-2.

GRANT, M. J. & BOOTH, A. 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal,* 26**,** 91-108.

GREENLAND, S. 1996. Basic methods for sensitivity analysis of biases. *International journal of epidemiology,* 25**,** 1107-1116.

GREENLAND, S., PEARL, J. & ROBINS, J. M. 1999. Causal diagrams for epidemiologic research. *Epidemiology,* 10**,** 37-48.

GRISWOLD, M. G., FULLMAN, N., HAWLEY, C., ARIAN, N., ZIMSEN, S. R., 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**,** 1015-1035.

GUILAMO-RAMOS, V., JACCARD, J., TURRISI, R. & JOHANSSON, M. 2005. Parental and school correlates of binge drinking among middle school students. *American journal of public health,* 95**,** 894-899.

HAFEMAN, D. M. & SCHWARTZ, S. 2009. Opening the Black Box: a motivation for the assessment of mediation. *International Journal of Epidemiology,* 38**,** 838-845.

HALL, W. D., PATTON, G., STOCKINGS, E., WEIER, M., LYNSKEY, M., MORLEY, K. I. & DEGENHARDT, L. 2016. Why young people's substance use matters for global health. *The Lancet Psychiatry,* 3**,** 265-279.

HANEWINKEL, R., MORGENSTERN, M., TANSKI, S. E. & SARGENT, J. D. 2008. Longitudinal study of parental movie restriction on teen smoking and drinking in Germany. *Addiction,* 103**,** 1722-30.

HANLEY, P., CHAMBERS, B. & HASLAM, J. 2016. Reassessing RCTs as the 'gold standard': synergy not separatism in evaluation designs. *International Journal of Research & Method in Education,* 39**,** 287-298.

HANSEN, K., JOHNSON, J. & JOSHI, H. 2012. Millennium Cohort Study first, second, third and fourth surveys: a guide to the datasets. *London: Centre for Longitudinal Studies, Institute of Education*.

HANSEN, P. & AINSWORTH, F. 2007. Parent blaming in child protection and health settings: A matter for concern. *Children Australia,* 32**,** 29-35.
HAYDEN, J. A., CÔTÉ, P. & BOMBARDIER, C. 2006. Evaluation of the quality of prognosis studies in systematic reviews. *Annals of internal medicine,* 144**,** 427-437.
HEMPHILL, S. A., HEERDE, J. A., HERRENKOHL, T. I., PATTON, G. C., TOUMBOUROU, J. W. & CATALANO, R. F. 2011. Risk and Protective Factors for Adolescent Substance Use in Washington State, the United States and Victoria, Australia: A Longitudinal Study. *Journal of Adolescent Health,* 49**,** 312-320.
HERNÁN, M. A. 2005. Invited commentary: hypothetical interventions to define causal effects—afterthought or prerequisite? *American Journal of Epidemiology,* 162**,** 618-620.
HERNÁN, M. A. 2014. Invited commentary: agent-based models for causal inference—reweighting data and theory in epidemiology. *American journal of epidemiology,* 181**,** 103-105.
HERNÁN, M. A. 2018. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *American Journal of Public Health,* 108**,** 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**,** 758-764.
HERNÁN, M. A. & ROBINS, J. M. 2019. *Causal Inference,* Boca Raton, Chapman & Hall/CRC.
HERNÁN, M. A. & TAUBMAN, S. L. 2008. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International journal of obesity,* 32**,** S8-S14.
HIGGINS, J. P., ALTMAN, D. G., GØTZSCHE, P. C., JÜNI, P., MOHER, D., OXMAN, A. D., SAVOVIĆ, J., SCHULZ, K. F., WEEKS, L. & STERNE, J. A. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj,* 343**,** d5928.
HIGGINS, J. P. & GREEN, S. 2011. *Cochrane Handbook for Systematic Reviews of Interventions*, The Cochrane Collaboration.
HILL, A. B. 1965. The environment and disease: association or causation? *Proceedings of the Royal society of Medicine,* 58**,** 295.
HOLLAND, P. W. 1986. Statistics and causal inference. *Journal of the American statistical Association,* 81**,** 945-960.
HOLMES, J., ANGUS, C., BUYKX, P., ALLY, A., STONE, T., MEIER, P. & BRENNAN, A. 2016. Mortality and morbidity risks from alcohol consumption in the UK: analyses using the Sheffield Alcohol Policy Model (v. 2.7) to inform the UK Chief Medical Officers' review of the UK lower risk drinking guidelines. ScHARR, University of Sheffield.
HOLMES, J., MEIER, P. S., ANGUS, C. & BRENNAN, A. 2018. Scotland's policy on minimum unit pricing for alcohol: the legal barriers are gone, so what are the implications for implementation and evaluation? *Addiction,* 113**,** 203-204.
HUANG, G. C., UNGER, J. B., SOTO, D., FUJIMOTO, K., PENTZ, M. A., JORDAN-MARSH, M. & VALENTE, T. W. 2014. Peer influences: the impact of online and offline friendship networks on adolescent smoking and alcohol use. *J Adolesc Health,* 54**,** 508-14.
HUNG, C., YEN, L. & WU, W. 2009. Association of parents' alcohol use and family interaction with the initiation of alcohol use by sixth graders: a preliminary study in Taiwan. *BMC Public Health,* 9**,** 172.
IMAI, K. & RATKOVIC, M. 2014. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 76**,** 243-263.
JACKSON, N., DENNY, S. & AMERATUNGA, S. 2014. Social and socio-demographic neighborhood effects on adolescent alcohol use: a systematic review of multi-level studies. *Soc Sci Med,* 115**,** 10-20.
JANSSEN, T., LARSEN, H., PEETERS, M., PRONK, T., VOLLEBERGH, W. A. M. & WIERS, R. W. 2014. Interactions between Parental Alcohol-Specific Rules and Risk Personalities in the Prediction of Adolescent Alcohol Use. *Alcohol and Alcoholism,* 49**,** 579-585.
KATIKIREDDI, S. V., BEESTON, C., MILLARD, A., FORSYTH, R., DELUCA, P., DRUMMOND, C., EADIE, D., GRAHAM, L., HILTON, S. & LUDBROOK, A. 2019. Evaluating possible intended and

unintended consequences of the implementation of alcohol minimum unit pricing (MUP) in Scotland: a natural experiment protocol. *BMJ Open,* 9**,** e028482.
KATIKIREDDI, S. V., HILTON, S., BONELL, C. & BOND, L. 2014. Understanding the development of minimum unit pricing of alcohol in Scotland: a qualitative study of the policy process. *PLoS One,* 9**,** e91185.
KENDLER, K. S., GARDNER, C. O., EDWARDS, A., HICKMAN, M., HERON, J., MACLEOD, J., LEWIS, G. & DICK, D. M. 2013. Dimensions of parental alcohol use/problems and offspring temperament, externalizing behaviors, and alcohol use/problems. *Alcohol Clin Exp Res,* 37**,** 2118-27.
KENDLER, K. S., GARDNER, C. O., EDWARDS, A. C., DICK, D. M., HICKMAN, M., MACLEOD, J. & HERON, J. 2018. Childhood Risk Factors for Heavy Episodic Alcohol Use and Alcohol Problems in Late Adolescence: A Marginal Structural Model Analysis. *Journal of Studies on Alcohol and Drugs,* 79**,** 370-379.
KHOUJA, J. N., MUNAFÒ, M. R., TILLING, K., WILES, N. J., JOINSON, C., ETCHELLS, P. J., JOHN, A., HAYES, F. M., GAGE, S. H. & CORNISH, R. P. 2019. Is screen time associated with anxiety or depression in young people? Results from a UK birth cohort. *BMC Public Health,* 19**,** 82.
KING, G. & NIELSEN, R. 2016. *Why propensity scores should not be used for matching* [Online]. http://j.mp/1sexgVw [Accessed 29 July 2019].
KLINE, R. B. 2015. *Principles and practice of structural equation modeling,* London, Guilford publications.
KONING, I. M., VAN DEN EIJNDEN, R. J., VERDURMEN, J. E., ENGELS, R. C. & VOLLEBERGH, W. A. 2012. Developmental alcohol-specific parenting profiles in adolescence and their relationships with adolescents' alcohol use. *Journal of Youth and Adolescence,* 41**,** 1502-1511.
KONING, I. M., VAN DEN EIJNDEN, R. J. J. M. & VOLLEBERGH, W. A. M. 2014. Alcohol-Specific Parenting, Adolescents' Self-Control, and Alcohol Use: A Moderated Mediation Model. *Journal of Studies on Alcohol and Drugs,* 75**,** 16-23.
KOUTAKIS, N., STATTIN, H. & KERR, M. 2008. Reducing youth alcohol drinking through a parent-targeted intervention: the Örebro Prevention Program. *Addiction,* 103**,** 1629-1637.
KRIEGER, N. & DAVEY SMITH, G. 2016. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol,* 45**,** 1787-1808.
KUNDI, M. 2006. Causality and the interpretation of epidemiologic evidence. *Environ Health Perspect,* 114**,** 969-74.
KUNTSCHE, E., VAN DER VORST, H. & ENGELS, R. 2009. The earlier the more? Differences in the links between age at first drink and adolescent alcohol use and related problems according to quality of parent–child relationships. *Journal of Studies on Alcohol and Drugs,* 70**,** 346-354.
LASSI, G., TAYLOR, A. E., MAHEDY, L., HERON, J., EISEN, T. & MUNAFÒ, M. R. 2019. Locus of control is associated with tobacco and alcohol consumption in young adults of the Avon Longitudinal Study of Parents and Children. *Royal Society Open Science,* 6**,** 181133.
LATENDRESSE, S. J., ROSE, R. J., VIKEN, R. J., PULKKINEN, L., KAPRIO, J. & DICK, D. M. 2008. Parenting mechanisms in links between parents' and adolescents' alcohol use behaviors. *Alcohol Clin Exp Res,* 32**,** 322-30.
LATENDRESSE, S. J., ROSE, R. J., VIKEN, R. J., PULKKINEN, L., KAPRIO, J. & DICK, D. M. 2010. Examining the etiology of associations between perceived parenting and adolescents' alcohol use: common genetic and/or environmental liabilities? *Journal of studies on alcohol and drugs,* 71**,** 313-325.
LEDERER, D. J., BELL, S. C., BRANSON, R. D., CHALMERS, J. D., MARSHALL, R., MASLOVE, D. M., OST, D. E., PUNJABI, N. M., SCHATZ, M. & SMYTH, A. R. 2019. Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals. *Annals of the American Thoracic Society,* 16**,** 22-28.

LILJESTROM, L., WIKSTROM, A. K. & JONSSON, M. 2018. Obstetric emergencies as antecedents to neonatal hypoxic ischemic encephalopathy, does parity matter? *Acta obstetricia et gynecologica Scandinavica,* 97**,** 1396-1404.

LIN, S. H., YOUNG, J., LOGAN, R., TCHETGEN, E. J. T. & VANDERWEELE, T. J. 2017. Parametric mediational g-formula approach to mediation analysis with time-varying exposures, mediators, and confounders. *Epidemiology (Cambridge, Mass.),* 28**,** 266.

LOWE, H. J. & BARNETT, G. O. 1994. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *Jama,* 271**,** 1103-1108.

MACARTHUR, G. J., HARRISON, S., CALDWELL, D. M., HICKMAN, M. & CAMPBELL, R. 2016. Peer-led interventions to prevent tobacco, alcohol and/or drug use among young people aged 11–21 years: a systematic review and meta-analysis. *Addiction,* 111**,** 391-407.

MACKINNON, D. P., KRULL, J. L. & LOCKWOOD, C. M. 2000. Equivalence of the mediation, confounding and suppression effect. *Prevention science : the official journal of the Society for Prevention Research,* 1**,** 173-181.

MAGGS, J. L., PATRICK, M. E. & FEINSTEIN, L. 2008. Childhood and adolescent predictors of alcohol use and problems in adolescence and adulthood in the National Child Development Study. *Addiction,* 103 Suppl 1**,** 7-22.

MAHEDY, L., MACARTHUR, G. J., HAMMERTON, G., EDWARDS, A. C., KENDLER, K. S., MACLEOD, J., HICKMAN, M., MOORE, S. C. & HERON, J. 2018. The effect of parental drinking on alcohol use in young adults: the mediating role of parental monitoring and peer deviance. *Addiction,* 113**,** 2041-2050.

MANIGLIO, R. 2009. The impact of child sexual abuse on health: a systematic review of reviews. *Clin Psychol Rev,* 29**,** 647-57.

MANIGLIO, R. 2010. Child sexual abuse in the etiology of depression: A systematic review of reviews. *Depression and Anxiety,* 27**,** 631-642.

MARCHI, J., BERG, M., DENCKER, A., OLANDER, E. K. & BEGLEY, C. 2015. Risks associated with obesity in pregnancy, for the mother and baby: a systematic review of reviews. *Obesity Reviews,* 16**,** 621-638.

MARES, S. H. W., VAN DER VORST, H., ENGELS, R. C. M. E. & LICHTWARCK-ASCHOFF, A. 2011. Parental alcohol use, alcohol-related problems, and alcohol-specific attitudes, alcohol-specific communication, and adolescent excessive alcohol use and alcohol-related problems: An indirect path model. *Addictive Behaviors,* 36**,** 209-216.

MARS, B., HERON, J., KLONSKY, E. D., MORAN, P., O'CONNOR, R. C., TILLING, K., WILKINSON, P. & GUNNELL, D. 2019. What distinguishes adolescents with suicidal thoughts from those who have attempted suicide? A population-based birth cohort study. *Journal of Child Psychology and Psychiatry,* 60**,** 91-99.

MARSHAL, M. P., FRIEDMAN, M. S., STALL, R., KING, K. M., MILES, J., GOLD, M. A., BUKSTEIN, O. G. & MORSE, J. Q. 2008. Sexual orientation and adolescent substance use: a meta-analysis and methodological review. *Addiction (Abingdon, England),* 103**,** 546-556.

MATTICK, R. P., CLARE, P. J., AIKEN, A., WADOLOWSKI, M., HUTCHINSON, D., NAJMAN, J., SLADE, T., BRUNO, R., MCBRIDE, N., KYPRI, K., VOGL, L. & DEGENHARDT, L. 2018. Association of parental supply of alcohol with adolescent drinking, alcohol-related harms, and alcohol use disorder symptoms: a prospective cohort study. *Lancet Public Health,* 3**,** E64-E71.

MCCANN, M., PERRA, O., MCLAUGHLIN, A., MCCARTAN, C. & HIGGINS, K. 2016. Assessing elements of a family approach to reduce adolescent drinking frequency: parent–adolescent relationship, knowledge management and keeping secrets. *Addiction,* 111**,** 843-853.

MCMORRIS, B. J., CATALANO, R. F., KIM, M. J., TOUMBOUROU, J. W. & HEMPHILL, S. A. 2011. Influence of Family Factors and Supervised Alcohol Use on Adolescent Alcohol Use and Harms: Similarities Between Youth in Different Alcohol Policy Contexts. *Journal of Studies on Alcohol and Drugs,* 72**,** 418-428.

MENEZES, A. M. B., MURRAY, J., LÁSZLÓ, M., WEHRMEISTER, F. C., HALLAL, P. C., GONÇALVES, H., ASSUNÇÃO, M. C. F., MENEZES, C. B. & BARROS, F. C. 2013. Happiness and depression in adolescence after maternal smoking during pregnancy: birth cohort study. *PloS one,* 8**,** e80370.

METHLEY, A. M., CAMPBELL, S., CHEW-GRAHAM, C., MCNALLY, R. & CHERAGHI-SOHI, S. 2014. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Serv Res,* 14**,** 579.

MITCHELL, R. J. 2001. Path analysis. *Design and analysis of ecological experiments. Oxford University Press, Oxford, UK***,** 217-234.

MODECKI, K. L., BARBER, B. L. & ECCLES, J. S. 2014. Binge Drinking Trajectories Across Adolescence: For Early Maturing Youth, Extra-Curricular Activities Are Protective. *Journal of Adolescent Health,* 54**,** 61-66.

MOEHRING, A., RUMPF, H., HAPKE, U., BISCHOF, G., JOHN, U. & MEYER, C. 2019. Diagnostic performance of the Alcohol Use Disorders Identification Test (AUDIT) in detecting DSM-5 alcohol use disorders in the General population. *Drug and alcohol dependence,* 204**,** 107530.

MOHER, D., LIBERATI, A., TETZLAFF, J. & ALTMAN, D. G. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine,* 151**,** 264-269.

MOORE, A. A., BECK, J. C., BABOR, T. F., HAYS, R. D. & REUBEN, D. B. 2002. Beyond alcoholism: identifying older, at-risk drinkers in primary care. *Journal of studies on alcohol,* 63**,** 316-324.

MOORE, T. H. M., ZAMMIT, S., LINGFORD-HUGHES, A., BARNES, T. R. E., JONES, P. B., BURKE, M. & LEWIS, G. 2007. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *The Lancet,* 370**,** 319-328.

MORENO-BETANCUR, M. & CARLIN, J. B. 2018. Understanding Interventional Effects: A More Natural Approach to Mediation Analysis? *Epidemiology,* 29**,** 614-617.

MORGAN, S. L. & WINSHIP, C. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research,* New York, Cambridge University Press.

MORGAN, S. L. & WINSHIP, C. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research, 2nd Edition,* New York, Cambridge University Press.

MUNAFÒ, M. R. & DAVEY SMITH, G. 2018. Robust research needs many lines of evidence. Nature Publishing Group.

MUNAFÒ, M. R., TILLING, K., TAYLOR, A. E., EVANS, D. M. & DAVEY SMITH, G. 2017. Collider scope: when selection bias can substantially influence observed associations. *International journal of epidemiology,* 47**,** 226-235.

MURAD, M. H., ASI, N., ALSAWAS, M. & ALAHDAB, F. 2016. New evidence pyramid. *Evidence Based Medicine,* 21**,** 125-127.

NAIMI, A. I., KAUFMAN, J. S. & MACLEHOSE, R. F. 2014. Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. *Int J Epidemiol,* 43**,** 1656-61.

NICHTER, M., NICHTER, M., CARKOGLU, A. & LLOYD-RICHARDSON, E. 2010. Smoking and drinking among college students: “It's a package deal”. *Drug and Alcohol Dependence,* 106**,** 16-20.

NOBLE, H. & SMITH, J. 2018. Reviewing the literature: choosing a review design. *Evidence Based Nursing,* 21**,** 39-41.

OAKES, J. M. & KAUFMAN, J. S. 2017. *Methods in social epidemiology,* San Francisco, John Wiley & Sons.

OFFICE FOR NATIONAL STATISTICS. 2011. *2011 Census: Tenure by Rural/Urban* [Online]. Available: http://www.nomisweb.co.uk/census/2011/QS405EW/view/2092957703?rows=c_tenhuk11&cols=rural_urban [Accessed 29 July 2019].

OLDHAM, M., HOLMES, J., WHITAKER, V., FAIRBROTHER, H. & CURTIS, P. 2018. Youth drinking in decline. ScHARR, University of Sheffield.

PAPE, H. & BYE, E. K. 2017. Drinking with parents: Different measures, different associations with underage heavy drinking? *Nordic Studies on Alcohol and Drugs,* 34**,** 445-455.
PEARL, J. 2009. *Causality,* New York, Cambridge University Press.
PEARL, J. 2018. Comments on: The tale wagged by the DAG. *Int J Epidemiol,* 47**,** 1002-1004.
PEARL, J., GLYMOUR, M. & JEWELL, N. P. 2016. *Causal Inference in Statistics: A Primer,* Chichester, Wiley.
PEARS, K., CAPALDI, D. M. & OWEN, L. D. 2007. Substance use risk across three generations: The roles of parent discipline practices and inhibitory control. *Psychology of Addictive Behaviors,* 21**,** 373-386.
PEDERSEN, W. & VON SOEST, T. 2013. Socialization to binge drinking: a population-based, longitudinal study with emphasis on parental influences. *Drug Alcohol Depend,* 133**,** 587-92.
PETRIE, J., BUNN, F. & BYRNE, G. 2007. Parenting programmes for preventing tobacco, alcohol or drugs misuse in children <18: a systematic review. *Health Educ Res,* 22**,** 177-91.
PHILPOT, M., PEARSON, N., PETRATOU, V., DAYANANDAN, R., SILVERMAN, M. & MARSHALL, J. 2003. Screening for problem drinking in older people referred to a mental health service: a comparison of CAGE and AUDIT. *Aging & Mental Health,* 7**,** 171-175.
PICCOLO, R. S., PEARCE, N., ARAUJO, A. B. & MCKINLAY, J. B. 2014. The contribution of biogeographical ancestry and socioeconomic status to racial/ethnic disparities in type 2 diabetes mellitus: results from the Boston Area Community Health Survey. *Annals of epidemiology,* 24**,** 648-654. e1.
PIETERS, S., BURK, W. J., VAN DER VORST, H., WIERS, R. W. & ENGELS, R. C. 2012. The moderating role of working memory capacity and alcohol-specific rule-setting on the relation between approach tendencies and alcohol use in young adolescents. *Alcoholism: Clinical and Experimental Research,* 36**,** 915-922.
POELEN, E. A., ENGELS, R. C., SCHOLTE, R. H., BOOMSMA, D. I. & WILLEMSEN, G. 2009. Predictors of problem drinking in adolescence and young adulthood. A longitudinal twin-family study. *Eur Child Adolesc Psychiatry,* 18**,** 345-52.
POELEN, E. A. P., SCHOLTE, R. H. J., WILLEMSEN, G., BOOMSMA, D. I. & ENGELS, R. C. M. E. 2007. Drinking by parents, siblings, and friends as predictors of regular alcohol use in adolescents and young adults: A longitudinal twin-family study. *Alcohol and Alcoholism,* 42**,** 362-369.
POLZER, I., SCHWAHN, C., VÖLZKE, H., MUNDT, T. & BIFFAR, R. 2012. The association of tooth loss with all-cause and circulatory mortality. Is there a benefit of replaced teeth? A systematic review and meta-analysis. *Clinical oral investigations,* 16**,** 333-351.
POULIN, F. & DENAULT, A. S. 2012. Other-sex friendships as a mediator between parental monitoring and substance use in girls and boys. *Journal of youth and adolescence,* 41**,** 1488-1501.
RAC FOUNDATION. 2012. *Car ownerhip rates per local authority in England and Wales* [Online]. Available: https://www.racfoundation.org/assets/rac_foundation/content/downloadables/car%20ownership%20rates%20by%20local%20authority%20-%20december%202012.pdf [Accessed 10 January 2019].
RANINEN, J., LIVINGSTON, M., KARLSSON, P., LEIFMAN, H., GUTTORMSSON, U., SVENSSON, J. & LARM, P. 2018. One explanation to rule them all? Identifying sub-groups of non-drinking Swedish ninth graders. *Drug and alcohol review,* 37**,** S42-S48.
REINERT, D. F. & ALLEN, J. P. 2007. The alcohol use disorders identification test: an update of research findings. *Alcoholism: Clinical and Experimental Research,* 31**,** 185-199.
REMES, O., BRAYNE, C., LINDE, R. & LAFORTUNE, L. 2016. A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain and Behavior,* 6**,** e00497.
RICHARDSON, W. S., WILSON, M. C., NISHIKAWA, J. & HAYWARD, R. S. 1995. The well-built clinical question: a key to evidence-based decisions. *ACP journal club,* 123**,** A12-A12.
RIOUX, C., CASTELLANOS-RYAN, N., PARENT, S., VITARO, F., TREMBLAY, R. E. & SEGUIN, J. R. 2016. Differential susceptibility to environmental influences: Interactions between child

temperament and parenting in adolescent alcohol use. *Development and psychopathology,* 28**,** 265-275.
ROBINS, J. M., HERNAN, M. A. & BRUMBACK, B. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology,* 11**,** 550-560.
ROBINS, J. M. & WASSERMAN, L. 1999. On the impossibility of inferring causation from association without background knowledge. *In:* COOPER, G. F. & GLYMOUR, C. (eds.) *Computation, causation, and discovery.* Cambridge: MIT Press.
ROBINS, J. M. & WEISSMAN, M. B. 2016. Commentary: Counterfactual causation and streetlamps: what is to be done? *Int J Epidemiol,* 45**,** 1830-1835.
ROOM, R. 2004. Smoking and drinking as complementary behaviours. *Biomedicine & Pharmacotherapy,* 58**,** 111-115.
ROOM, R., BABOR, T. & REHM, J. 2005. Alcohol and public health. *The Lancet,* 365**,** 519-530.
ROSENBAUM, P. R. & RUBIN, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70**,** 41-55.
ROTHMAN, K. J. & GREENLAND, S. 2004. Causation and causal inference. *In:* DETELS, R. (ed.) *Oxford Textbook of Public Health.* 4 ed. New York: Oxford University Press.
ROTHMAN, K. J. & GREENLAND, S. 2005a. Causation and causal inference in epidemiology. *Am J Public Health,* 95 Suppl 1**,** S144-50.
ROTHMAN, K. J. & GREENLAND, S. 2005b. Causation and Causal Inference in Epidemiology. *American Journal of Public Health,* 95**,** S144-S150.
ROTHMAN, K. J., GREENLAND, S. & LASH, T. L. 2008. *Modern epidemiology,* Philadelphia, Lippincott Williams & Wilkins.
ROYSTON, P. 2004. Multiple imputation of missing values. *The Stata Journal,* 4**,** 227-241.
ROYSTON, P. & WHITE, I. R. 2011. Multiple imputation by chained equations (MICE): implementation in Stata. *J Stat Softw,* 45**,** 1-20.
RUBIN, D. B. 1997. Estimating causal effects from large data sets using propensity scores. *Annals of internal medicine,* 127**,** 757-763.
RYAN, S. M., JORM, A. F. & LUBMAN, D. I. 2010. Parenting factors associated with reduced adolescent alcohol use: a systematic review of longitudinal studies. *Aust N Z J Psychiatry,* 44**,** 774-83.
SALVATORE, J., ALIEV, F., EDWARDS, A., EVANS, D., MACLEOD, J., HICKMAN, M., LEWIS, G., KENDLER, K., LOUKOLA, A. & KORHONEN, T. 2014. Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. *Genes,* 5**,** 330-346.
SANTIS, R., GARMENDIA, M. L., ACUNA, G., ALVARADO, M. E. & ARTEAGA, O. 2009. The Alcohol Use Disorders Identification Test (AUDIT) as a screening instrument for adolescents. *Drug Alcohol Depend,* 103**,** 155-8.
SARTOR, C. E., LYNSKEY, M. T., BUCHOLZ, K. K., MCCUTCHEON, V. V., NELSON, E. C., WALDRON, M. & HEATH, A. C. 2007. Childhood sexual abuse and the course of alcohol dependence development: findings from a female twin sample. *Drug Alcohol Depend,* 89**,** 139-44.
SAUER, B. C., BROOKHART, M. A., ROY, J. & VANDERWEELE, T. 2013. A review of covariate selection for non-experimental comparative effectiveness research. *Pharmacoepidemiol Drug Saf,* 22**,** 1139-45.
SAUNDERS, J. B., AASLAND, O. G., BABOR, T. F., DE LA FUENTE, J. R. & GRANT, M. 1993. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction,* 88**,** 791-804.
SCHMIDT, S. C. E. & WOLL, A. 2017. Longitudinal drop-out and weighting against its bias. *BMC Med Res Methodol,* 17**,** 164.
SCHÜNEMANN, H., HILL, S., GUYATT, G., AKL, E. A. & AHMED, F. 2011. The GRADE approach and Bradford Hill's criteria for causation. *Journal of Epidemiology & Community Health,* 65**,** 392-395.

SCOTTISH GOVERNMENT 2018. Delivering for today, investing for tomorrow: the Government's programme for Scotland 2018-2019. Edinburgh.
SELJAMO, S., AROMAA, M., KOIVUSILTA, L., RAUTAVA, P., SOURANDER, A., HELENIUS, H. & SILLANPAA, M. 2006. Alcohol use in families: a 15-year prospective follow-up study. *Addiction,* 101**,** 984-92.
SHARMIN, S., KYPRI, K., KHANAM, M., WADOLOWSKI, M., BRUNO, R., ATTIA, J., HOLLIDAY, E., PALAZZI, K. & MATTICK, R. P. 2017a. Effects of parental alcohol rules on risky drinking and related problems in adolescence: Systematic review and meta-analysis. *Drug and alcohol dependence,* 178**,** 243-256.
SHARMIN, S., KYPRI, K., KHANAM, M., WADOLOWSKI, M., BRUNO, R. & MATTICK, R. P. 2017b. Parental Supply of Alcohol in Childhood and Risky Drinking in Adolescence: Systematic Review and Meta-Analysis. *Int J Environ Res Public Health,* 14.
SHEA, B. J., BOUTER, L. M., PETERSON, J., BOERS, M., ANDERSSON, N., ORTIZ, Z., RAMSAY, T., BAI, A., SHUKLA, V. K. & GRIMSHAW, J. M. 2007a. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PloS one,* 2**,** e1350.
SHEA, B. J., GRIMSHAW, J. M., WELLS, G. A., BOERS, M., ANDERSSON, N., HAMEL, C., PORTER, A. C., TUGWELL, P., MOHER, D. & BOUTER, L. M. 2007b. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology,* 7**,** 10.
SHEA, B. J., HAMEL, C., WELLS, G. A., BOUTER, L. M., KRISTJANSSON, E., GRIMSHAW, J., HENRY, D. A. & BOERS, M. 2009. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of clinical epidemiology,* 62**,** 1013-1020.
SHEA, B. J., REEVES, B. C., WELLS, G., THUKU, M., HAMEL, C., MORAN, J., MOHER, D., TUGWELL, P., WELCH, V., KRISTJANSSON, E. & HENRY, D. A. 2017. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ,* 358.
SHORTT, A. L., HUTCHINSON, D. M., CHAPMAN, R. & TOUMBOUROU, J. W. 2007. Family, school, peer and individual influences on early adolescent alcohol use: first-year impact of the Resilient Families programme. *Drug and Alcohol Review,* 26**,** 625-634.
SMIT, E., VERDURMEN, J., MONSHOUWER, K. & SMIT, F. 2008. Family interventions and their effect on adolescent alcohol use in general populations; a meta-analysis of randomized controlled trials. *Drug Alcohol Depend,* 97**,** 195-206.
SMITH, G. C. S. & PELL, J. P. 2003. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *British Medical Journal,* 327**,** 1459-1461.
SMITH, V., DEVANE, D., BEGLEY, C. M. & CLARKE, M. 2011. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Medical Research Methodology,* 11**,** 15.
SPIJKERMAN, R., VAN DEN EIJNDEN, R. J. J. M., OVERBEEK, G. & ENGELS, R. C. M. E. 2007. The impact of peer and parental norms and behavior on adolescent drinking: The role of drinker prototypes. *Psychology & Health,* 22**,** 7-29.
STEEN, J., LOEYS, T., MOERKERKE, B. & VANSTEELANDT, S. 2017. Flexible Mediation Analysis With Multiple Mediators. *Am J Epidemiol,* 186**,** 184-193.
STERNE, J. A., HERNÁN, M. A., REEVES, B. C., SAVOVIĆ, J., BERKMAN, N. D., VISWANATHAN, M., HENRY, D., ALTMAN, D. G., ANSARI, M. T. & BOUTRON, I. 2016. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj,* 355**,** i4919.
STICE, E., BARRERA JR, M. & CHASSIN, L. 1998. Prospective differential prediction of adolescent alcohol use and problem use: Examining the mechanisms of effect. *Journal of abnormal psychology,* 107**,** 616.

STRANDBERG, A. K., BODIN, M. C. & ROMELSJO, A. 2014. Gender differences in the prediction of parental servings of alcohol to adolescents and youth drunkenness. *Subst Use Misuse,* 49**,** 1857-66.
SUSSER, M. 1986. The logic of Sir Karl Popper and the practice of epidemiology. *American Journal of Epidemiology,* 124**,** 711-718.
SUTTON, A. J., COOPER, N. J. & JONES, D. R. 2009. Evidence synthesis as the key to more coherent and efficient research. *BMC medical research methodology,* 9**,** 29.
TENNANT, P. W. G., TEXTOR, J., GILTHORPE, M. S. & ELLISON, G. T. H. 2017. OP87 Dagitty and directed acyclic graphs in observational research: a critical review. *Journal of Epidemiology and Community Health,* 71**,** A43.2-A43.
TEUNISSEN, H. A., KUNTSCHE, E., SCHOLTE, R. H., SPIJKERMAN, R., PRINSTEIN, M. J. & ENGELS, R. C. 2016. Friends' drinking norms and male adolescents' alcohol consumption: The moderating role of performance-based peer influence susceptibility. *Journal of adolescence,* 53**,** 45-54.
TEXTOR, J., HARDT, J. & KNUPPEL, S. 2011. DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology,* 22**,** 745.
TEXTOR, J., VAN DER ZANDER, B., GILTHORPE, M. S., LISKIEWICZ, M. & ELLISON, G. T. 2016. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol,* 45**,** 1887-1894.
THOEMMES, F. & ONG, A. D. 2015. A Primer on Inverse Probability of Treatment Weighting and Marginal Structural Models. *Emerging Adulthood,* 4**,** 40-59.
TILLING, K., WILLIAMSON, E. J., SPRATT, M., STERNE, J. A. & CARPENTER, J. R. 2016. Appropriate inclusion of interactions was needed to avoid bias in multiple imputation. *J Clin Epidemiol,* 80**,** 107-115.
TOD, E., GRANT, I., WYPER, G., MESALLES-NARANJO, O., STOCKTON, D., ROBINSON, M., MCCARTNEY, G., FISCHBACHER, C., DOBBIE, R. & CRAIG, N. 2018. *Hospital admissions, deaths and overall burden of disease attributable to alcohol consumption in Scotland***,** Scottish Public Health Observatory/NHS Health Scotland.
TSAI, A. & GOODMAN, M. 2013. Achieving consensus on terminology describing multivariable analyses. *American journal of public health,* 103**,** E1.
TUCKER, J. S., ORLANDO, M. & ELLICKSON, P. L. 2003. Patterns and correlates of binge drinking trajectories from early adolescence to young adulthood. *Health Psychology,* 22**,** 79.
VAN DEN EIJNDEN, R., VAN DE MHEEN, D., VET, R. & VERMULST, A. 2011. Alcohol-specific parenting and adolescents' alcohol-related problems: The interacting role of alcohol availability at home and parental rules. *Journal of studies on alcohol and drugs,* 72**,** 408-417.
VAN DER VORST, H., ENGELS, R. C. & BURK, W. J. 2010. Do parents and best friends influence the normative increase in adolescents' alcohol use at home and outside the home? *Journal of studies on alcohol and drugs,* 71**,** 105-114.
VAN DER VORST, H., ENGELS, R. C., DEKOVIĆ, M., MEEUS, W. & VERMULST, A. A. 2007. Alcohol-specific rules, personality and adolescents' alcohol use: A longitudinal person–environment study. *Addiction,* 102**,** 1064-1075.
VAN DER VORST, H., ENGELS, R. C., MEEUS, W. & DEKOVIĆ, M. 2006. Parental attachment, parental control, and early development of alcohol use: A longitudinal study. *Psychology of addictive behaviors,* 20**,** 107.
VAN DER VORST, H., KRANK, M., ENGELS, R. C., PIETERS, S., BURK, W. J. & MARES, S. H. 2013. The mediating role of alcohol-related memory associations on the relation between perceived parental drinking and the onset of adolescents' alcohol use. *Addiction,* 108**,** 526-533.
VAN DER ZWALUW, C. S., OTTEN, R., KLEINJAN, M. & ENGELS, R. C. 2014. Different trajectories of adolescent alcohol use: testing gene–environment interactions. *Alcoholism: clinical and experimental research,* 38**,** 704-712.

VAN DER ZWALUW, C. S., SCHOLTE, R. H., VERMULST, A. A., BUITELAAR, J. K., VERKES, R. J. & ENGELS, R. C. 2008. Parental problem drinking, parenting, and adolescent alcohol use. *J Behav Med,* 31**,** 189-200.

VANDENBROUCKE, J. P., BROADBENT, A. & PEARCE, N. 2016. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol,* 45**,** 1776-1786.

VANDERVEEN, J. W., COHEN, L. M. & WATSON, N. L. 2013. Utilizing a multimodal assessment strategy to examine variations of impulsivity among young adults engaged in co-occurring smoking and binge drinking behaviors. *Drug and Alcohol Dependence,* 127**,** 150-155.

VANDERWEELE, T. 2015. *Explanation in causal inference: methods for mediation and interaction,* New York, Oxford University Press.

VANDERWEELE, T. J. 2016. Commentary: On Causes, Causal Inference, and Potential Outcomes. *Int J Epidemiol,* 45**,** 1809-1816.

VANDERWEELE, T. J. 2019. Principles of confounder selection. *European journal of epidemiology***,** 1-9.

VANDERWEELE, T. J., VANSTEELANDT, S. & ROBINS, J. M. 2014. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology,* 25**,** 300-6.

VANSTEELANDT, S. & DANIEL, R. M. 2017. Interventional Effects for Mediation Analysis with Multiple Mediators. *Epidemiology,* 28**,** 258-265.

VERMEULEN-SMIT, E., KONING, I. M., VERDURMEN, J. E. E., VAN DER VORST, H., ENGELS, R. C. M. E. & VOLLEBERGH, W. A. M. 2012. The influence of paternal and maternal drinking patterns within two-partner families on the initiation and development of adolescent drinking. *Addictive Behaviors,* 37**,** 1248-1256.

VISSER, L., DE WINTER, A. F. & REIJNEVELD, S. A. 2012. The parent–child relationship and adolescent alcohol use: a systematic review of longitudinal studies. *BMC Public Health,* 12**,** 1-16.

VISSER, L., DE WINTER, A. F., VOLLEBERGH, W. A. M., VERHULST, F. C. & REIJNEVELD, S. A. 2013. The Impact of Parenting Styles on Adolescent Alcohol Use: The TRAILS Study. *European Addiction Research,* 19**,** 165-172.

VON HIPPEL, P. T. 2009. How to Impute Interactions, Squares, and Other Transformed Variables. *Sociological Methodology***,** 265-291.

WAGENAAR, A. C. & TOOMEY, T. L. 2002. Effects of minimum drinking age laws: review and analyses of the literature from 1960 to 2000. *Journal of Studies on Alcohol, Supplement***,** 206-225.

WANG, A. & ARAH, O. A. 2015. G-computation demonstration in causal mediation analysis. *Eur J Epidemiol,* 30**,** 1119-27.

WANG, C., HIPP, J. R., BUTTS, C. T., JOSE, R. & LAKON, C. M. 2017. Peer influence, peer selection and adolescent alcohol use: a simulation study using a dynamic network model of friendship ties and alcohol use. *Prevention science,* 18**,** 382-393.

WEICHOLD, K., WIESNER, M. F. & SILBEREISEN, R. K. 2014. Childhood Predictors and Mid-Adolescent Correlates of Developmental Trajectories of Alcohol Use among Male and Female Youth. *Journal of Youth and Adolescence,* 43**,** 698-716.

WEITZMAN, E. R. & CHEN, Y. 2005. The co-occurrence of smoking and drinking among young adults in college: National survey results from the United States. *Drug and Alcohol Dependence,* 80**,** 377-386.

WESTREICH, D. & GREENLAND, S. 2013. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol,* 177**,** 292-8.

WHITE, I. R., ROYSTON, P. & WOOD, A. M. 2011. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine,* 30**,** 377-399.

WINDLE, M., SPEAR, L. P., FULIGNI, A. J., ANGOLD, A., BROWN, J. D., PINE, D., SMITH, G. T., GIEDD, J. & DAHL, R. E. 2008. Transitions into underage and problem drinking: developmental processes and mechanisms between 10 and 15 years of age. *Pediatrics,* 121**,** S273-S289.

WOLKE, D., WAYLEN, A., SAMARA, M., STEER, C., GOODMAN, R., FORD, T. & LAMBERTS, K. 2009. Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *Br J Psychiatry,* 195**,** 249-56.

WORLD HEALTH ORGANISATION 2014. Global status report on alcohol and health 2014.
WORLD HEALTH ORGANIZATION 2010. Global strategy to reduce the harmful use of alcohol.
WRIGHT, S. 1960. Path Coefficients and Path Regressions - Alternative or Complementary Concepts. *Biometrics,* 16**,** 189-202.
YAP, M. B. H., CHEONG, T. W. K., ZARAVINOS-TSAKOS, F., LUBMAN, D. I. & JORM, A. F. 2017. Modifiable parenting factors associated with adolescent alcohol misuse: a systematic review and meta-analysis of longitudinal studies. *Addiction,* 112**,** 1142-1162.
ZAMPIERI, F. G., AGUIAR, F. J., BOZZA, F. A., SALLUH, J. I., SOARES, M. & INVESTIGATORS, O. S. 2019. Modulators of systemic inflammatory response syndrome presence in patients admitted to intensive care units with acute infection: a Bayesian network approach. *Intensive care medicine***,** 1-3.