

Original Article

Cite this article: Halladay J, Visontay R, Slade T, Devine EK, Smout S, Andrews JL, Champion KE, Teesson M, Sunderland M (2024). Across the multiverse: exploring a diverse set of specifications related to cross-sectional and prospective associations between adolescent alcohol use and emotional problems. *Psychological Medicine* **54**, 4585–4599. <https://doi.org/10.1017/S0033291724002502>

Received: 20 April 2024

Revised: 26 July 2024

Accepted: 17 September 2024

First published online: 16 December 2024

Keywords:

adolescent; alcohol drinking; anxiety; depression; multiverse; psychological distress

Corresponding author:

J. Halladay;

Email: halladje@mcmaster.ca;

jillian.halladay@sydney.edu.au

Across the multiverse: exploring a diverse set of specifications related to cross-sectional and prospective associations between adolescent alcohol use and emotional problems

J. Halladay^{1,2,3} , R. Visontay², T. Slade², E. K. Devine², S. Smout², J. L. Andrews^{2,4}, K. E. Champion^{2,5}, M. Teesson² and M. Sunderland²

¹School of Nursing, McMaster University, Ontario, Canada; ²The Matilda Centre for Research in Mental Health and Substance Use, University of Sydney, New South Wales, Australia; ³Peter Boris Centre for Addictions Research, McMaster University/St. Joseph's Healthcare Hamilton, Ontario, Canada; ⁴Department of Experimental Psychology, University of Oxford, Oxford, UK and ⁵School of Public Health, University of Sydney, New South Wales, Australia

Abstract

Background. The relationship between adolescent alcohol use and emotional problems remains unclear and contradictory. These inconsistencies may in part be due to differences in the measurement and operationalization of alcohol use and emotional problems across studies, as well as confounder selection and missing data decisions. This study explores the associations between common specifications of adolescent alcohol use and emotional problems in a large sample of adolescents.

Methods. A multiverse analysis (also known as specification curve analysis or vibration of effects) was done with 7680 unique model specifications in a large longitudinal sample of 6639 Australian adolescents (aged ~14.7–15.7, 2021–2022).

Results. While alcohol use and emotional problems nearly universally co-occurred in minimally adjusted cross-sectional models (98–99%), the operationalization of emotional problems, temporality of prospective relationships, and choice of confounders substantially impacted findings. Emotional problems appeared to predict later alcohol use more-so than the reverse, depression-focused measures yielded more consistent associations with alcohol use than anxiety-focused measures, and certain confounders (i.e. conduct, ADHD, smoking) explained most of the associations between adolescent alcohol use and emotional problems. Missing data decisions and whether outcomes were modelled continuously *v.* dichotomously had minimal impact on findings.

Conclusions. While adolescent alcohol use and emotional problems commonly co-occur, inconsistencies in the magnitude, direction, and significance of effects are closely tied to researcher decisions that are often made arbitrarily.

Background

Adolescence is a sensitive period during which most substance use (particularly alcohol use) and emotional problems (such as depression, anxiety, distress, and general internalizing symptoms) emerge (Solmi et al., 2021). These issues often co-occur in general population and clinical adolescent samples (Halladay, MacKillop, Munn, Amlung, & Georgiades, 2022; Hawke, Koyama, & Henderson, 2018; Suntharalingam et al., 2022), though the specific nature of the association between alcohol use and emotional problems remains unclear and contradictory. These inconsistencies may, in part, be due to differing researcher decisions regarding the conceptualization, measurement, and directionality of associations between alcohol use and emotional problems across studies. Contemporary statistical frameworks have been built to account for and uncover the impact of these researcher choices – which can be arbitrary – on the stability or robustness of effects across different specifications between alcohol use and emotional problems (Patel, Burford, & Ioannidis, 2015; Simonsohn, Simmons, & Nelson, 2020; Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). Given the global importance of preventing alcohol use harms and emotional problems (United Nations Department of Economic and Social Affairs, 2015), a more robust understanding of their relationship is needed, considering factors like the timing and direction of associations, choice of confounders, and operationalization of alcohol use and emotional problems.

Large longitudinal datasets are powerful tools for exploring these relationships. However, the hypothesized causal relationship between alcohol use and emotional problems, which informs a study's chosen statistical model, may impact the detection and magnitude of associations. Three

core hypotheses explain the co-occurrence of these problems. First, emotional problems may lead to alcohol use, often attributed to psychosocial pathways. Some studies suggest certain types of emotional problems predict early alcohol use and related problems later in life, though these associations are nuanced and inconsistent (Dyer, Easey, Heron, Hickman, & Munafò, 2019; Hussong, Ennett, Cox, & Haroon, 2017). Second, adolescent alcohol use may contribute to the development or worsening of emotional problems, often attributed to social, cognitive, or neurobiological pathways. Existing longitudinal studies reveal weak/negligible associations between adolescent alcohol quantity-frequency measures and later depression and anxiety (Cochrane Canada, 2022; McCabe, Brumback, Brown, & Meruelo, 2023). Third, alcohol use and emotional problems may share common risk and protective factors, leading to their co-occurrence due to confounding. Common confounders across studies include demographics, other substance use, and externalizing problems (Cochrane Canada, 2022; Dyer et al., 2019; Hussong et al., 2017; Ning, Gondek, Patalay, & Ploubidis, 2020). As such, researcher decisions pertaining to the selection of independent, dependent, and confounding variables may contribute to existing inconsistencies.

The operationalization of key constructs may also drive inconsistencies. Across existing studies, adolescent alcohol use is measured in diverse ways, from initial sipping to the diagnosis of alcohol use disorder (AUD). Common ways of operationalizing adolescent alcohol use include full standard drink consumption (prevalence and frequency), binge drinking (often defined as 5+ drinks/occasion), alcohol volume (frequency $\times$ quantity), and alcohol-related problem scales and diagnoses (Cochrane Canada, 2022; Dyer et al., 2019; Ning et al., 2020). Some researchers place emphasis on the age of initiation, particularly by age 14, as studies have shown earlier onset to strongly correlate with suicidality and longer-term alcohol problems (Ahuja, Awasthi, Records, & Lamichhane, 2021; Gardner, Stockings, Champion, Mather, & Newton, 2024; Lee et al., 2024). Additional inconsistencies may arise due to considerable variability in recall periods.

Existing research suggests that decisions about how to operationalize alcohol use may impact the nature of the associations found. For instance, a previous systematic review found a more consistent link between early initiation of alcohol use and depression compared to other aspects of alcohol use, such as alcohol problems, quantity-frequency, and AUD (Hussong et al., 2017). Conversely, another review found adolescent anxiety to be associated with later AUD, but found inconsistent evidence regarding anxiety's connection to other alcohol quantity-frequency variables (Dyer et al., 2019). A recent meta-analysis, predominantly focused on adults, revealed that individuals with mood or anxiety disorders had twice the odds of AUD compared to those without such disorders (Puddephatt, Irizar, Jones, Gage, & Goodwin, 2022). However, the direction and magnitude of effects was inconsistent for other quantity-frequency measures of alcohol use. This highlights the need for further investigation into the causes, correlates, and consequences of various facets of adolescent alcohol use in relation to emotional problems.

The term 'emotional problems', defined in the current study as encompassing depression, anxiety, psychological distress, and general internalizing-related factors, can be measured using various symptom scales and/or diagnostic assessments. While evidence suggests that these emotional problems may be better understood as a general internalizing factor (Watson et al., 2022), various sub-domains of emotional problems exhibit distinct associations with alcohol related factors (Dyer et al., 2019;

Hussong et al., 2017; Ning et al., 2020). For instance, measures of depression predict later alcohol use and related problems more consistently than anxiety and general internalizing related measures (Greenwood et al., 2021; Hussong et al., 2017; Ning et al., 2020). The nature of these relationships may also depend on how emotional problems are assessed, such as differentiating between symptoms, a spectrum of severity, clinical thresholds, or meeting diagnostic criteria (Dyer et al., 2019; Hussong et al., 2017; Ning et al., 2020).

To date, no study has quantitatively evaluated the potential impact of different researcher decisions on the significance and magnitude of the association between adolescent alcohol use and emotional problems. In the present study we therefore explore the overall and specific associations between adolescent alcohol use and emotional problems in a large sample of adolescents across Australia. We do this by applying a contemporary framework for quantifying sensitivity to alternative specifications known as multiverse analysis (Steenen et al., 2016), with complementary approaches existing within specification curve (Simonsohn et al., 2020) and vibration of effects analyses (Patel et al., 2015). These allow us to examine and report all non-redundant, reasonable, and justifiable measurement and analytic specifications, and identify the consequences of these specification decisions. Ultimately, these three analytical approaches have a shared goal of summarizing effects across various sets of sensitivity analyses based on varying research or design decisions. In brief, the underlying motivation of multiverse analysis (Steenen et al., 2016) is to explore effects across the 'multiverse' of possible combinations to increase transparent reporting and identify key choices or aspects of relationships. Within this framework, each unique combination of specifications is considered to be an analytic 'universe' (which corresponds to a single regression model), and here, alcohol use-emotional problems analyses are explored for each universe before interpreting the overall set of results. For specification curve analysis (Simonsohn et al., 2020), the focus is on visualizing the range of all estimated effects in a 'curve.' For vibration of effects analysis (Patel et al., 2015), the main focus is on exploring all justifiable sets of confounders and how different confounder adjustments impact effect sizes. All of these approaches are applied in a single dataset, so we can analyze and directly compare a range of specifications (or models) within the same dataset. Specifically, this study identifies: (1) the overall association between alcohol and emotional problems among adolescents using common measurement and modelling specifications; (2) specifications that yield the strongest (and weakest) association(s); (3) the impact of frequently referenced confounding variables; and (4) the magnitude of difference between associations modelled cross-sectionally *v.* prospectively.

Methods

Data

This is a secondary analysis of data from the Health4Life Study, a cluster-randomized controlled trial of a school-based eHealth intervention targeting lifestyle risk behaviors. The Health4Life study recruited 6639 students aged 11–14 (average 12.7) in 71 schools across three Australian states (New South Wales, Queensland, and Western Australia) (Teesson et al., 2020). Baseline data were collected in 2019 (T1) with follow-ups at post-intervention (~7 weeks, T2), 12 (T3), 24 (T4), and 36 months (T5). Given the peak age of onset for mental disorders is 14.5

years of age (Solmi et al., 2021) and the average age of onset of alcohol use among young Australian's was 16.2 in 2019 (Australian Institute of Health and Welfare, 2020), this paper focused on data collected when students were on average 14.7–15.7 years of age to maximize the prevalence and variability in alcohol use and emotional problems. Prospective analyses were conducted with 24-month data (T4, mean age = 14.7, ~year 9) predicting 36-month outcomes (T5, mean age = 15.7, ~year 10) and cross-sectional analyses were based on 36-month data (T5). The overall response rate was 75.4 and 66.8% at 24-months and 36-months, respectively. Further, given the intervention did not demonstrate effects in modifying alcohol use or emotional problems by 24-months (Champion et al., 2023; Smout et al., 2024), students participating in both control and intervention arms were included, adjusting for trial arm.

Parameters of interest and their specifications

Alcohol

Alcohol related specifications in this study included: past 6-months full standard drink (yes/no), past 6-months monthly or more drinking (yes/no), past 6-months frequency of drinking (never, <monthly, 1–2/month, 2–3/month, weekly, daily/almost daily), past 6-months binge drinking (yes/no), past 6-months monthly or more binge drinking (yes/no), past 6-months alcohol volume (frequency $\times$ quantity), alcohol-related harms as per a summative score on the Brief Rutgers Alcohol Problems Index (Earleywine, LaBrie, & Pedersen, 2008), and endorsing drinking a full standard drink in the past 6-months at any point in the study ≤ 14 years of age (yes/no).

Emotional problems

Four domains of emotional problems were captured continuously and dichotomously by four commonly used scales. First, non-specific psychological distress was measured using the 6-item Kessler-6 (K6) that asks about frequency of feeling nervous, hopeless, restless or fidgety, depressed, that everything was an effort, and worthless over the past 4 weeks (Kessler et al., 2002). Item responses were summed ranging from 0–24 where higher scores reflect greater distress, with scores ≥ 13 indicative of serious psychological distress. General internalizing problems were measured with the 5-item emotion symptoms subscale on the Strengths and Difficulties Questionnaire (SDQ-E; (Goodman, Meltzer, & Bailey, 1998). Scores were summed ranging from 0–10, with scores ≥ 7 indicative of a problematic level of symptoms. Symptoms of depression were measured with an adapted 8-item version of Patient Health Questionnaire for adolescents (PHQ-A) that asks about the frequency of depressive symptoms over the past 7 days (Johnson, Harris, Spitzer, & Williams, 2002). The 9th item regarding suicidal ideation was dropped as per requests from ethics; notably, previous evaluations indicate comparable results and psychometric properties for the 8- and 9-item versions (Wu et al., 2020). Item responses were summed ranging from 0–24 where higher scores indicate more symptoms, with scores ≥ 10 indicative of moderate to severe depressive symptoms. Symptoms of anxiety were measured with the 13-item PROMIS-Anxiety Pediatric Scale asking about frequency of symptoms over the past 7 days (Irwin et al., 2010). Item responses were summed ranging from 13–65 where higher scores indicate greater severity of anxiety, with scores ≥ 34 indicative of moderate to severe anxiety symptoms.

Confounders

Five sets of confounders were created based on the most commonly included confounders in studies included in related systematic reviews (Cochrane Canada, 2022; Dyer et al., 2019; Hussong et al., 2017; Ning et al., 2020) and the foundational goals and hypotheses of the Health4Life trial (Teesson et al., 2020). Sets of confounders include: (1) demographics, (2) smoking, (3) conduct problems, (4) ADHD symptoms, and (5) other health behaviors including physical activity, screen time, sleep, and diet. Additionally, in prospective models, T4 emotional problems were controlled for when examining associations between T4 alcohol and T5 emotional problems (e.g. controlling for autocorrelation, or pre-existing levels). Similarly, T4 alcohol variables were controlled for when examining association between T4 emotional problems and T5 alcohol-related outcomes. In total, we evaluated 16 different confounder combinations (i.e. all combinations where demographics are always included, apart from the unadjusted model). See Table 1 for more details.

Missing data

Similar to methods used in a prior multiverse analysis (Barendse et al., 2022), missing data were multiply imputed ($n = 20$) and then combined into a single dataset for inclusion in the multiverse analysis by taking the mean across imputations. Missing data was imputed using a multilevel fully conditional specification approach in BLIMP imputation software (Keller & Enders, 2021) with imputation models including all multiverse variables across all Health4Life timepoints. As such, a missing data specification was included to explore models with complete cases only v. multiple imputed data.

Statistical analysis

We conducted a multiverse analysis (Steege et al., 2016) supplemented with tools from similar statistical frameworks (i.e. specification curve analysis (Simonsohn et al., 2020) and vibration of effects (Patel et al., 2015)). Methods and code used for this paper come from previous multiverses (Barendse et al., 2022; Visontay et al., 2023) using R version 4.3.2 with packages *multiverse* (Sarma, 2023) and *spectr* (Masur & Scharkow, 2023). Two multiverses were estimated based on different implied directions of associations: (A) alcohol variables predicting emotional problems, and (B) emotional problems predicting alcohol variables. See details in Table 1. All of the reasonable and possible 'universes' are modeled and then pooled for interpretation, which is called a 'multiverse.' The total multiverse of specifications for Multiverse A included 4096 unique combinations of variables derived from all possible combinations of the available: 8 measures of alcohol use, 8 measures of emotional problems, 16 confounder combinations, 2 types of missing data, and 2 time points. Multiverse B had 3584 specifications as early alcohol use was not included as an outcome in this set of analyses. Each specification (otherwise known as a universe or an individual regression model) is analyzed separately through linear (for continuous outcome specifications) or logistic (for binary outcome specifications) regression with standard errors adjusted for school clustering.

All continuous variables were standardized before analysis to enable comparisons across specifications. Standardized effects of 0.01, 0.2, 0.5, and 0.8 are often interpreted as very small, small, medium, and large effects respectively (Matthay et al., 2021; Sullivan & Feinn, 2012). Given the scaling of residuals changes across nested logistic models depending on the model fit, which

Table 1. Summary of multiverse specifications

Alcohol Specifications		
1	Past 6-months any alcohol use	<i>'Have you had a full standard alcoholic drink (see alcohol chart) in the past 6 months?'</i> (0 = no, 1 = yes)
2	Past 6-months ~monthly alcohol use	<i>'How often did you have a standard alcoholic drink of any kind in the past 6 months?'</i> (0 = never or less than monthly, 1 = once a month or more)
3	Past 6-months frequency of alcohol use	Standardized: <i>'How often did you have a standard alcoholic drink of any kind in the past 6 months?'</i> (0 = never, 1 = less than monthly, 2 = once a month, 3 = 2-3 times, 4 = weekly, 5 = daily or almost daily)
4	Past 6-months binge drinking	<i>'How often did you have 5 or more standard alcoholic drinks on one occasion in the past 6 months?'</i> (0 = never, 1 = less than monthly or more)
5	Past 6-months ~monthly binge drinking	<i>'How often did you have 5 or more standard alcoholic drinks on one occasion in the past 6 months?'</i> (0 = never or less than monthly, 1 = once a month or more)
6	Past 6-months average monthly alcohol volume (in grams/centilitres)	Natural log transformed and standardized: Frequency $\times$ quantity $\times$ 10 g of alcohol per standard drink - Frequency: <i>'How often did you have a standard alcoholic drink of any kind in the past 6 months?'</i> 0 = never, 0.5 = less than monthly, 1 = once a month, 2.5 = 2-3 times, 4 = weekly, 28 = daily or almost daily (note: frequency coded as the mid-point of the response option) - Quantity: <i>'In the past 6 months, how many standard drinks did you have on a typical day when you were drinking alcohol?'</i> 0 = none, 1.5 = 1-2, 3.5 = 3-4, 5.5 = 5-6, 8 = 7-9, 10 = 10 + note: frequency coded as the mid-point of the response option)
7	Past 6-months alcohol problems	Standardized: Brief Rutgers Alcohol Problems Index score
8	Early age of onset	Any alcohol use in the past 6 months reported ≤ 14 years of age at any point in the study (0 = no, 1 = yes)
Emotional Problems Specifications		
1	Psychological distress symptoms	Standardized: K6 continuous score
2	High psychological distress	K6 ≥ 13 (0 = No, 1 = Yes)
3	Emotional symptoms	Standardized: SDQ-E emotional symptoms continuous score
4	Abnormal emotional symptoms	SDQ-E emotional symptoms ≥ 7 (0 = No, 1 = Yes)
5	Depressive symptoms	Standardized: PHQ-A continuous score
6	Moderate to severe depressive symptoms	PHQ-A ≥ 10 (0 = No, 1 = Yes)
7	Anxiety symptoms	Standardized: PROMIS-A continuous score
8	Moderate to severe anxiety symptoms	PROMIS-A ≥ 34 (0 = No, 1 = Yes)
Confounder Specifications		
Sets for analysis include: none (unadjusted), 1 (minimally adjusted), 1 + 2, 1 + 3, 1 + 4, 1 + 5, 1 + 2 + 3, 1 + 2 + 4, 1 + 2 + 5, 1 + 3 + 4, 1 + 3 + 5, 1 + 4 + 5, 1 + 2 + 3 + 4, 1 + 2 + 3 + 5, 1 + 3 + 4 + 5, 1 + 2 + 3 + 4 + 5		
1	Demographics	Sex, family affluence, cultural and linguistic diversity
2	Smoking	<i>'In the past 6 months, have you tried cigarette smoking, even one or two puffs?'</i> (0 = No, 1 = Yes)
3	Conduct	SDQ conduct problems subscale continuous score
4	ADHD	SDQ hyperactivity/inattention subscale continuous score
5	Health Behaviors	- Physical activity: Number of days engaging in 60 min or more of moderate to vigorous physical activity in the past 7 days (0-7). - Sedentary screen time: Average time (hours) in the past 7 days spent sitting or lying down while using an electronic device during free time, excluding homework. - Sleep: Whether adolescents were meeting sleep recommendations of 8-10 h per night based on mean sleep duration derived from a 6-item scale asking about bedtime, time attempted sleep, time taken to fall asleep, time awake, and time out of bed. - Sugar-sweetened beverages: Usual consumption of soft drinks, cordials, or sports drinks per week operationalized as 0 = <1 cup per week, 1 = 2-4 cups per week, 2 = 5 or more cups per week.
Missing data		
1	Complete case	Listwise deletion
2	Multiply imputed	Averaged across 20 imputations
Timing and implied direction of associations		
1A	Cross-sectional	T5 alcohol variables predicting emotional problem variables
2A	Prospective	T4 alcohol variables predicting T5 emotional problems, adjusting for T4 emotional problems
1B	Cross-sectional	T5 emotional problems variables predicting alcohol variables
2B	Prospective	T4 emotional problems predicting T5 alcohol variables, adjusting for T4 alcohol variables

To note, each 'universe' includes 1 alcohol specification, 1 emotional problem specification, 1 set of confounder(s), and 1 missing data strategy.

rescales coefficients accordingly (resulting in non-collapsibility; [Schuster, Twisk, Ter Riet, Heymans, & Rijnhart, 2021]), the log-odds obtained from logistic models were γ -standardized to ensure appropriate comparability across nested models (Williams & Jorgensen, 2023). This was accomplished through dividing the raw logistic regression coefficients by the estimated standard deviation of γ^* (i.e. the continuous latent dependent variable assumed to underly the dichotomous variable) (Huang, 2023).

First, the proportion of specifications where $p < 0.05$ are reported (Steenen et al., 2016). From the SCA framework, results are visualized through specification 'curves' that demonstrate the direction and (in)consistency in the magnitude of effects across specifications by plotting effects in order of magnitude (Simonsohn et al., 2020). From the VoE framework (Patel et al., 2015), volcano plots demonstrate the degree of (in)consistency in the direction of effects (provided in online Supplementary materials), and descriptive summary statistics are presented including the median beta, Range of Betas (RBs) depicting the range of standardized beta coefficients, the median p value (50th%), and the Range of $-\log_{10}(p)$ values (RPs) presenting the p values between the 1st and 99th percentiles. Larger ranges in both RBs and RPs suggest greater variability across the universe. Further, the variances in effects were decomposed by parameter specification, by: (1) calculating intra-class correlation coefficients (ICCs) for each overarching specification, and (2) plotting median and interquartile range of betas for each specification using box and whisker plots. These approaches protect against selective reporting and p hacking by presenting all justifiable specifications available.

Results

Descriptive characteristics

The sample includes 49% female and 12% culturally and linguistically diverse adolescents. The prevalence of most alcohol specifications nearly doubled from T4 to T5. Emotional problems also showed a slight increase between T4 and T5. See Table 2 for details.

Multiverse A: alcohol use predicting emotional problems

Overall, there was a considerable range of effects (RBs = 0.67; RPs = 29.38) with less than half of the models yielding significant alcohol effects (45.56% significant; 38.11% positive, 7.45% inverse) on emotional problems. Effects not only ranged in magnitude but also direction, with the full RBs from -0.41 to 0.69 (See Fig. 1a and Extended Data 4A).

Variability across these models was attributed to whether the model was specified cross-sectionally or prospectively (ICC = 0.30), the operationalization of emotional problems (ICC = 0.16) and alcohol use (ICC = 0.15), and the confounder adjustments (ICC = 0.11). Cross-sectional models consistently yielded larger significant positive effects while prospective models (after adjusting for prior levels of emotional problems)^{†1} were largely null with most significant effects found in the inverse direction (See Figs 2a, 3a, and Extended Data 4A). The alcohol specifications that had the smallest relative change in effects from cross-sectional to prospective specifications were: (1) any full drink in the past 6-months, and (2) early alcohol use. Operationalization

of emotional problems that focused on depression (i.e. PHQ-A), or had more depression than anxiety items (i.e. K6), yielded consistently larger significant positive effects compared to those focused on anxiety (i.e. PROMIS-A), or with more anxiety than depression items (i.e. SDQ-E). The only outcome that yielded more significant inverse effects (23–25%) than positive effects (13–17%) was the SDQ-E.

Further, when examining the median and IQR across confounder specifications (See Fig. 2a), adjusting for conduct symptoms had the largest impact on mitigating the magnitude of alcohol effects, followed by ADHD symptoms and smoking. When adjusting for either no confounders or only demographic confounders, 61% of effects across all specifications were significantly positively related (98% of cross-sectional, 24% of prospective; See Fig. 3a and Extended Data 4A). In fully adjusted models, only 27% were significantly related with just over half of these effects (14% overall) being inversely related; for cross-sectional 34% were significant in fully adjusted models with 16% positively and 19% inversely related, while in prospective models 20% were significant with 9% positively and 11% inversely related.

Multiverse B: emotional problems predicting alcohol use

Overall, there was a considerable range of effects (RBs = 0.57; RPs = 20.65) with only half of the models yielding significant emotional problem effects (50.14% significant; 45.54% positive, 4.6% inverse) on alcohol use. Again, effects ranged in both magnitude and direction, with the full RBs going from -0.31 to 0.69 (See Fig. 1b and Extended Data 4B).

Variability across these models was attributed to whether the model was specified cross-sectionally or prospectively (ICC = 0.22), the operationalization of emotional problems (ICC = 0.25), and the confounder adjustments (ICC = 0.25). In this multiverse, the operationalization of alcohol use had little impact on the variability of effects across models (ICC = 0.04). Here, cross-sectional models consistently yielded larger significant positive effects while prospective models yielded attenuated effects (after adjusting for prior levels of alcohol use) that were largely null (See Figs 2b, 3b, and Extended Data 4B). Most depression-focused emotional problem specifications (i.e. total and binary PHQ-A, and total K6 specifications) still yielded ~40% significant positive effects in prospective models and <1% of any model with a depression-focused emotional specification had significant inverse effects. Effects related to anxiety-focused specifications were mixed and inconsistent. Similar to multiverse A, SDQ-E specifications yielded a non-negligible proportion of significant inverse effects in both cross-sectional and prospective models (8–24%). However, while cross-sectional effects related to PROMIS-A specifications were attenuated in prospective models with some significant inverse effects (3–9%), the pattern of associations with PROMIS-A as compared to SDQ-E scores were quite distinct.

Further, when examining the median and IQR across confounder specifications (See Fig. 2b), it appears as though adjusting for conduct symptoms had the largest impact on mitigating the magnitude of emotional problem effects, followed by ADHD symptoms. When adjusting for either no confounders or only demographic confounders, 88% of effects across all specifications were significant positively related (99% of cross-sectional, 77% of prospective; See Fig. 3b and Extended Data 4B). In fully adjusted models, only 28% were significantly related with a majority of these effects (19% overall) being inversely related; for cross-

[†]The notes appear after the main text.

Table 2. Descriptive Alcohol Use and Emotional Problem Sample Characteristics

	Complete Case				Multiply Imputed (mean across imputations)	
	T4	%missing from total T4 sample	T5	%missing from total T5 sample	T4	T5
Alcohol Use						
Past 6-months any alcohol use	16.0%	29.2%	29.9%	36.4%	17.4%	38.2%
Past 6-months ~monthly alcohol use	7.3%		14.1%		5.3%	9.5%
Past 6-months frequency of alcohol use	0.3 (0.9)		0.5 (1.1)		0.3 (0.8)	0.6 (0.9)
Past 6-months binge drinking	7.0%	29.7%	14.4%	37.0%	7.0%	14.2%
Past 6-months ~monthly binge drinking	3.5%		6.7%		2.6%	4.4%
Past 6-months average monthly alcohol volume (grams/centiliters, log transformed)	0.4 (1.3)	29.2%	0.9 (1.7)	36.4%	0.5 (1.1)	0.9 (1.4)
Past 6-months alcohol problems	1.4 (4.9)	29.3%	1.9 (5.3)	36.5%	1.5 (4.2)	2.1 (4.3)
Early alcohol use (≤ 14 years of age)	12.3%	15.0%	12.3%	14.9%	12.5%	12.6%
Emotional Problems						
Psychological distress symptoms (K6)	8.1 (6.2)	29.9%	8.4 (6.4)	36.8%	8.2 (5.5)	8.7 (5.5)
High psychological distress (K6 ≥ 13)	22.5%		23.8%		19.6%	21.0%
Emotional symptoms (SDQ-E)	3.4 (2.8)	31.8%	3.9 (2.7)	39.0%	3.5 (2.5)	3.9 (2.3)
Abnormal emotional symptoms (SDQ-E ≥ 7)	15.3%		18.2%		12.2%	14.0%
Depressive symptoms (PHQ-A)	6.6 (6.5)	30.4%	6.8 (6.5)	37.2%	6.7 (5.7)	7.1 (5.5)
Moderate to severe depressive symptoms (PHQ-A ≥ 10)	27.5%		28.1%		25.2%	27.7%
Anxiety symptoms (PROMIS-A)	24.9 (12.7)	30.6%	25.2 (12.6)	37.3%	25.1 (11.2)	25.8 (10.8)
Moderate to severe anxiety symptoms (PROMIS-A ≥ 34)	22.9%		24.2%		20.0%	21.3%

sectional 36% were significant in fully adjusted models with 17% positively and 19% inversely related, while in prospective models 18% were significant with nearly all inversely related.

Discussion

This study explored nearly 8000 different ways of modelling the relationship between alcohol use and emotional problems, leveraging a recent sample of over 6000 adolescents. By using contemporary statistical frameworks to compare various specifications of these relationships within the same sample, we can draw conclusions about the co-occurrence of alcohol use and emotional problems among adolescents (regardless of measurement) and identify which measurement and analysis choices made by researchers impact findings. Echoing inconsistencies observed across different samples and studies in existing literature (Cochrane Canada, 2022; Dyer *et al.*, 2019; Greenwood *et al.*, 2021; Hussong *et al.*, 2017; McCabe *et al.*, 2023; Ning *et al.*, 2020; Puddephatt *et al.*, 2022; Watson *et al.*, 2022), this multiverse analysis found notable inconsistencies within a single sample, depending on specifications. Methodologically, researcher decisions that appeared to have the biggest impact on findings included the operationalization of emotional problems, temporality of relationships, and choice of confounders. Researcher decisions that had minimal impact on findings were related to

missing data strategies and whether outcomes were modelled continuously *v.* dichotomously. Inconsistencies in the magnitude, direction, and significance of effects between alcohol use and emotional problems appear closely tied to researcher decisions that are often regarded as relatively arbitrary.

The operationalization of emotional problems impacted the direction and magnitude of effects found, regardless of the direction of analysis, with depression-related measures more consistently positively related to alcohol use than anxiety-related measures. This suggests that there may be distinct associations between alcohol use and depression *v.* anxiety, indicating a broad ‘internalizing’ factor may inadequately capture these relationships during mid-adolescence. This is similar to previous studies that have found more consistent positive associations between adolescent alcohol and depression when compared to anxiety or general internalizing measures (Greenwood *et al.*, 2021; Hussong *et al.*, 2017; Ning *et al.*, 2020). This may be because adolescent alcohol use typically occurs in a social context (Brooks-Russell, Simons-Morton, Haynie, Farhat, & Wang, 2014), which may be a barrier for use among adolescents with high levels of anxiety. The specific measures also seemed to operate differently, particularly the SDQ-E. However, whether emotional problems were operationalized using symptom scores or binary clinical cut-points did not play a large role in the magnitude and significance of effects.

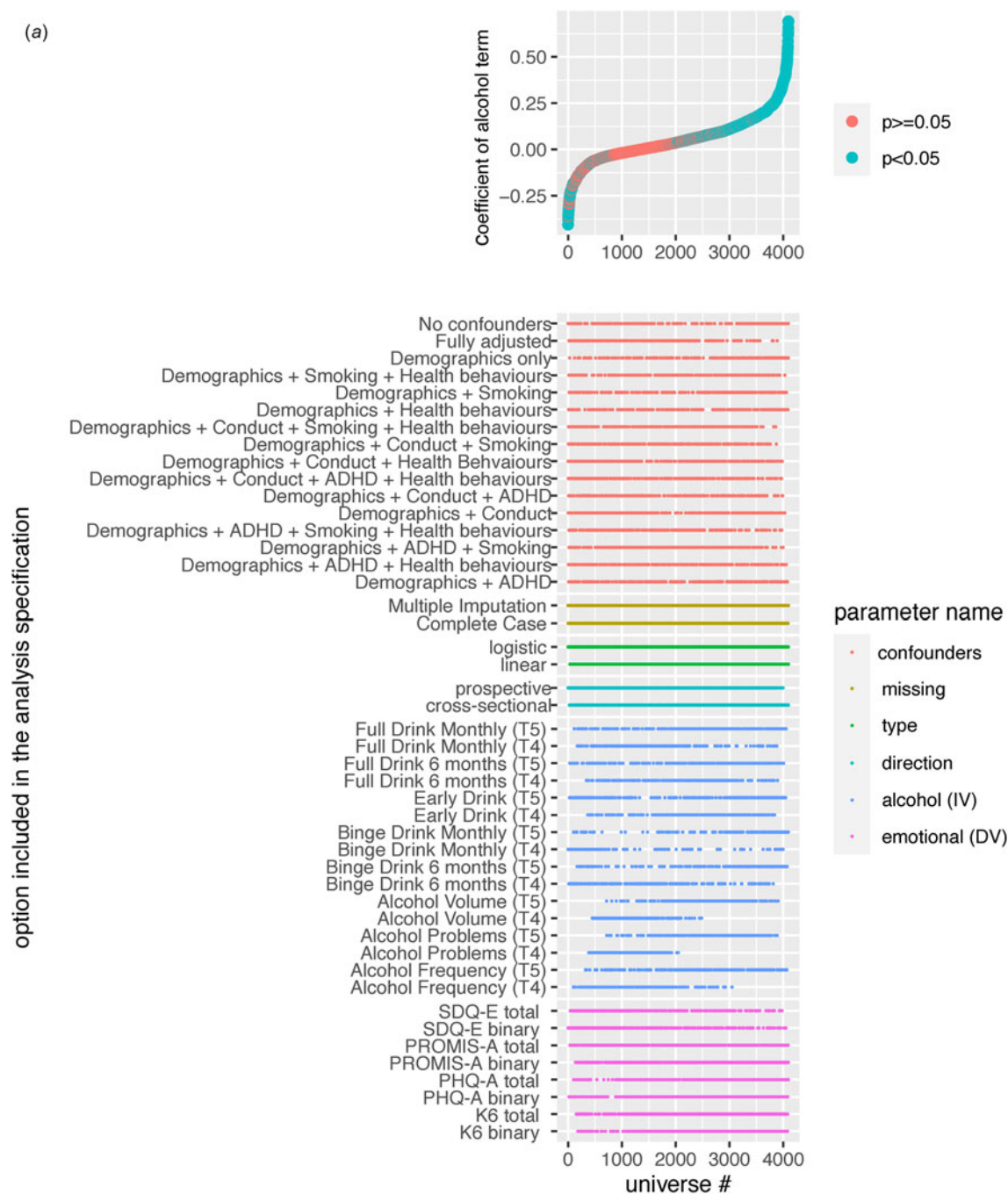


Figure 1. (a) Specification Curve Multiverse A. The specification curve plot in the top part of the figure plots standardized alcohol parameter estimates from each universe in multiverse A from smallest to largest. The blue dots represent parameter estimates that are statistically significant at a $p < 0.05$. The red dots represent parameter estimates that are not statistically significant. Dots below “0” indicate models in which higher levels of the alcohol variable are associated with lower emotional problems. Dots above “0” indicate models in which higher alcohol is associated with higher emotional problems. The detail below the graph shows the universe where each specification listed on the left is included (i.e., the dot in the top figure lines up with dots in the detailed section below to indicate which specifications were in that particular model).

The operationalization of alcohol use appeared to impact effects when exploring whether alcohol use predicted emotional problems, but not vice versa. Where emotional problems predicted alcohol use, there appeared to be general effects whereby the type, pattern, or measure of alcohol use did not seem to meaningfully impact the direction, magnitude, and significance of results. However, when alcohol use predicted emotional problems, associations across different operationalizations became more nuanced. For example,

cross-sectionally, binge drinking had the strongest association with emotional problems, though these largely became null (or inverse) when prospectively explored. While it is possible that binge drinking may confer social and emotional benefits due to typical use in social settings (Brooks-Russell et al., 2014), most prospective models were null (83–87% null) and co-occurrence cross-sectionally was common. Prospectively, any past 6-month full standard drink consumption and early drinking seemed to remain

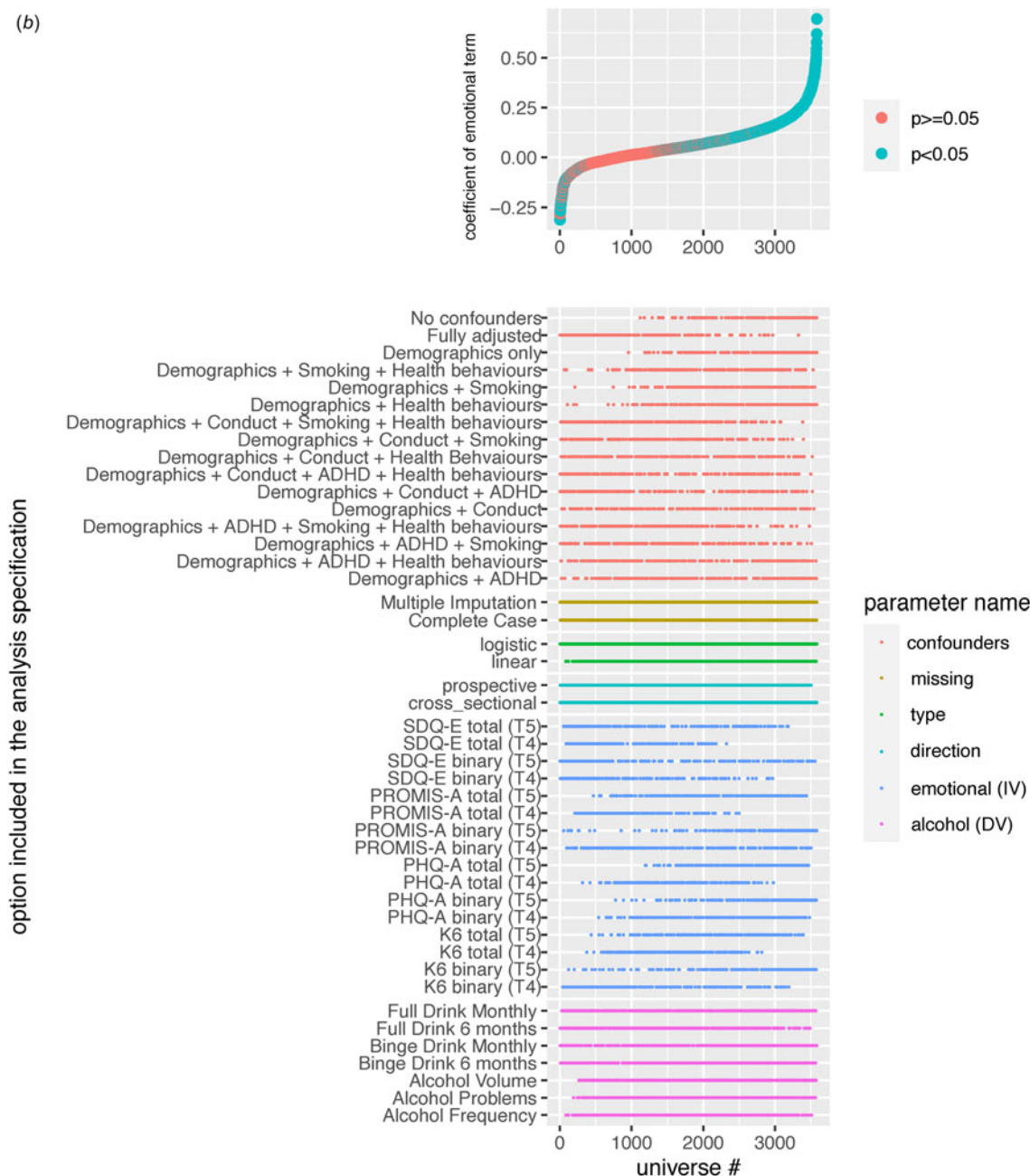


Figure 1. Specification Curve Multiverse B. The specification curve plot in the top part of the figure plots standardized emotional problems parameter estimates from each universe in multiverse B from smallest to largest. The blue dots represent parameter estimates that are statistically significant at a $p < 0.05$. The red dots represent parameter estimates that are not statistically significant. Dots below “0” indicate models in which higher values on the emotional problems variable are associated with lower values on the alcohol variable. Dots above “0” indicate models in which higher emotional problems is associated with higher values on the alcohol variable. The detail below the graph shows the universe where each specification listed on the left is included (i.e., the dot in the top figure lines up with dots in the detailed section below to indicate which specifications were in that particular model).

the most consistently, positively related to later emotional problems (though still $<50\%$ significant). While previous literature has suggested differential associations depending on the operationalization of alcohol use (Dyer et al., 2019; Hussong et al., 2017; Puddephatt et al., 2022), this may be impacted by the stage of development (e.g. stronger prospective relationships earlier in adolescence) and direction of specified effects (Lees, Meredith, Kirkland, Bryant, & Squeglia, 2020; Spear, 2018).

The strongest sources of inconsistency across models related to the temporality, directionality, and included confounders of the

specified model. Nearly universally, alcohol use and emotional problems cross-sectionally co-occurred (e.g. 98–99% of minimally adjusted cross-sectional models had significant, positive effects). Prospectively, alcohol use did not typically predict emotional problems 1 year later after accounting for baseline emotional problems (e.g. 76% null effects), and while prospective associations between emotional problems and alcohol use 1 year later were more consistent (e.g. 77% positive significant effects), these effects were strongly influenced by other researcher decisions (e.g. confounders, operationalization of outcomes). This aligns with

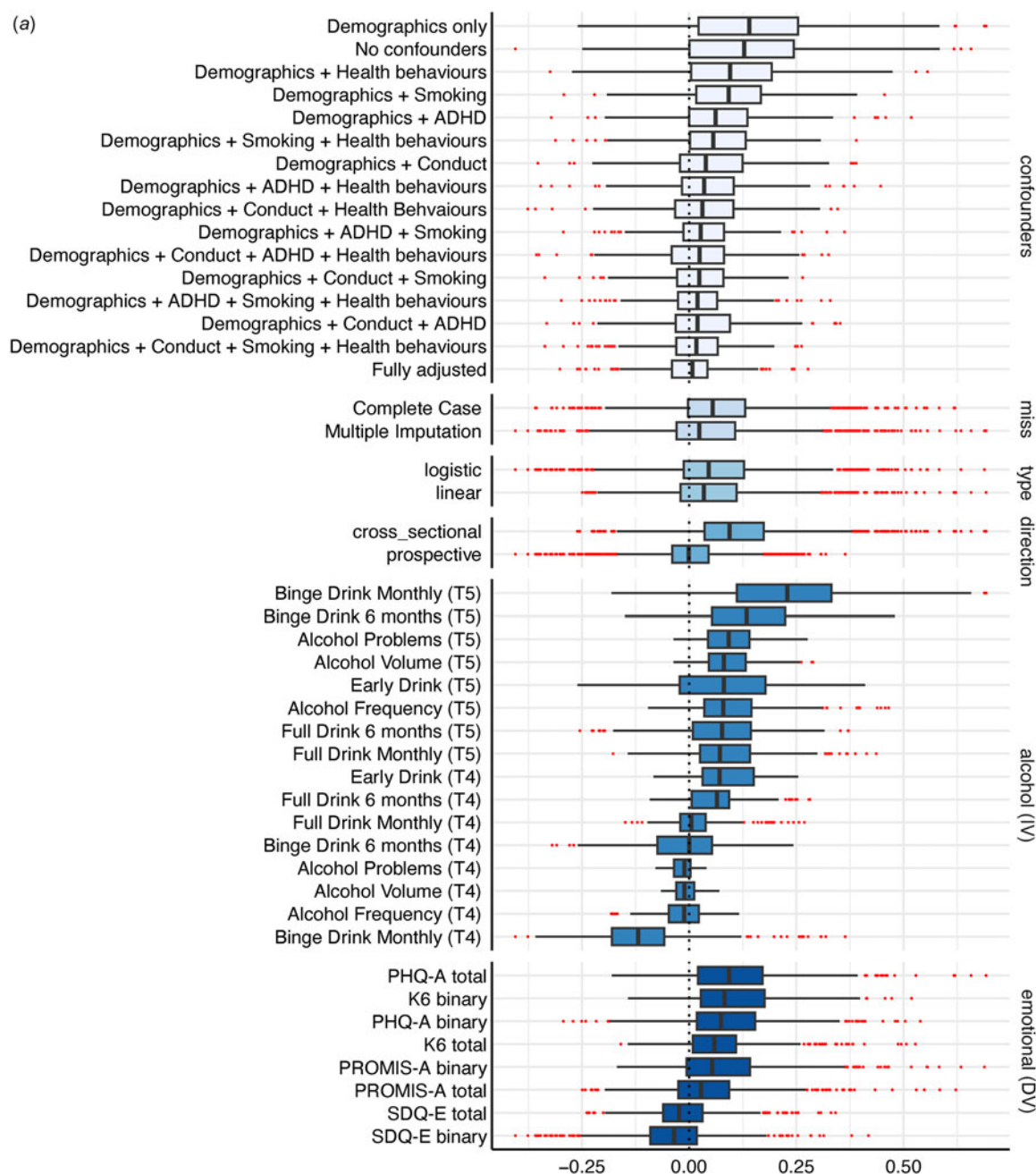


Figure 2. (a) Box Plot Multiverse A. The box plot shows the median eLect, interquartile range, and full range of all universes (or models) where the specification indicated on the left side of the figure was included. For example, Binge Drink Monthly (T5) has a black line at the median eLect of 0.23, reflecting the median eLect across all crosssectional specifications where Binge Drinking Monthly was included. The blue box around the median eLect is bound by the eLect at the 25th% (here, 0.11) and the 75th% (here, 0.33). The whiskers reflect the lower and upper data points within 1.5*the interquartile range. For Binge Drink Monthly (T5) this is -0.18 (as that is the lowest observed value) at the lower end and 0.66 at the upper end. The red dots reflect outliers beyond these edges.

existing literature suggesting weak or null prospective relationships between alcohol and emotional problems (Cochrane Canada, 2022; McCabe et al., 2023) with more evidence, though still nuanced, for emotional problems predicting alcohol use (Dyer et al., 2019; Hussong et al., 2017). What was evident was that the relationship between adolescent alcohol use and emotional problems was strongly influenced by choice of confounders, which may be due to true confounding or overcontrolling of higher-level constructs (e.g. 72–73% of fully adjusted models

yielded null effects). For example, externalizing symptoms (e.g. conduct and ADHD symptoms) and other substance use (i.e. smoking) explained a large proportion (or all) of the shared variance between alcohol use and emotional problems.

Observational data presents a significant challenge due to the potential risk of residual confounding. The selection of appropriate confounders is thus crucial, though often underappreciated, researcher decision (Digitale, Martin, & Glymour, 2022; Herbert, 2020; Von Elm et al., 2007). There is both a risk for

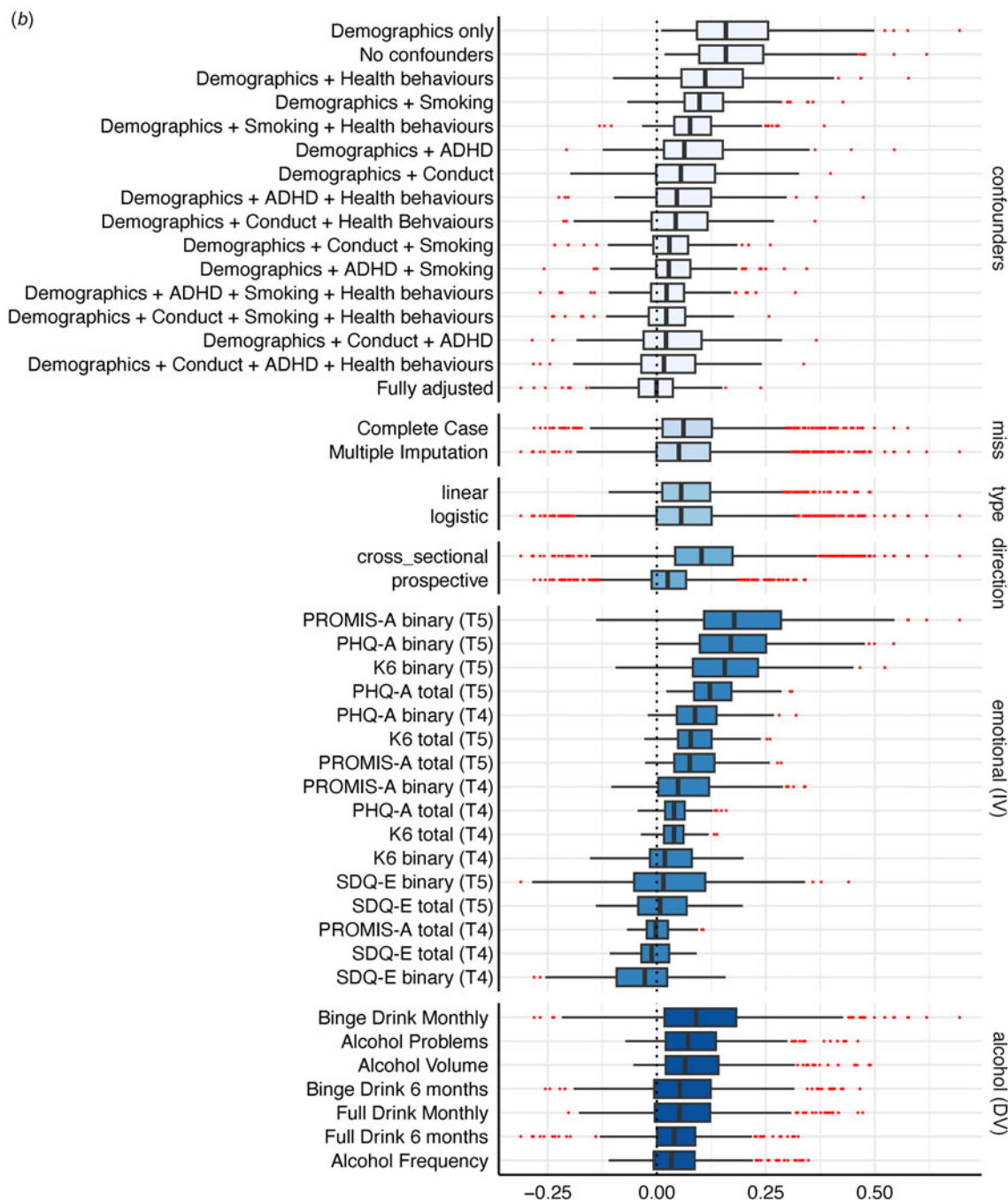


Figure 2. The box plot shows the median eLect, interquartile range, and full range of all universes (or models) where the specification indicated on the left side of the figure was included. For example, PHQ-A binary (T5) has a black line at the median eLect of 0.17, reflecting the median eLect across all cross-sectional specifications where PHQ-A binary was included. The blue box around the median eLect is bound by the eLect at the 25th% (here, 0.10) and the 75th% (here, 0.25). The whiskers reflect the lower and upper data points within 1.5*the interquartile range. For PHQ-A binary (T5) this is 0.002 at the lower end (as that is the lowest observed value) and 0.48 at the upper end. The red dots reflect outliers beyond these edges.

under- and over-controlling, with the ultimate goal to control for shared causal factors (e.g. factors that, when unaccounted for, suggest causal relationships that do not exist) while avoiding controlling for factors that lie on the causal pathway (e.g. mediators, or factors that explain causal relationships that do exist) (Herbert, 2020). While the selection of confounders for this multiverse adhered to established guidelines, in that they were selected *a priori* based on existing literature where they are commonly

included and justified as confounders (Hernán, 2018; Larsson, 2022; Lederer et al., 2019), it is possible some specifications are at risk for over-controlling. For example, conduct symptoms, ADHD symptoms, and smoking appeared to explain a substantial amount of the relationship between alcohol use and emotional problems. While some researchers consider these critical confounders for alcohol-emotional relationships (Hussong et al., 2017), others view them as subdomains of overarching constructs

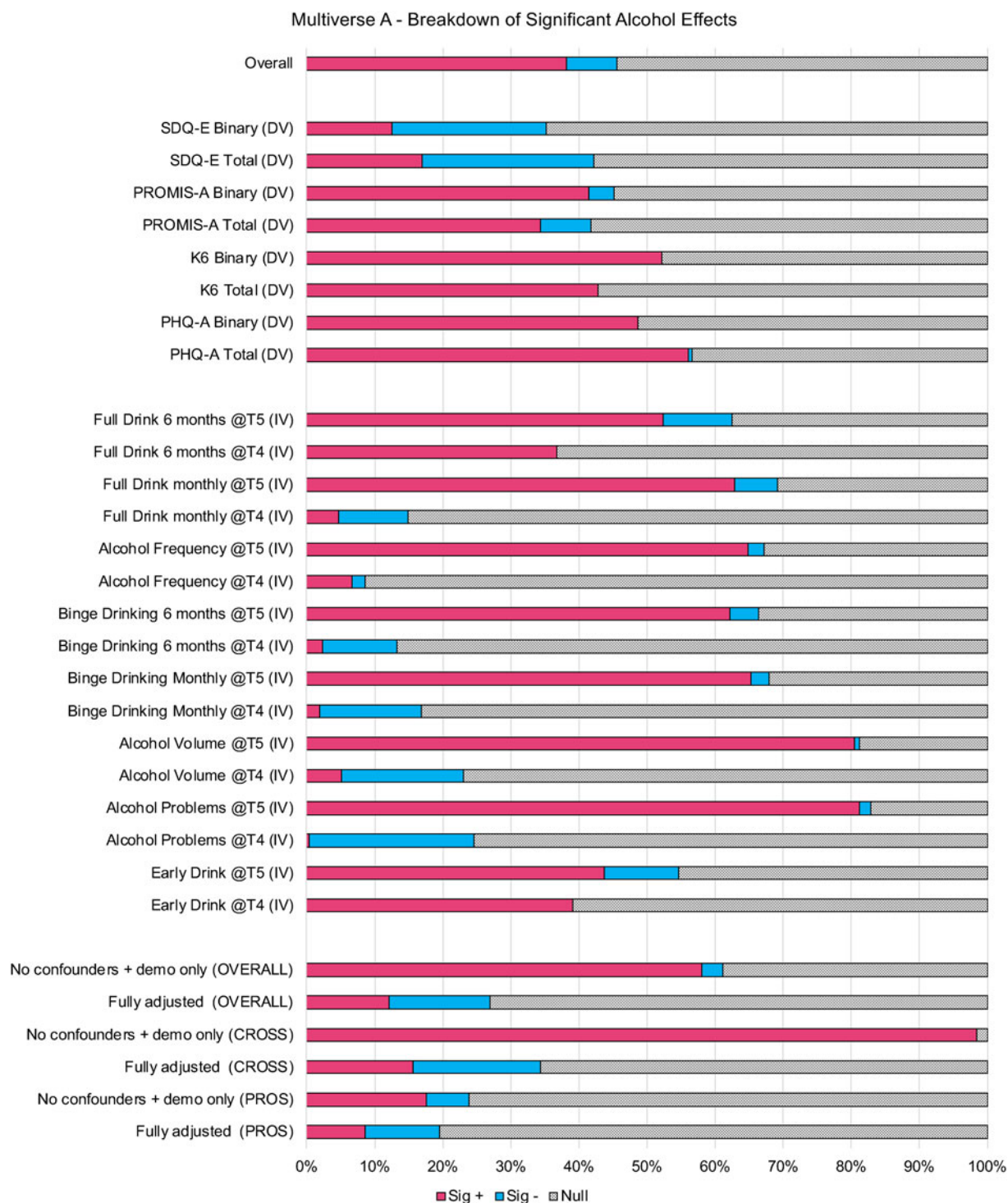


Figure 3. (a) *p* value Plot Multiverse A. This figure shows the proportion of statistically significant alcohol effects ($p < 0.05$) that are positively related to emotional problems (Sig +) and inversely related to emotional problems (Sig -) of all universes (or models) where the specification indicated on the left side of the figure was included.

(e.g. general externalizing, general substance use) suggesting controlling for these factors may inadvertently hide causal effects (Krueger et al., 2021; Vanyukov et al., 2012). To note, the aforementioned nuanced differences for anxiety and depression (*v.* general internalizing) provide evidence against higher-order constructs (Watson et al., 2022)—thus, further exploration into

specific lower and higher-order constructs across different developmental stages is needed.

Two types of modeling specifications that often receive considerable attention by epidemiologists had minimal to no impact on effects in these data: (1) the approach to dealing with missing data, and (2) logistic *v.* linear models. First, the multiverse

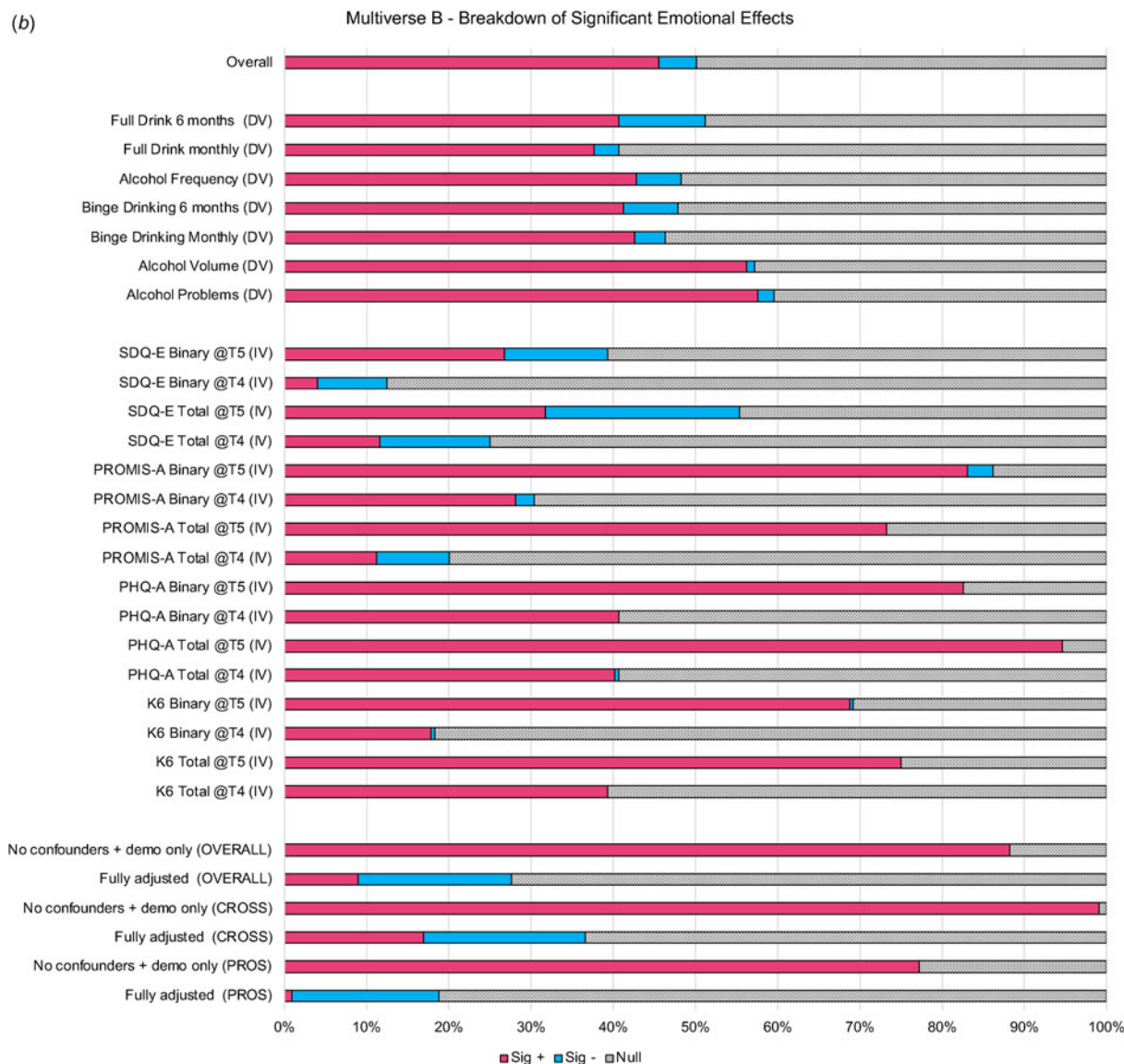


Figure 2. P-value Plot Multiverse B. This figure shows the proportion of statistically significant emotional problems (lects ($p < 0.05$) that are positively related to alcohol variables (Sig +) and inversely related to alcohol variables (Sig -) of all universes (or models) where the specification indicated on the left side of the figure was included.

explored what is typically regarded as poor practice (i.e. complete cases analysis) compared to best practice (i.e. multiple imputation) (Enders, 2017). In our sample, for this question, how we treated missing data only explained <2% of the variability in the magnitude and direction of effects across specifications. As such, the effects did not seem to be biased depending on missing data practices and/or we did not have good enough auxiliary variables to predict missing data (e.g. it is possible data were missing not at random [MNAR]). There are other contemporary missing data approaches that may have been better suited to filling in the missing data in this sample, though these are often model-specific making them not feasible to apply to a multiverse of models (Enders, 2023). Second, once logistic effect estimates were y^* -standardized, the model function had largely null effect on the magnitude of effects across models. Notably, when y^* -standardization was not applied, logistic models seemingly yielded significantly

larger effects than linear models. As such, while correct interpretation of beta coefficients from logistic models is critical (Williams & Jorgensen, 2023), dichotomization of variables of interest did not appear to lose meaningful explanatory power.

There are limitations to consider when interpreting this multiverse analysis. First, while the Health4Life dataset is a large, contemporary sample spanning three Australian states, it is not a representative sample. Second, data were collected in 2021 and 2022; though 2022 was not impacted by COVID19 lockdowns, there were a number of lockdowns in 2021 which may have impacted prospective associations. Third, these analyses give equal weighting to all specifications (Simonsohn et al., 2020) and there are several circumstances where the same question was operationalized in multiple different ways (e.g. full alcohol drink past 6 months, past month, frequency). Fourth, several significant findings may be a reflection of type

1 errors due to multiple testing; as these approaches look at overall distributions of effects and p values, rather than specific models, this is not a major concern. Fifth, this multiverse does not include *all* conceivable specifications of this relationship. For example, the Health4Life study did not include any diagnostic assessments of emotional disorders, measures of suicidality, or specific types of anxiety disorders. We also only explored one cut-off for each measure. Certain confounders were also not measured in Health4Life and thus could not be evaluated such as cannabis use, e-cigarette use, or family history. We also did not explore other types of covariates, such as moderators or mediators and only used one (cross-sectional) or two (prospective) time points – reflecting the most commonly evaluated models in the broader literature – precluding adjusting for confounders measured prior to the exposure and time-varying confounders (Cinelli, Forney, & Pearl, 2022; Clare, Dobbins, & Mattick, 2019). Notably, these multiverse models did not follow a formal causal inference framework, as the goals were to explore the impact of different specifications that are and have been commonly reported on in the literature. Further, outcomes were analyzed using linear and logistic models, though it is possible other approaches may have fit certain specifications better (e.g. Poisson, log-binomial).

The key findings of this multiverse are: (1) alcohol use and emotional problems commonly co-occur among adolescents; (2) emotional problems, particularly depression, may predict later alcohol use among adolescents, but there is limited evidence for the reverse; (3) conduct symptoms, ADHD symptoms, and smoking explain most of the associations between adolescent alcohol use and emotional problems; and (4) researcher decisions related to the operationalization of variables, inclusion of confounders, and choice in temporality and underlying causal hypotheses influence the magnitude, direction, and significance of relationships found between adolescent alcohol use and emotional problems. Whether these are causal relationships requires the application of formal causal modelling approaches. Based on the current findings, practitioners and policymakers should consider both: (1) the most consistent findings (i.e. co-occurrence was common and emotional problems often preceded alcohol use among adolescents) and (2) the degree of inconsistency in findings and possible reasons for inconsistency where it exists (i.e. the relationships may be explained by other substance use or behavioral problems). This enables cautious and accurate interpretations, while remaining open to further research to clarify our understanding. The contemporary analytical multiverse framework used in this paper needs to be applied to different samples of adolescents to uncover methodological and substantive reasons for the current inconsistencies in the literature.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291724002502>

Acknowledgments. We sincerely thank Abhraneel Sarma, developer of the *multiverse* R package, who provided technical and troubleshooting support for our analyses. We also thank the Health4Life Team (collaborative group) including Nicola C Newton, Tim Slade, Katherine Mills, Matthew Sunderland, Belinda Partmenter, Bonnie Spring, David Lubans, Steve Allsop, Leanne Hides, Nyanda McBride, Lexine Stapinski, and Louise Birrell.

Author contributions. All authors contributed to the conceptualization, interpretation, and editing of this manuscript. JH led the conceptualization, design, data analysis, and writing of this manuscript. RV co-analyzed the data, providing support and verification on multiverse methods and code. ED and SS supported and verified data cleaning. TS and MS provided

substantive and methodological supervision. All authors had full access to the data in the study and all had final responsibility for the decision to submit for publication.

Funding statement. JH is funded by a Health Systems Impact Embedded Early Career Researcher award co-funded by the Canadian Institutes of Health Research, McMaster University, and St. Joseph's Healthcare Hamilton (HS3-191640). JLA is supported by a Wellcome Trust Early-Career Award (227640/Z/23/Z). KC is funded by a University of Sydney Horizon Fellowship. The Health4Life Study was funded by the Paul Ramsay Foundation and the Australian National Health and Medical Research Council (Fellowship to KC, APP1120641; MT, APP1078407; Centre of Research Excellence Grant in Prevention and Early Intervention in Mental Illness and Substance Use, PREMISE, APP1134909). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests. MT is the co-Director of Climate Schools Pty Ltd and OurFutures Institute Ltd. There are no other potential conflicts of interest to report.

Ethical standards. The Health4Life study has been approved by the University of Sydney Human Research Ethics Committee (2018/882), University of Queensland Human Research Ethics Committee (2019000037), Curtin University Human Research Ethics Committee (HRE2019-0083), NSW State Education Research Applications Process (2019006), Catholic Education Diocese of Bathurst, Catholic Schools Office Diocese of Maitland-Newcastle, Edmund Rice Education Australia, Brisbane Catholic Education Committee (373), and Catholic Education Western Australia (RP2019/07).

Code availability. The code for this analysis is available open access of open science framework at DOI 10.17605/OSF.IO/G2EQD.

Note

¹ Post-hoc exploratory analysis showed that when prior levels of emotional problems were not adjusted for in prospective models, alcohol variables were largely positively related to emotional problems. After adjusting for prior levels of emotional problems, they largely reversed in direction of effects.

References

- Ahuja, M., Awasthi, M., Records, K., & Lamichhane, R. R. (2021). Early age of alcohol initiation and its association with suicidal behaviors. *Substance Use & Misuse*, 56(9), 1332–1338. doi: 10.1080/10826084.2021.192245
- Australian Institute of Health and Welfare. (2020). *National drug strategy household survey 2019*. Canberra. doi: 10.25816/e42p-a447.
- Barendse, M. E., Byrne, M. L., Flournoy, J. C., McNeilly, E. A., Guazzelli Williamson, V., Barrett, A.-M. Y., ... Pfeifer, J. H. (2022). Multimethod assessment of pubertal timing and associations with internalizing psychopathology in early adolescent girls. *Journal of Psychopathology and Clinical Science*, 131(1), 14. doi: 10.1037/abn0000721
- Brooks-Russell, A., Simons-Morton, B., Haynie, D., Farhat, T., & Wang, J. (2014). Longitudinal relationship between drinking with peers, descriptive norms, and adolescent alcohol use. *Prevention Science*, 15(4), 497–505. <https://doi.org/10.1007/s11121-013-0391-9>
- Champion, K. E., Newton, N. C., Gardner, L. A., Chapman, C., Thornton, L., Slade, T., ... O'Dean, S. (2023). Health4Life eHealth intervention to modify multiple lifestyle risk behaviours among adolescent students in Australia: A cluster-randomised controlled trial. *The Lancet Digital Health*, 5(5), e276–e287. doi: 10.1016/S2589-7500(23)00028-6
- Cinelli, C., Forney, A., & Pearl, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research*, 53(3), 1071–1104. <https://doi.org/10.1177/00491241221099552>
- Clare, P. J., Dobbins, T. A., & Mattick, R. P. (2019). Causal models adjusting for time-varying confounding – a systematic review of the literature.

- International Journal of Epidemiology*, 48(1), 254–265. <https://doi.org/10.1093/ije/dyy300>
- Cochrane Canada. (2022). Update of a systematic review of the effect of alcohol consumption on the development of depression, anxiety and suicidal ideation. <https://ccsa.ca/effect-alcohol-consumption-development-depression-anxiety-and-suicidal-ideation-update-systematic>
- Digitale, J. C., Martin, J. N., & Glymour, M. M. (2022). Tutorial on directed acyclic graphs. *Journal of Clinical Epidemiology*, 142, 264–267. doi: 10.1016/j.jclinepi.2021.08.001
- Dyer, M. L., Easey, K. E., Heron, J., Hickman, M., & Munafò, M. R. (2019). Associations of child and adolescent anxiety with later alcohol use and disorders: A systematic review and meta-analysis of prospective cohort studies. *Addiction*, 114(6), 968–982. doi: 10.1111/add.14575
- Earleywine, M., LaBrie, J. W., & Pedersen, E. R. (2008). A brief Rutgers alcohol problem Index with less potential for bias. *Addictive Behaviors*, 33(9), 1249–1253. doi: 10.1016/j.addbeh.2008.05.006
- Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour Research and Therapy*, 98, 4–18. doi: 10.1016/j.brat.2016.11.008
- Enders, C. K. (2023). Missing data: An update on the state of the art. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000563>
- Gardner, L. A., Stockings, E., Champion, K. E., Mather, M., & Newton, N. C. (2024). Alcohol initiation before age 15 predicts earlier hazardous drinking: A survival analysis of a 7-year prospective longitudinal cohort of Australian adolescents. *Addiction*, 119(3), 518–529. <https://doi.org/10.1111/add.16376>
- Goodman, R., Meltzer, H., & Bailey, V. (1998). The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry*, 7(3), 125–130. doi: 10.1007/s007870050057
- Greenwood, C., Youssef, G., Fuller-Tyszkiewicz, M., Letcher, P., Macdonald, J., Hutchinson, D., ... Biden, E. (2021). Psychosocial predictors of binge-drinking residual harm in adolescence and young adulthood: Findings from the Australian temperament project. *Drug and Alcohol Dependence*, 226, 108864. doi: 10.1016/j.drugalcdep.2021.108864
- Halladay, J., MacKillop, J., Munn, C., Amlung, M., & Georgiades, K. (2022). Individual- and school-level patterns of substance use and mental health symptoms in a population-based sample of secondary students: A multilevel latent profile analysis. *Drug and Alcohol Dependence*, 240, 109647. doi: 10.1016/j.drugalcdep.2022.109647
- Hawke, L. D., Koyama, E., & Henderson, J. (2018). Cannabis use, other substance use, and co-occurring mental health concerns among youth presenting for substance use treatment services: Sex and age differences. *Journal of Substance Abuse Treatment*, 91, 12–19. doi: 10.1016/j.jsat.2018.05.001
- Herbert, R. D. (2020). Research note: Causal inference. *Journal of Physiotherapy*, 66(4), 273–277. doi: 10.1016/j.jphys.2020.07.010
- Hernán, M. A. (2018). The C-word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health*, 108(5), 616–619. doi: 10.2105/AJPH.2018.304337
- Huang, F. L. (2023). Comparing coefficients across logistic regression models. Retrieved February from <https://franchish.net/post/comparing-coefficients-across-logistic-regression-models/#:~:text=Y%20standardization&text=Basically%2C%20this%20takes%20the%20variance,which%20is%20the%20standard%20deviation>
- Hussong, A. M., Ennett, S. T., Cox, M. J., & Haroon, M. (2017). A systematic review of the unique prospective association of negative affect symptoms and adolescent substance use controlling for externalizing symptoms. *Psychology of Addictive Behaviors*, 31(2), 137. doi: 10.1037/adb0000247
- Irwin, D. E., Stucky, B., Langer, M. M., Thissen, D., DeWitt, E. M., Lai, J.-S., ... DeWalt, D. A. (2010). An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Quality of Life Research*, 19, 595–607. doi: 10.1007/s11136-010-9619-3
- Johnson, J. G., Harris, E. S., Spitzer, R. L., & Williams, J. B. (2002). The patient health questionnaire for adolescents: Validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *Journal of Adolescent Health*, 30(3), 196–204. doi: 10.1016/s1054-139x(01)00333-0
- Keller, B. T., & Enders, C. K. (2023). *Blimp user's guide (Version 3)*. Retrieved from www.appliedmissingdata.com/blimp
- Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L., ... Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959–976. doi: 10.1017/s0033291702006074
- Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., ... Latzman, R. D. (2021). Validity and utility of hierarchical taxonomy of psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry*, 20(2), 171–193. doi: 10.1002/wps.20844
- Larsson, H. (2022). Transparent and complete reporting of confounding in observational research. In (Vol. 2, pp. e12086): Wiley Online Library. doi: 10.1002/jcv2.12086
- Lederer, D. J., Bell, S. C., Branson, R. D., Chalmers, J. D., Marshall, R., Maslove, D. 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(1), 22–28. doi: 10.1513/AnnalsATS.201808-564PS
- Lee, Y. Y., Slade, T., Chatterton, M. L., Le, L. K.-D., Perez, J. K., Faller, J., ... Mihalopoulos, C. (2024). Age at first drink and its influence on alcohol use behaviours in young adulthood: Evidence from an Australian household-based panel study. *Preventive Medicine*, 107898. Online ahead of print. <https://doi.org/10.1016/j.ypmed.2024.107898>
- Lees, B., Meredith, L. R., Kirkland, A. E., Bryant, B. E., & Squeglia, L. M. (2020). Effect of alcohol use on the adolescent brain and behavior. *Pharmacology Biochemistry and Behavior*, 192, 172906. doi: 10.1016/j.pbb.2020.172906
- Masur, P. K., & Scharkow, M. (2023). specr: Conducting and Visualizing Specification Curve Analyses. In.
- Matthay, E. C., Hagan, E., Gottlieb, L. M., Tan, M. L., Vlahov, D., Adler, N., & Glymour, M. M. (2021). Powering population health research: Considerations for plausible and actionable effect sizes. *SSM Popul Health*, 14, 100789. doi: 10.1016/j.ssmph.2021.100789
- McCabe, C. J., Brumback, T., Brown, S. A., & Meruelo, A. D. (2023). Assessing cross-lagged associations between depression, anxiety, and binge drinking in the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) study. *Drug and Alcohol Dependence*, 243, 109761. doi: 10.1016/j.drugalcdep.2022.109761
- Ning, K., Gondek, D., Patalay, P., & Ploubidis, G. B. (2020). The association between early life mental health and alcohol use behaviours in adulthood: A systematic review. *PLoS One*, 15(2), e0228667. doi: 10.1371/journal.pone.0228667
- Patel, C. J., Burford, B., & Ioannidis, J. P. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology*, 68(9), 1046–1058. doi: 10.1016/j.jclinepi.2015.05.029
- Puddephatt, J. A., Irizar, P., Jones, A., Gage, S. H., & Goodwin, L. (2022). Associations of common mental disorder with alcohol use in the adult general population: A systematic review and meta-analysis. *Addiction*, 117(6), 1543–1572. doi: 10.1111/add.15735
- Sarma, A. (2023). Package ‘Multiverse’. In.
- Schuster, N. A., Twisk, J. W., Ter Riet, G., Heymans, M. W., & Rijnhart, J. J. (2021). Noncollapsibility and its role in quantifying confounding bias in logistic regression. *BMC Medical Research Methodology*, 21, 1–9. <https://doi.org/10.1186/s12874-021-01316-8>
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11), 1208–1214. <https://doi.org/10.1038/s41562-020-0912-z>
- Smout, S., Champion, E. E., O'Dean, S., Teesson, M., Gardner, L. A., & Newton, N. C. (2024). Anxiety, depression and distress outcomes from the Health4Life intervention for adolescent mental health: A cluster randomised controlled trial. *Nature Mental Health*, 2, 818–827. <https://doi.org/10.1038/s44220-024-00246-w>
- Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., ... Fusar-Poli, P. (2021). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, 27(1), 281–295. <https://doi.org/10.1038/s41380-021-01161-7>
- Spear, L. P. (2018). Effects of adolescent alcohol consumption on the brain and behaviour. *Nature Reviews Neuroscience*, 19(4), 197–214. doi: 10.1038/nrn.2018.10

- Steegeen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712. doi: 10.1177/1745691616658637
- Sullivan, G. M., & Feinn, R. (2012). Using effect size-or why the *p* value is not enough. *Journal of Graduate Medical Education*, 4(3), 279–282. <https://doi.org/10.4300/JGME-D-12-00156.1>
- Suntharalingam, S., Johnson, D., Suresh, S., Thierault, F. L., De Sante, S., Perinpanayagam, P., ... Pajer, K. (2022). Rates of dual diagnosis in child and adolescent psychiatric inpatients: A scoping review. *Journal of Addiction Medicine*, 16(1), 101–109. doi: 10.1097/ADM.0000000000000818
- Teesson, M., Champion, K. E., Newton, N. C., Kay-Lambkin, F., Chapman, C., Thornton, L., ... Gardner, L. A. (2020). Study protocol of the Health4Life initiative: A cluster randomised controlled trial of an eHealth school-based program targeting multiple lifestyle risk behaviours among young Australians. *BMJ Open*, 10(7), e035662. doi: 10.1136/bmjopen-2019-035662
- United Nations Department of Economic and Social Affairs. (2015). Goal 3: Ensure healthy lives and promote well-being for all at all ages, Sustainable Development Goals. <https://sdgs.un.org/goals/goal3>
- Vanyukov, M. M., Tarter, R. E., Kirillova, G. P., Kirisci, L., Reynolds, M. D., Kreek, M. J., ... Bierut, L. (2012). Common liability to addiction and “gateway hypothesis”: Theoretical, empirical and evolutionary perspective. *Drug and Alcohol Dependence*, 123, S3–S17. doi: 10.1016/j.drugalcdep.2011.12.018
- Visontay, R., Mewton, L., Sunderland, M., Bell, S., Britton, A., Osman, B., ... Slade, T. (2023). A comprehensive evaluation of the longitudinal association between alcohol consumption and a measure of inflammation: Multiverse and vibration of effects analyses. *Drug and Alcohol Dependence*, 247, 109886. doi: 10.1016/j.drugalcdep.2023.109886
- Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *The Lancet*, 370(9596), 1453–1457. doi: 10.1016/S0140-6736(07)61602-X
- Watson, D., Levin-Aspenson, H. F., Waszczuk, M. A., Conway, C. C., Dalglish, T., Dretsch, M. N., ... Hobbs, K. A. (2022). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): III. Emotional dysfunction superspectrum. *World Psychiatry*, 21(1), 26–54. doi: 10.1002/wps.20943
- Williams, R., & Jorgensen, A. (2023). Comparing logit & probit coefficients between nested models. *Social Science Research*, 109, 102802. doi: 10.1016/j.ssresearch.2022.102802
- Wu, Y., Levis, B., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., ... Gilbody, S. (2020). Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: A systematic review and individual participant data meta-analysis. *Psychological Medicine*, 50(8), 1368–1380. doi: 10.1017/S0033291719001314