Examining vaping's possible unintended consequences on cannabis initiation and the initiation of other substances

Alexander S. Perlmutter

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023

© 2023

Alexander S. Perlmutter

All Rights Reserved

Abstract

Examining vaping's possible unintended consequences on cannabis initiation and the initiation of other substances

Alexander S. Perlmutter

Electronic nicotine delivery systems emerged during the 2010s as a novel way to consume (i.e., vape) nicotine. Public health authorities became concerned that vaping could cause nicotine-naïve youth to begin using tobacco products and that a new generation of youth could become tobacco-dependent. Though millions of youth have vaped, authorities' fears about a new generation of youth tobacco dependence has not materialized. A more recent concern is nicotine vaping's potential effects on cannabis use and the use of other substances. An increase in cannabis use among some adolescent groups and young adults could be because of nicotine vaping's rise. Additionally, cannabis can be vaped, so transitioning from nicotine vaping to cannabis vaping may be easier than transitioning from nicotine vaping to other forms of cannabis use. Furthermore, nicotine product use was historically associated with later use of cannabis and other substances; this trend may be renewed with the advent of nicotine vaping. To date, most studies on the associations between nicotine vaping and cannabis/other substance use are cross-sectional, so more longitudinal evidence is needed. If evidence suggests that nicotine vaping does affect the use of cannabis and other substances, specifying a mechanism would help with developing potential interventions and with testing the validity of total effects. The overarching goal of this dissertation is to advance evidence of nicotine vaping's potential harmful effects on youth and young adults, which could be used to support

interventions aimed at reducing the burden of nicotine vaping's outcomes. *First*, I conducted a systematic review in which I examined the extent to which confounding, measurement errors, and loss to follow-up could alternatively explain reported longitudinal effects of nicotine vaping on cannabis use or other substance use. I also identified studies that tested effect modification and mediation. This systematic review revealed that nicotine vaping likely increases the risk of subsequent cannabis use and other substance use for up to 24 months. It also revealed that some studies evaluated effect measure modification, while no study assessed mechanisms. These observations suggest that future studies should assess long-term effects on initiation and evaluate potential mechanisms. *Second*, I evaluated whether nicotine vaping affected the initiation of cannabis and other substances over a six-year period among adolescents as they age into adulthood. Results suggested that nicotine vaping had harmful effects on both outcomes over the six-year period. I also found evidence that nicotine vaping's harmful effects in later years appeared stronger than in earlier years; the absence of age effects suggest the absence of cohort effects. Furthermore, I found that effects appeared stronger among individuals who had a history of non-vaping tobacco product use than among individuals without a history of non-vaping tobacco product use, suggesting that tobacco use is key to nicotine vaping's harms *Finally*, I evaluated possible mechanisms of the effects based on a theory that I developed from prior empirical literature and behavioral theory. I posited that nicotine vaping caused deviant peer affiliation, which caused conduct problems and subsequently, the outcomes. I found no evidence that three conduct problems (considered together) were mechanisms of the effects. Future studies of mechanisms can reveal potential intervention targets, lead to studies of other potential mechanisms, and help test the validity of total effects. This dissertation achieved its goal of advancing evidence that nicotine vaping may harm youth and young adults. Public health bodies tasked with addressing potential public health concerns about nicotine vaping products should consider evidence from this dissertation.

Table of Contents

Acknowledgments xi

Chapter 1: Introduction 1

1.1 Background 1

1.2 Dissertation overview 2

Chapter 2: Does current evidence support harmful effects of nicotine vaping on cannabis and other substance use risks? A systematic review 5

2.1 Abstract 5

2.2 Introduction 7

2.3 Methods 8

2.3.1 Data sources and searches 9

2.3.2 Study selection 9

2.3.3 Data extraction 10

2.3.4 Data synthesis and analysis 10

2.4 Results 14

2.4.1 General Characteristics of included studies 14

2.4.2 Association 16

2.4.3 Sources of bias 17

2.4.4 Effect modification and mediation 31

2.5 Discussion 35

2.5.1 Conclusions 39

Chapter 3: Does nicotine vaping affect cannabis and other substance initiation from adolescence into young adulthood? Results from Wave 1 (2013-14) to Wave 5 (2018-19) of the Population Assessment of Tobacco and Health Study 40

3.1 Abstract 40

3.2 Introduction 42

3.3 Methods 44

3.3.1 Population Assessment of Tobacco and Health Study (PATH) 44

3.3.2 Sampling design 45

3.3.3 Data collection 45

3.3.4 Study population 46

3.3.5 Estimand 46

3.3.6 Identification 50

3.3.7 Measures 50

3.3.8 Statistical analysis 52

3.4 Results 54

3.4.1 Distributions of exposure and confounders at baseline 54

3.4.2 Distributions of baseline confounders stratified by baseline nicotine vaping status 57

3.4.3 Nicotine vaping prevalence over time 65

3.4.4 Risks of cannabis initiation by Waves 2 through 5 65

3.4.5 Risks of other substance initiation by Waves 2 through 5 67

3.4.6 Incremental propensity score effects on cannabis initiation by Waves 2 through 5 . . . 67

3.4.7 Incremental propensity score effects on other substance initiation by Waves 2 through 5 . . . 67

3.4.8 Potential effect modification by age . . . 69

3.4.9 Potential effect modification by time . . . 72

3.4.10 Potential effect modification by tobacco product co-use . . . 76

3.4.11 Sensitivity analyses . . . 76

3.5 Discussion . . . 78

Chapter 4: Do conduct problems mediate the effects of nicotine vaping on cannabis and other substance initiation? Results from Wave 1 (2013-14) to Wave 5 (2018-19) of the Population Assessment of Tobacco and Health Study . . . 82

4.1 Abstract . . . 82

4.2 Introduction . . . 84

4.3 Methods . . . 87

4.3.1 Population Assessment of Tobacco and Health (PATH) . . . 87

4.3.2 Sampling design . . . 87

4.3.3 Data collection . . . 88

4.3.4 Study populations . . . 88

4.3.5 Estimand . . . 89

4.3.6 Identification assumptions . . . 93

4.3.7 Measures . . . 94

4.3.8 Statistical analysis . . . 97

4.4 Results . . . 99

4.4.1 Distributions of baseline confounders . . . 99

4.4.2 Distributions of nicotine vaping and posited mediators stratified by outcome status . . . 100

4.4.3 Distribution of nicotine vaping stratified by posited mediators . . . 102

4.4.4 Stochastic interventional indirect effects . . . 104

4.5 Discussion . . . 106

4.5.1 Conclusions . . . 110

Chapter 5: Conclusions . . . 111

5.1 Summary of findings . . . 111

5.2 Reflection on the dissertation's results and achievements . . . 116

5.2.1 Longitudinal effects . . . 116

5.2.2 Mechanisms . . . 118

5.2.3 Methodological advancements . . . 119

5.3 Reflection on the dissertation's challenges . . . 121

5.4 Future directions . . . 125

5.5 Final thoughts . . . 127

References . . . 131

Appendix A: Aim 1 . . . 149

A.1 Search equation . . . 149

A.1.1 Google Scholar . . . 149

A.1.2 Pubmed (MEDLINE) and PsychINFO . . . 149

A.2 Covidence extraction form . . . 150

Appendix B: Aims 2 and 3 180

B.1 Approach to confounding 181

B.2 Revised incremental propensity score interventions code
(based on Kennedy's **npcausal::ipsi**) 188

B.3 Stochastic direct and indirect effects code
(based on Hejazi's **pub_medshift_interv_biostats::example**) 196

List of Figures

2.1 Diagram depicting correlations between evidence-based confounders and their likely effects on nicotine vaping and cannabis and other substance use (other confounders are ones controlled in studies and seen from Table A.7) 12

2.2 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Flowchart 16

2.3 Proportion of 44 "main effect" estimates and 32 lower confidence interval bounds that could truly be null if controlling for unmeasured confounding of varying strengths, indexed by deciles of E-values (first and ninth are furthest left and right points, respectively, and vertical line is median) 23

3.1 Causal diagram depicting assumed causal structure, where L_1 is a set of baseline confounders, A_t is time-varying nicotine vaping at Waves 1 through 4, nodes L_2, L_3, and L_4 comprise sets of time-varying confounders W_t and observed outcomes Y_t at Waves 2 through 4, and Y_5 is the observed outcome at Wave 5, and blue and black lines depict exposure and other effects, respectively 46

3.2 Nicotine vaping prevalence across PATH Waves 1 through 4 (N=9,571 baseline participants 12 to 16 years old) 65

3.3 Past-year observed cannabis initiation risks and past-year observed other substance initiation risks and 95% confidence intervals (right-most dots) by each follow-up wave and those risks had nicotine vaping odds been 10% to 90% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.1$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old 66

3.4 Risk differences and 95% confidence intervals comparing past-year observed cannabis initiation risks and past-year observed other substance initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 90% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.1$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old. Example: where $\delta = 0.1$ on plot, risk difference shown is $\mathbb{E}\left(Y_t - Y_t^{\delta=0.1}\right)$ 68

3.5 Age differences in risk differences comparing observed past-year cannabis initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 80% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.2$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old 70

3.6 Age differences in risk differences comparing observed past-year other substance initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 80% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.2$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old 71

3.7 Wave differences in risk differences comparing observed past-year cannabis initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 80% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.2$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old 73

3.8 Wave differences in risk differences comparing observed past-year other substance initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 80% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.2$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old 75

3.9 Observed Wave 5 risks and 95% confidence intervals of past-year observed cannabis initiation and past-year observed other substance initiation and those risks had nicotine vaping odds been 80% lower in Wave 4 and differences in those risks, all stratified by history of non-vaping tobacco product use at Waves 1 through 4, and controlling for baseline and time-varying confounding, N=9,571 baseline participants 12 to 16 years old 77

4.1 Causal diagram depicting assumed causal structure, where L is a set of baseline confounders, A is nicotine vaping at Waves 1, 2, or 3, Z is a non-vaping tobacco use at Wave 3, M is a set of mediators at Wave 4 (lying or conning to get something, bullying perpetration, starting fights), and Y is observed cannabis initiation or other substance initiation at Wave 5 89

4.2 Stochastic interventional indirect effects: effects of shifting nicotine vaping odds on cannabis initiation (left) and the initiation of other substances (right) if lower nicotine vaping odds changed only the conditional distribution of three conduct problems (lying or conning to get something, bullying perpetration, and starting fights at least twice in the past 12 months). 105

B.1 Plot showing little overlap of conditional probabilities of nicotine vaping (i.e., propensity scores) for observed nicotine vaping and non-vaping groups (i.e., positivity violation) at Wave 1 of PATH, 9,572 baseline participants aged 12 to 16 years old 181

B.2 E-values for risk differences from Figure 3.4 by each follow-up Wave 193

B.3 Conceptual model showing nicotine vaping's posited indirect effects on cannabis initiation and other substance initiation through its effects on conduct problems within the context of a deviant peer environment 194

B.4 Plot showing little overlap of conditional probabilities of nicotine vaping (i.e., propensity scores) for observed nicotine vaping and non-vaping groups (i.e., positivity violation) at Waves 1, 2, or 3 of PATH, 2,949 and 2,995 baseline participants aged 12 to 16 years old in cannabis initiation cohort (top) and other substance initiation cohort (bottom), respectively 195

B.5 Plot showing overlap of propensity scores for any conduct problem, lying or conning to get something, bullying perpetration, and starting fights by corresponding observed groupings (i.e., present at least twice in the past year and otherwise) in Wave 4 of PATH, 2,949 baseline participants aged 12 to 16 years old in cannabis outcome cohort 211

List of Tables

2.1 Descriptive characteristics of 16 studies included for quantitative and qualitative analysis 14

2.2 Abbreviated assessment of covariate or confounding control in 17 included studies 18

2.3 Nine unique sets of confounder control from four groupings of potential unmeasured confounders, aggregated across 17 included studies 22

2.4 Assessment of approaches to address selection bias from potential differential loss to follow-up in 17 included studies 26

2.5 Assessment of effect modification in 17 included studies 32

3.1 Distribution of baseline confounders (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) 54

3.2 Distribution of baseline confounders by baseline nicotine vaping status (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) 59

4.1 Table assessing Wave 5 outcomes' associations with nicotine vaping at Waves 1, 2, or 3 and conduct problems at Wave 4 (i.e., posited mediators) in cannabis initiation cohort (N = 2,949) and other substance initiation cohort (N = 2,995), Population Assessment of Tobacco and Health 101

4.2 Table assessing associations between three conduct problems at Wave 4 (i.e., posited mediators) and nicotine vaping at Waves 1, 2, or 3 in cannabis initiation cohort (N = 2,949) and other substance initiation cohort (N = 2,995), Population Assessment of Tobacco and Health 103

A.1 Table used for extracting exposure information 151

A.2 Table used for extracting outcome information 151

A.3 Table used for extracting covariate information 152

A.4 Table used for extracting bias information 152

A.5 First author, title, and year of publication of 17 included studies 153

A.6 Assessment of associations in 17 studies included 156

A.7 Assessment of covariate or confounding control in 17 included studies 161

B.1 Dates and numbers of interviews from data collection waves from PATH restricted data files user guide 180

B.2 List of confounders selected for incremental propensity score intervention and stochastic interventional in(direct) effects analyses 183

B.3 Aim 2 missing exposure data 192

B.4 Aim 2 missing outcome data (cannabis initiation) 192

B.5 Aim 2 Missing outcome data (prohibited substance initiation) 192

B.6 Aim 3 missing data among 2,949 participants who never reported conduct problems before Wave 4 and who never reported cannabis use before Wave 5 196

B.7 Aim 3 missing data among 2,995 baseline participants aged 12 to 16 years old who never reported conduct problems before Wave 4 and who never reported prohibited substance use before Wave 5 196

B.8 Distribution of baseline confounders among baseline 12- to 16-year-old youth in cannabis initiation (N = 2,949) and other substance initiation (N=2,995) outcome cohorts, Population Assessment of Tobacco and Health 212

B.9 Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health . . 219

Acknowledgements

I would like to sincerely thank my dissertation committee for all of their time and attention to this dissertation. In particular, I would like to thank Dr. Kara Rudolph, who pushed me to learn many new methods and challenge my previously held beliefs about them. Thank you for also helping me become a better writer and to think more independently than I ever have. Thank you to Dr. Katherine Keyes, whose incisive critiques often made me reevaluate the framing of this dissertation, but who always offered words of encouragement that it was moving in the right direction. Thank you to Dr. Sharon Schwartz for the five-minute to hour-plus-long deskside conversations that I missed so much during the pandemic and which made the doctoral experience at Columbia so meaningful to me. Your eagle eye for identifying countless deviations from arguments I tried making in this dissertation (and otherwise) helped me pay greater attention to my thought process when writing and to self-critique it often; however, I am very much still a work in progress.

A special thanks to Dr. Silvia Martins for her unwavering support during my career, first by helping me launch it and then supporting the several directions I sought to take it. Thanks to Dr. Moïse Desvarieux for helping me learn about the challenges of primary data collection, advocating for my admission to the doctoral program, and his enduring friendship. Liliane Zaretsky and Elizabeth Ferrari, the department would not run without you. Thank you for so effectively advocating for the doctoral students.

I would like to thank my fellow doctoral students, particularly Dr. Luis Segura, Dr. Paul Reeping, Dr. Diana Garofalo, Dr. Hillary Samples, and the soon to be Dr. Natalie Levy for their friendships. I would not have finished this degree if it were not for you. Additionally, thank you to Anna Krasnova for helping with my systematic review.

Finally, thank you to my supportive and loving family who put me in a position to succeed despite some early challenges, and to my fiancée, Camille, for her unconditional love and support these past four years.

Chapter 1: Introduction

1.1 Background

Nicotine vaping, or the consumption of aerosolized nicotine via electronic nicotine delivery systems has become a common way for youth to consume nicotine. Nationally, nicotine vaping in the past 30 days increased between 2017 and 2019 from 4% to 9% among 8th graders, from 8% to 20% among 10th graders, and from 11% to 21% among 12th graders).[1] Among high school students, nicotine vaping prevalence in the past 30 days increased sharply from 21% in 2018 to 27.5% in 2019, then decreased to 19.6% in 2020.[2–4] Despite this decrease, nicotine vaping on 20 or more days in the past 30 days among individuals who vaped in the past 30 days continued to increase from 27.7% in 2018 to 34.2% in 2019 to 39% in 2020.[2–4] This pattern of nicotine vaping suggests that there may be many high schoolers who only experiment with nicotine vaping and some who vape frequently.

Nicotine vaping among youth is associated with numerous health hazards, including future use of cannabis and other substances, such as cocaine and heroin.[5–12] Regular use of these substances may increase the risk for mental health problems,[13] substance dependence,[14] injection drug use and associated disease,[15,16] sleep problems,[17] deterioration of interpersonal relationships,[18] poor job and school performance,[19] and unemployment.[20] Even trialing the substances may indirectly lead to these problems through increasing the probability of habitual use.[21]

There are still many unknowns concerning longitudinal associations between nicotine vaping and the use of cannabis and other substances. First, no one has adjudicated the extent to

which the longitudinal evidence is consistent with causal effects. Second, even if the evidence suggested that nicotine vaping increased the risk of cannabis use and the use of other substances, to my knowledge there are no studies of long-term effects among adolescents as they age into adulthood. Studying the long-term effects could tell us whether nicotine vaping remains a public health threat many years later. Third, to my knowledge there is no evidence concerning the mechanisms of any possible effects. Studying mechanisms could reveal potential intervention targets and other mechanisms that could be studied, and it could help test the validity of any total effects.

1.2 Dissertation overview

I achieved three aims with this dissertation. In the first aim I conducted a systematic review about the degree to which nicotine vaping's longitudinal associations with use of cannabis and other substances published between 2013 and 2021 were consistent with causal effects. Specifically, I examined the extent to which confounding, measurement error, and selection bias could alternatively explain results. I also examined whether studies evaluated effect measure modification and mediation.

The first aim revealed evidence of longitudinal associations but no evidence of effects that lasted longer than 24 months; therefore, in the second aim I examined whether nicotine vaping affected the initiation of cannabis and other substances among youth as they aged into young adulthood over six years. Evidence of these effects would suggest that nicotine vaping prevention should occur from adolescence into adulthood. I conducted a study to examine the effects using data from the Population Assessment of Tobacco and Health Study (hereon, PATH). Unfortunately, the positivity assumption for causal identification was violated; that is, a large proportion participants had extremely low probabilities of exposure, conditional on measured confounders. Therefore, I estimated incremental propensity score effects, which guarantee positivity. These effects were comparisons between observed longitudinal outcome

risks with the outcome risks had the probability of nicotine vaping been lower than observed in prior years.

The first aim also revealed that no study examined potential mechanisms of longitudinal effects; therefore, in the third aim I examined whether three conduct problems—lying or conning to get something, bullying perpetration, and starting fights—were mechanisms of nicotine vaping's effects on the initiation of cannabis and other substances. Conduct problems were not the originally posited mediators. Rather, I posited deviant peer affiliation as the mediator of true interest, but I could not find longitudinal data to estimate this possible indirect effect. I chose to estimate indirect effects through conduct problems because evidence suggests that youth and young adults may have conduct problems because of the influence of their deviant peers. In short, the conduct problems were posited as downstream mediators nicotine vaping's effect on the initiation of cannabis and other substances. Evidence of positive indirect effects through conduct problems would suggest that deviant peer affiliation or peer-influenced conduct problems are possible intervention target for blunting nicotine vaping's harmful effects. It would also be evidence supporting the validity of nicotine vaping's total effects on the outcomes.

Ideally, I would have estimated natural indirect effects to test the mediational hypothesis, but they cannot be identified in the presence of intermediate confounder(s) (i.e., common cause(s) of the mediator and outcome that are caused by the exposure). Another option was to estimate interventional indirect effects (of a static exposure), which can be identified in the presence of intermediate confounding. Unfortunately, these effects could not be identified because I encountered positivity violations as in the second aim; therefore, I estimated stochastic interventional indirect effects because they are identifiable in the presence of intermediate confounding and positivity violations. A stochastic interventional indirect effect can be expressed as the effect of a shift in the odds of nicotine vaping on cannabis initiation or the

initiation of other substances had the shift in nicotine vaping odds changed the conditional joint distribution of conduct problems. It tells us whether nicotine vaping affects the mediators together and whether the mediators together affect the outcomes. I comprehensively controlled for measured confounding of nicotine vaping and the outcomes, nicotine vaping and the three conduct problems, and the conduct problems and the mediator, and I tried to prevent selection bias by imputing missing data.

This dissertation is the first to evaluate the extent to which nicotine vaping's associations with use of cannabis and other substances use can be explained by bias. It is also the first to study whether nicotine vaping has long-term effects on the initiation of cannabis and other substances, and it is the first to test mechanisms of the potential effects.

Chapter 2: Does current evidence support harmful effects of nicotine vaping on cannabis and other substance use risks? A systematic review

2.1 Abstract

Background: The rapid uptake of nicotine vaping and simultaneous increase in cannabis use among some adolescent groups and young adults are associated. I conducted a systematic review to clarify the extent to which longitudinal associations were compatible with effects of nicotine vaping on cannabis and other substance use.

Methods: I systematically searched Pubmed (MEDLINE), PsychINFO, and Google Scholar for studies that examined longitudinal associations between nicotine vaping and substance use outcomes (i.e., use of cannabis, methamphetamine, speed, heroin, inhalants, solvents, hallucinogens, cocaine, or crack, or nonmedical use of prescription substances), published between 2013 and 2021. Using simulations, I examined confounding, measurement error, and selection bias as possible alternative explanations for associations. I then assessed evidence of effect modification as potentially providing context for possible effects and evidence of mediation as potentially supporting the validity of effects.

Results: A total of 17 studies with a median of 4,670 participants (IQR: 1,342-7,842) were identified from searches. These studies reported 46 associations between nicotine vaping and outcomes, of which six were negative and 40 were positive. Based on simulations, a set of unmeasured confounders that multiplies the risk of nicotine vaping and outcomes by 2.36 could alternatively explain 77% of these associations. The median odds ratio (OR=1.46) could be biased towards (from OR=1.53) and away from the null (from OR=1.40) due to different

measurement errors. Differential loss to follow-up likely biases effects towards the null if there is disproportionately greater loss to follow-up among people with greater substance use involvement. Possible effect modifiers are age group, sensation-seeking proclivity, cigarette smoking, and gender. Mediation hypotheses remain untested.

Conclusions: Nicotine vaping likely positively affects use of cannabis and other substances despite potential biases in observed associations, but effects are likely smaller than associations (ORs around 2) reported in even the best conducted studies. Evidence of effect modification by age group, smoking status, and sensation-seeking proclivity adds context to possible effects. Future investigations of mechanisms could help test the validity of the effects.

2.2 Introduction

Nicotine vaping has rapidly increased in the past 10 years. From 2011 to 2020 past-month e-cigarette use among nationally representative samples of high school students increased from 1.5% to 19.6%, peaking in 2019 (27.5%).[3,4,22] By 2020, 10.5%, 19.3%, and 24.7% percent of eighth, 10th, and 12th graders reported past-month nicotine vaping.[23] These trends led the US Surgeon General to label youth nicotine vaping an epidemic.[24] The Surgeon General's report cited numerous potential harms, such as lung damage from inhalation of aerosolized constituents (e.g., heavy metals, volatile organic compounds, ultrafine particles), impeded brain development, future tobacco smoking, and addiction to other drugs.[24]

Increasing nicotine vaping is concerning because it appears to be longitudinally associated with initiation or use of cannabis and other substances with negative health effects, methamphetamine, speed, heroin, inhalants, solvents, hallucinogens, cocaine, or crack, or nonmedical use of prescription substances.[8,10,25–28] Sustained use of cannabis and other substances can contribute to ongoing problems like increasing substance use disorders,[29] as well as to psychosis, behavioral and cognition problems,[13] infectious disease,[15,16] sleep problems,[17] worsening interpersonal relationships,[18] poor job and school performance,[19] and unemployment.[20] The earliest research using data from 2013 to 2015 documented nicotine vaping's positive longitudinal associations with cannabis initiation and use among adolescents and young adults.[26–28,30] More recent research using data from 2013 to 2016 documented mostly positive but some inverse longitudinal associations with initiation or use of cannabis (including cannabis vaping) and other substances.[10,31]

Whether nicotine vaping affects cannabis and other substance use is unclear due to potential bias. A systematic review assessed that nicotine vaping's effects on cannabis use may have been overestimated because studies lacked control for important confounders.[25] Another

review specified living with people who smoke, peer smoking, school performance or commitment, likelihood of trying cigarettes in the future, alcohol and other drug use, and deviant behaviors could be potential unmeasured confounders that positively bias nicotine vaping's positive effects on smoking.[32] These potential unmeasured confounders could also positively bias nicotine vaping's effects on cannabis and other substance use. Measurement error is another form of bias that could alternatively explain effects. Imperfect self-reported substance use, such as misreporting the timing of binary nicotine vaping or cannabis use, is expected to bias effects towards the null whereas deception likely leads to bias away from the null.[33,34] Differential loss to follow-up is third form of bias that could alternatively explain effects; it is expected to bias effects towards the null, particularly when loss is disproportionately greater among participants with greater substance use involvement (i.e., more substances and more frequent use).[35–38] How much and in which direction the admixture of confounding, measurement errors, and differential loss may have biased observed effect estimates is unclear. Aside from evaluating how much effect estimates may be biased, evaluating effect modification (for whom or when effects occur) may add context to potential effects, while evaluating mechanisms (how effects occur) may test the validity of the effects;[39] these issues have also not been reviewed.

I conducted a systematic review in which I examined the potential impact of unmeasured confounding, measurement errors, and differential loss to follow-up on nicotine vaping's reported longitudinal effects on cannabis or other substance use. I also identified studies that tested effect modification and mediation.

2.3 Methods

I followed standard procedures for systematic reviews according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).[40]

2.3.1 Data sources and searches

I systematically searched Pubmed (MEDLINE), PsychINFO, and Google Scholar for eligible studies published between January 1, 2013, and December 31, 2021. The Pubmed and PsychINFO searches had a combination of MeSH terms of many substances and free text words, while the Google Scholar search only included the latter (Appendix A). I also searched reference lists of included studies. I excluded alcohol use from my searches (i.e., it was not an outcome of interest) because I intended to identify studies that estimated whether nicotine vaping affected initiation of outcomes. Alcohol initiation usually precedes nicotine vaping initiation, or both are initiated at the same age or in the same grade,[41–43] which makes estimating nicotine vaping's effects on alcohol initiation challenging. This challenge is evident in longitudinal studies that reported positive cross-sectional or longitudinal associations with non-incident alcohol use and no effect on incident alcohol use.[6,31,44]

2.3.2 Study selection

I included published reports of studies that estimated one or more longitudinal associations between nicotine vaping and subsequent use of cannabis or other substances among study participants aged 25 years old or younger at baseline. Other substance use was defined as prescription drugs used nonmedically, such as use of Ritalin, Adderall, painkillers, sedatives, and tranquilizers, or use of methamphetamine, speed, heroin, inhalants, solvents, hallucinogens, cocaine, or crack. Self-reported nicotine vaping was defined as any occurrence within a given period (past-month, past-year, ever) or a frequency within a given period (number of times in past year). Similarly, self-reported cannabis and other substance use was defined as any consumption of those products within a given period (past-month, past-year, ever) or a frequency within a given period (number of times in past year). Studies were eligible if they compared summary measures (odds, risk, probability, etc.) of cannabis or other substance use for a particular history of nicotine vaping with an alternative history of nicotine vaping. I included only longitudinal studies because cross-sectional studies lack temporality. Unlike

longitudinal studies, cross-sectional studies are not usually designed to address bias due to confounding or loss to follow-up, or investigate mediation, which were three of the five concepts I aimed to examine (the others being effect modification and measurement error).

I used Covidence, a program that facilitates management of and collaboration on literature reviews, in the screening and eligibility confirmation phases. Two investigators (AP and AK) screened titles and abstracts for irrelevant publications. AP and AK confirmed the eligibility of all screened-in studies based on the articles' full-text and noted the reasons for not meeting eligibility.

2.3.3 Data extraction

I used a standardized extraction form created on Covidence to extract studies' general characteristics (e.g., authors, title, number of participants, exposures, outcomes, covariates, and other elements seen in the full extraction form in Table A.2). The following were also extracted if available:

1. measures of association (e.g., odds ratio (OR), risk ratio (RR) or difference, etc.), point estimates and confidence intervals (CIs), and if available, prevalence estimates of the outcome;
2. confounders controlled;
3. approaches towards assessing the impact of differential loss to follow-up or avoiding it;
4. information on effect modification by third variables; and
5. information on mediation by third variables.

2.3.4 Data synthesis and analysis

I exported the extracted data into a Microsoft Excel spreadsheet for harmonization and to conduct descriptive and qualitative analyses. Then, I summarized the characteristics of

included studies with proportions for categorical variables and medians and interquartile ranges (IQR) for continuous variables.

Next, I evaluated the correspondence between ORs and RRs to see if I could use ORs as approximations of RRs because ORs can often exaggerate RRs when RRs and/or baseline risks are very high. Assuming no bias, I considered ORs to exaggerate RRs when RRs calculated from the ORs were greater than 2 or less than 0.5 and when baseline risks were greater than 10%.[45] I also considered ORs to exaggerate RRs when RRs calculated from the ORs were greater than 1.5 or less than 0.6 and when baseline risks were greater than 20%.[45] To calculate RRs from ORs when baseline risks (p) were available, I used the formula $RR = OR/(1 - p + (p \times OR))$. When RRs were reported, I did not calculate or form any conclusions about ORs overestimating RRs.

I then analyzed potential unmeasured confounding. First, I evaluated whether studies controlled for known confounders based on several theories about substance use etiology and models of delinquency, such as Problem Behavior Theory, Interactive Theory, and socio-developmental and socio-ecological models (Figure 2.1).[46–50] The theories, models, and evidence consistently specify substance use history, family and friends' influences on substance use, psychopathology, and school or work factors (e.g., academic performance, employment status) as the strongest causes of substance use.[32,46–61] In other words, these causes could alternatively explain associations between substance use behaviors. I evaluated the strengths of these factors' effects on nicotine, cannabis, and other substance use in prior studies and ordered them as more important if they had stronger effects.[32,51–62] Then, I assessed whether there was a relationship between the importance of confounder groupings and the number of studies controlling for variables from more important confounder groupings. Next, I identified sets of studies that controlled for variables from each of the four groupings of potential unmeasured confounders to see if there was a systematic relationship between controlling for variables from more confounder groupings or more important confounder groupings. I refer to the factors as groupings of

potential unmeasured confounders because each grouping could contain specific variables. For each set of studies, I calculated the median associations and median lower 95%CIs (when available). I expected to see smaller median associations when studies controlled for variables from more confounder groupings or more important groupings. Finally, I calculated E-values for all effect estimates, and when available, their lower 95%CI bound after converting inverse estimates to be positive (e.g., $\frac{1}{0.5} = 2$). An E-value is the effect a set of unmeasured confounders must have on both a binary exposure and a binary outcome to explain the entire association between the exposure and the outcome.[63,64] I assessed the extent to which unmeasured confounding of varying strengths, indexed by the first to ninth deciles of all E-values, could have alternatively explained entire associations. I assumed there were multiple unmeasured confounders in all studies that lacked control for variables from at least one grouping. To interpret E-values when there are multiple unmeasured confounders, one must make one of many possible restrictive monotonicity assumptions, all of which require binary exposures and outcomes.[64] I assumed *distributional monotonicity*, which requires that nicotine vaping does not decrease the risks of cannabis or other substance use within all confounder strata (i.e., effects are either positive or null).

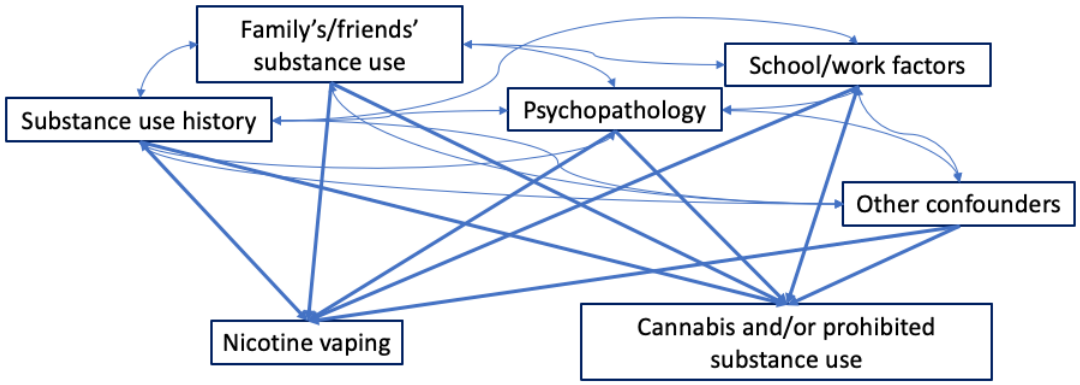

Figure 2.1: Diagram depicting correlations between evidence-based confounders and their likely effects on nicotine vaping and cannabis and other substance use (other confounders are ones controlled in studies and seen from Table A.7)

Next, I examined the most likely direction that measurement errors could have biased possible

effects. I assumed a hypothetical study with measurement error had OR=1.46 (the median across studies) and an observed baseline risk of 10%. I then simulated scenarios assuming different kinds of measurement error, namely independent and dependent nondifferential misclassification, and varying amounts of each kind of error. For the former, I simulated scenarios with varying sensitivity and specificity estimates from prior research.[34,65,66] Sensitivities ranged from 0.55 to 1, and specificities ranged from 0.90 to 0.99. For dependent nondifferential misclassification, I simulated scenarios with varying proportions of correctly classified participants positively reporting both nicotine vaping and an outcome substance or positively reporting one and negatively reporting the other. These proportions ranged from 0.55 to 1. I assumed perfect classification of participants negatively reporting both behaviors. After simulating the scenarios most compatible with prior research,[34,65,66] I gauged how much and in which direction the combination of measurement error could bias effect estimates. These Shiny apps show all simulations results.

To analyze potential selection bias due to differential loss to follow-up, I grouped studies by whether they analyzed and/or took measures to prevent potential bias due to differential loss to follow-up or did neither of these. I then calculated the median association for each group of studies and compared them. I expected the median association to be greater among studies that analyzed and/or took measures to prevent potential bias due to differential loss compared with those that did neither, which could indicate disproportionately greater loss to follow-up among people with greater substance use involvement.[36,37,67,68]

Then, I examined whether authors investigated effect modification and which variables were tested as effect modifiers. As with the "main effect" estimates, I examined baseline risks (when available) to see if ORs could have exaggerated RRs. Finally, I assessed whether authors investigated mediation, and if so, which variables were tested as mediators.

2.4 Results

2.4.1 General Characteristics of included studies

I identified 17 studies that met inclusion criteria (Figure 2.2 and Table A.5).[5–12,26–28,69–74] The median number of participants was 4,670 (IQR 1,342-7,842). Most studies were conducted in the United States (n=14, 82%), while one each occurred in Canada,[9] Germany,[11] and Mexico.[5] The earliest and latest data collection was in 2013 and 2019, respectively. Studies included school-based adolescents or children (n=6, 35%),[5,8,9,11,27,74] college students at one or more universities (n=4, 24%),[7,69,71,73] nationally representative samples of adolescents (n=3, 18%)[10,12,26] and adults (n=1, 6%),[70] current young adults from a high school sample (n=1, 6%),[28] community-based adolescents (n=1, 6%),[6] and a nationally representative sample of 12th grade students (n=1, 6%).[8] There were 46 "main effect" associations across studies (excluding interaction or mediation by third variables), 32 of which had available CIs. Most outcomes were any cannabis use or initiation, while five outcomes were specifically cannabis vaping or initiation, and six outcomes were varying operationalizations of nonmedical prescription drug or polysubstance use (Table 2.1). Many studies (41%) had prevalent outcomes. Half of the exposures were operationalized as lifetime (ever) nicotine vaping, and a plurality (33%) of outcome substances were operationalized as use in the past year (Table 2.1). Follow-up durations were mostly 2 to < 3 years after baseline (53%) (Table 2.1).

Table 2.1: Descriptive characteristics of 16 studies included for quantitative and qualitative analysis

	n	%
Number of participants (median and interquartile range)	4,670	1,342-7,842
Data collection ended	17	
2017 and later	8	47%
2016 and earlier	8	47%
Unknown	1	6%
Publication year	16	
2019 and later	11	65%
2018 and earlier	6	35%

Continued on next page

Table 2.1: Descriptive characteristics of 16 studies included for quantitative and qualitative analysis (cont.)

	n	%
Study Population	17	
College students at a university or universities	4	24%
School-based studies of adolescents (or younger)	6	35%
Nationally representative population-based studies among adolescents	3	18%
Nationally representative population-based studies among adults	1	6%
High school-based study of current adults	1	6%
Community-based study of young adolescents	1	6%
Nationally representative school-based study of 12th grade students	1	6%
Direction of results	17	
Positive	14	82%
Negative	1	6%
Mixed	2	12%
Baseline nicotine vaping/e-cigarette/ENDS measures[a]	17	
Lifetime (ever)	9	53%
Past-year (-12-month)	1	6%
Past-month (-30-day)	4	24%
Past-six-month	1	6%
Various frequencies in past year	2	12%
Outcome measures	26	
Any marijuana/cannabis use		
Initiation between baseline and follow-up (varies)	3	12%
Past-year (-12-month)	4	15%
Past-month (-30-day)	5	19%
Past-six-month	1	4%
Various frequencies in past year	2	8%
Vaping marijuana/cannabis		
Initiation between baseline and follow-up	3	12%
Past-year (-12-month)	1	4%
Past-30-day month nonprescribed prescription drug use	1	4%
Past-year (-12-month) polysubstance use	2	8%
Past-year (-12-month) nonprescribed Ritalin or Adderall use	2	8%
Past-year (-12-month) nonprescribed painkiller, sedative, or tranquilizer use	2	8%
Types of outcomes grouped by study	17	
Incident	9	53%
Prevalent	7	41%
Both	1	6%
Loss to follow-up durations (years)	17	
< 1	1	6%
1 to < 2	5	29%
2 to < 3	10	59%
3 to < 4	1	6%

[a] ENDS: Electronic nicotine delivery system(s)

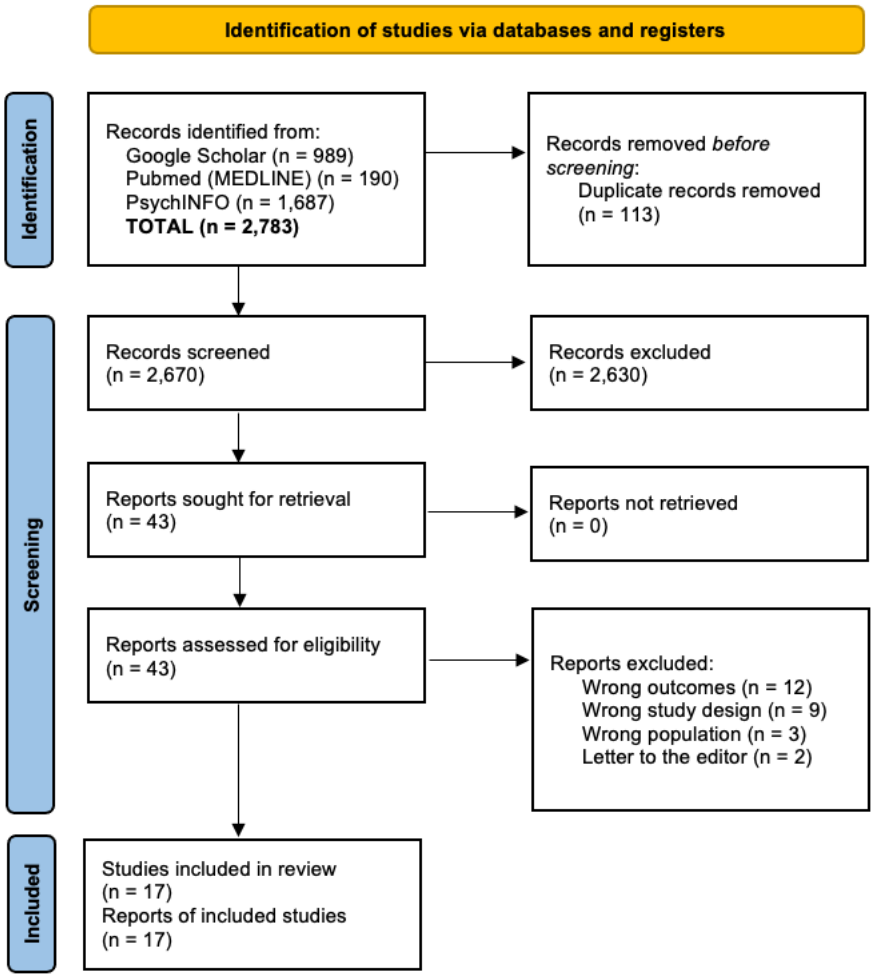

Figure 2.2: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Flowchart

2.4.2 Association

All associations in 14 of 17 studies (82%) were positive.[8–12,26–28,69–73,75] One (6%)[74] and two studies (12%)[6,7] reported inverse and mixed associations, respectively (Table 2.2). I found no evidence that ORs were substantially different than computed RRs based on the formula $RR = OR/(1 - p + (p \times OR))$.

2.4.3 Sources of bias

Confounding

Based on substance use theory and prior studies,[32,51–62] substance use history (e.g., alcohol use, prior use of an outcome substance) was the most important potential confounder grouping. Family's or friends' influences on substance use, psychopathology (i.e., behavioral or internalizing disorders or symptoms from DSM-IV or DSM-5), and school or work factors followed in importance in that order.[32,51–62] Ten studies controlled for substance use history (Table 2.2). Six and five studies controlled for family's or friends' influences and psychopathology, respectively. Four studies controlled for school and work factors. In other words, the number of studies controlling for confounders from the groupings decreases as the importance of confounder groupings decreases.

Table 2.2: Abbreviated assessment of covariate or confounding control in 17 included studies

Study	Estimate(s) (95%CI)[a]	E-value(s) (95%CI)[a]	Substance use history[b]	Family's/ friends' influences[b]	Psycho-pathology[b]	School/ employment factors[b]	Total
Cassidy	2.13 (1.20 -3.79)	2.28 (1.42)	0	0	0	0	**0**
Lozano	1.42 (0.84-2.37)	2.19 (1.00)	1	1	0	0	**2**
Dai	1.9 (1.4-2.5) 2.5 (1.2-5.3) 1.3 (1.1-1.5) 1.2 (1.0-1.5)	3.21 (2.14) 3.26 (1.69) 1.92 (1.43) 1.69 (1.00)	1	0	0	1	**2**
Audrain-McGovern	3.67 (2.51-5.36) 3.63 (2.69-4.90)	6.8 (4.46) 3.22 (2.66)	0	1	1	0	**2**
Unger	1.97 (1.01, 3.86)	3.35 (1.11)	1	0	0	0	**1**
Park	0.64[c] 0.94[c] 1.15[c] 1.15[c] 1.23[c] 1.05[c] 1.43[c] 1.37[c] 0.97[c]	2.45 1.32 1.15 1.14 1.46 1.18 1.67 1.62 1.14	0	0	0	0	**0**
Wong	2.08 (1.11-3.90) 1.37 (1.01-1.87) 1.19 (0.95-1.42)	3.58 (1.46) 2.08 (1.11) 1.67 (1.00)	1	1	0	1	**3**

Continued on next page

Table 2.2: Abbreviated assessment of confounding control in 17 included studies (cont.)

Study	Estimate(s) (95%CI)[a]	E-value(s) (95%CI)[a]	Substance use history[b]	Family's/ friends' influences[b]	Psycho-pathology[b]	School/ employment factors[b]	Total
Rogers	0.87 (0.79-0.95) 1.24 (1.16-1.32) 1.09 (1.05-1.14) 1.20 (1.16-1.25)	1.35 (1.19) 1.47 (1.37) 1.26 (1.18) 1.42 (1.37)	0	0	0	0	**0**
Pokhrel	4.63 (1.98-10.8) 13.4 (5.37-33.3)	8.73 (3.37) 26.29 (10.21)	1	1	0	0	**2**
Evans-Polce	2.03 (1.07-3.86) 2.57 (1.16-5.71)	2.20 (1.22) 4.58 (1.59)	1	0	0	1	**2**
Lee	2.16 (1.20-3.89)	3.74 (1.69)	1	1	1	0	**3**
Butler	3.09 (2.45-3.91)	5.63 (2.45)	0	0	0	0	**0**
Bentivegna	2.59 (1.90-3.52) 1.89 (1.09-3.28) 2.09 (1.43-3.05) 2.26 (1.56-3.27) 1.21 (0.79-1.87)	2.6 (2.1) 3.19 (1.40) 3.60 (2.21) 3.95 (2.49) 1.71 (1.00)	1	0	1	1	**3**
Ksinan	1.09[c] 1.01[c] 1.09[c]	1.40 1.11 1.40	1	0	1	0	**2**
Ortega[d]	NA NA	NA NA	0	0	0	0	**0**
Seidel	1.83 (1.48-2.25)	3.06 (2.32)	1	1	0	0	**2**

Continued on next page

Table 2.2: Abbreviated assessment of confounding control in 17 included studies (cont.)

Study	Estimate(s) (95%CI)[a]	E-value(s) (95%CI)[a]	Substance use history[b]	Family's/ friends' influences[b]	Psycho-pathology[b]	School/ employment factors[b]	Total
Silveira	1.9 (1.6-2.2)	2.10 (1.84)	1	0	1	0	**2**
	3.2 (2.3-4.4)	5.85 (4.03)					
	1.3 (1.0-1.8)	1.92 (1.00)					
	2.5 (1.7-3.5)	4.44 (2.79)					
Total			**10**	**6**	**5**	**4**	

[a] CI: Confidence interval. All estimates were odds ratios except Seidel *et al.* and Lozano *et al.*, who reported risk ratios, and Ortega *et al.*, who reported Probit regression coefficients.
[b] 1 and 0 represent that the confounding variable was adjusted and not adjusted for, respectively.
[c] Transformed from logarithmic scale.
[d] NA: Not applicable. Study estimated Probit regression coefficients, for which E-value cannot be calculated.

Forty-four "main effect" estimates could be evaluated (Ortega *et al.*'s could not because they estimated two Probit regression coefficients). A set of unmeasured confounders that multiplied the risk of nicotine vaping and the outcomes by 2.36 and 1.40 (median and first decile E-values) could have alternatively explained 34 (77%) and 20 (45%) associations, respectively (Figure 2.3). The median and first decile E-values across 32 "main effect" lower CI bounds were 1.69 and 1.28, respectively (Figure 2.3). A set of unmeasured confounders that multiplied the risk of nicotine vaping and outcomes by 1.69 and 1.28 could have rendered null 13 (75%) and eight (59%) lower CI bounds, respectively). There were nine unique sets of four groupings of potentially unmeasured confounders (Table 2.3); however, there was no systematic relationship between control for potential confounders from more groupings and a lower magnitude of associations. The median OR of studies that did not control for potentially unmeasured confounders from any grouping (OR=1.22) was unexpectedly lower than the median OR of studies that controlled for variables from only one grouping (OR=1.56) (Table 2.3). There were five sets of studies that controlled for variables from two groupings. The median ORs of these studies (max OR=3.65) were, as expected, generally higher than the median OR of studies that controlled for confounders from three groupings (OR=2.16), but they were unexpectedly higher than median ORs of studies that controlled for variables from one and no grouping (Table 2.3). The unexpected results could be due to chance or other bias towards the null (e.g., measurement error) in studies that had unsatisfactory confounding control.

Table 2.3: Nine unique sets of confounder control from four groupings of potential unmeasured confounders, aggregated across 17 included studies

Set	N	Substance use history[a]	Family's/ friends' influences[a]	Psycho-pathology[a]	School/ employment factors[a]	Median effect estimate[b] (95%CI)[c]
1	3	0	0	0	0	1.22 (1.16)
2	2	1	0	0	0	1.56 (1.03)
3	2	0	1	1	0	3.65 (2.60)
4	3	1	1	0	0	3.23 (1.73)
5	1	1	0	1	0	1.09[d]
6	1	1	0	1	0	2.20 (1.65)[d]
7	2	1	0	0	1	1.97 (1.13)
8	1	1	1	1	0	2.16 (1.20)
9	1	1	0	1	1	2.09 (1.43)
10	1	1	1	0	1	1.37 (1.01)

[a] 1 and 0 represent that the potential confounding variable was adjusted and not adjusted for, respectively. Combinations that did not exist are not shown.

[b] 42 evaluable effect estimates were odds ratios, and two were risk ratios.

[c] CI: Confidence interval.

[d] Two studies had the 1 0 1 0 pattern, one of which did not report confidence intervals or information to compute them, so I considered them separately.

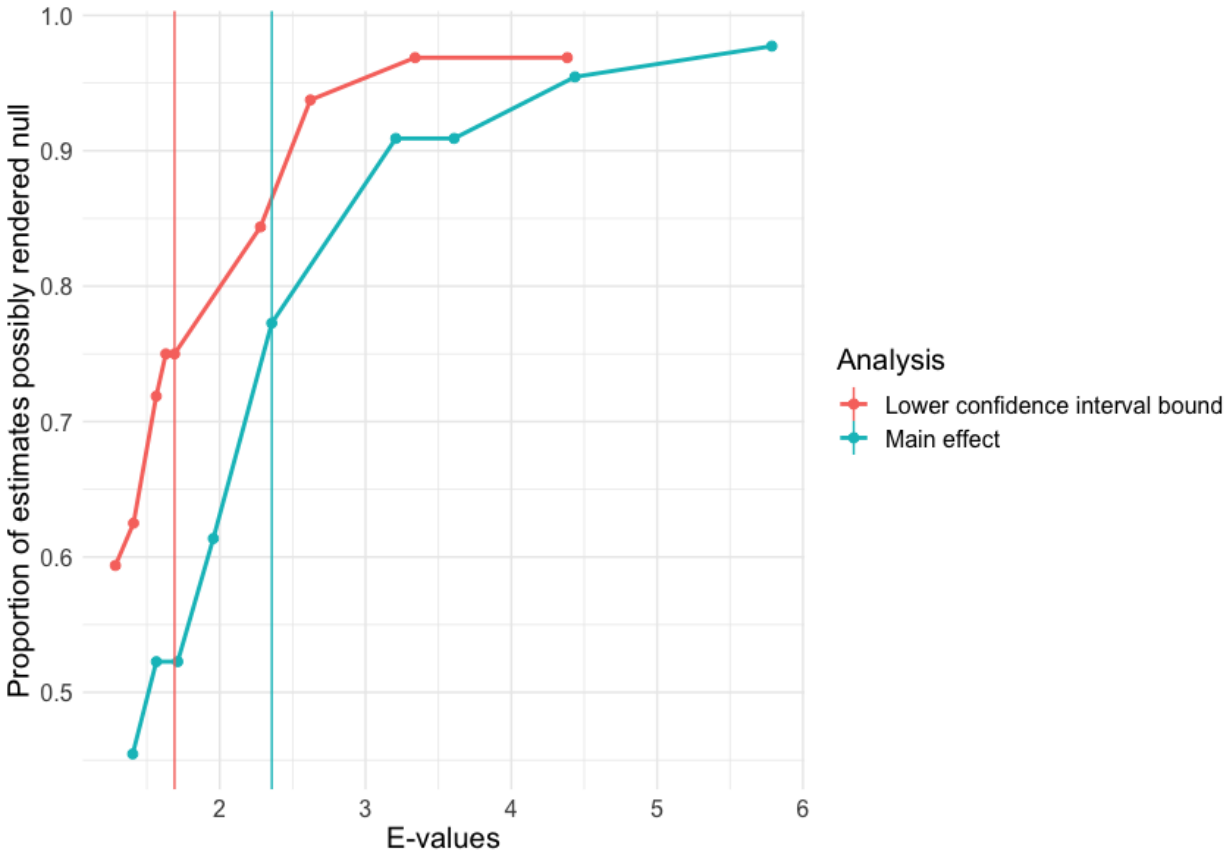

Figure 2.3: Proportion of 44 "main effect" estimates and 32 lower confidence interval bounds that could truly be null if controlling for unmeasured confounding of varying strengths, indexed by deciles of E-values (first and ninth are furthest left and right points, respectively, and vertical line is median)

Measurement error

Based on prior reports of the validity of self-reported substance use and given the prospective designs of the included studies,[33,66,67,76–78] the likeliest measurement errors are due to independent and dependent nondifferential misclassification of nicotine vaping and cannabis and other substance use. Independent nondifferential misclassification, where measurement errors of binary exposures and outcomes are uncorrelated, is generally expected to bias effects towards the null in all studies. This kind of measurement error could occur if respondents could not recall their last use of a particular substance relative to a time period or understand survey questions.[79–81] For instance, 35% of survey respondents reported ever nicotine vaping, but

51% of the same respondents responded affirmatively when questioned about use of specific nicotine vaping products (e.g., JUUL).[77] Independent nondifferential misclassification could also occur if respondents did not know which drugs corresponded to specific drug classes (e.g., that Adderall is an amphetamine) or did not know that they were exposed to one.[67,82–84] Assuming an observed OR=1.46 and 80% and 99% sensitivity and specificity, respectively, as well as no confounder misclassification, the corrected OR would have been 1.53. Please view my Shiny apps for more results.

Dependent nondifferential misclassification occurs when the sources of measurement error in the exposures and outcomes are correlated.[85,86] A possible cause of correlated measurement errors is the socially desirable response style.[86,87] Examples of possible social desirability in these studies were if youth or young adult participants misreported true nicotine vaping and/or true use of cannabis or other substances because of associated stigma and/or their illicit status for people under 21 years old.[88–90] I expected positive bias given the tendency for youth participants to underreport all kinds of substance use, and cannabis and other substance use to a greater extent than nicotine vaping.[88–90] I assumed correct classification of:

- 70% of observations reporting nicotine vaping and use of outcome substances,
- 80% of observations reporting nicotine vaping and no use of outcome substances,
- 75% of observations reporting no nicotine vaping and use of outcome substances,
- and 100% of observations reporting no nicotine vaping and no use of outcome substances.

I also assumed that 100% of people who misreported nicotine vaping and/or use of outcome substances reported using neither. The corrected OR would have been 1.40, indicating bias away from the null. Averaging this OR (1.40) and the OR corrected for independent nondifferential misclassification (1.53, sensitivity=80%, specificity=99%) produces an identical OR to the one simulated for this exercise with measurement error (1.46). These results indicate that the two

kinds of nondifferential misclassification could balance if assumptions hold. All results can be seen in these Shiny apps.

Selection bias due to differential loss to follow-up

Seven studies, including two that estimated RRs, analyzed the potential for selection bias due to differential loss to follow-up or used methods to control for its effects (Table 2.4), while the other ten studies took none of these approaches. The studies that used inverse probability of censoring weights, conducted multiple imputation by chained equations, or analyzed potential differential loss to follow-up were likely at lower risk of bias than the studies that did not address differential loss to follow-up. As expected, studies that analyzed potential bias due to differential loss to follow up or took measures to avoid its effects had a larger median effect estimate (OR=1.97) than studies that did not take these measures (OR=1.20). Removing the two studies that estimated RRs made no difference (median ORs were 2.09 and 1.20, respectively) (Table 2.4). Differential loss to follow-up would have biased effects towards the null.

Table 2.4: Assessment of approaches to address selection bias from potential differential loss to follow-up in 17 included studies

Study	Approach to address potential differential loss to follow-up	Assessment
Cassidy	None	If there is multiplicative interaction between e-cigarette use and outcome in causing loss to follow-up, OR[a] is potentially biased, otherwise no bias.
Lozano	Authors conducted a sensitivity analysis to access the potential bias from differential loss to follow-up using a propensity score analysis. The patterns of results for this sensitivity analysis were similar in direction, magnitude and statistical significance when compared to the results from my main analysis.	It is unclear what the sensitivity analysis accomplished. If exposure groups were retained differently by outcome status, or there was multiplicative interaction between e-cigarette use and outcome in causing selection, RR[b] is potentially biased, otherwise no bias.
Dai	None, but authors found no evidence that history of substance use was associated with loss to follow-up.	Possibly no selection bias.
Audrain-McGovern	Multiple imputation with Markov Chain Monte Carlo method of data augmentation	No bias if authors MAR[c] assumption was met.
Unger	None	If there is multiplicative interaction between e-cigarette use and outcome in causing loss to follow-up, OR[a] is potentially biased, otherwise no bias.

Continued on next page

Table 2.4: Assessment of approaches to address selection bias from potential differential loss to follow-up in 17 included studies (cont.)

Study	Approach to address potential differential loss to follow-up	Assessment
Park	None	Authors stated that loss to follow-up was associated with substance use suggesting differential loss to follow-up based on study variables. If there is multiplicative interaction between e-cigarette use and outcome in causing loss to follow-up, OR[a] is potentially biased, otherwise no bias.
Wong	None	If there is multiplicative interaction between e-cigarette use and outcome in causing loss to follow-up, OR[a] is potentially biased, otherwise no bias.
Rogers	None	Authors reported <1% missing data, so it is unlikely there is much bias in the OR[a]. If there is, it would be because of multiplicative interaction between e-cigarette use and outcome in causing loss to follow-up, otherwise no bias.
Pokhrel	None	If there is multiplicative interaction between e-cigarette use and outcome in causing loss to follow-up, OR[a] is potentially biased, otherwise no bias.
Evans-Polce	All analyses used sample weights, and longitudinal analyses used follow-up loss to follow-up weights (inverse probability of participation) to account for potential bias due to loss to follow-up.	Unlikely that there is selection bias.

Continued on next page

Table 2.4: Assessment of approaches to address selection bias from potential differential loss to follow-up in 17 included studies (cont.)

Study	Approach to address potential differential loss to follow-up	Assessment
Lee	Authors applied weights for survey design, which may have included loss to follow-up weights.	Unsure if weights applied considered loss to follow-up. If not, and there is multiplicative interaction between e-cigarette use and outcome in causing loss to follow-up, OR[a] is potentially biased, otherwise no bias.
Butler	None	If there is multiplicative interaction between e-cigarette use and outcome in causing loss to follow-up, OR[a] is potentially biased, otherwise no bias.
Bentivegna	PATH[d] survey weights were used to compensate for variable probabilities of selection, differential nonresponse rates, and some deficiencies in the sampling frame (e.g., under coverage of some population groups). The weights addressed sampling design and loss to follow-up due to nonresponse.	Unlikely that there is selection bias.
Ksinan	None	Unsure whether estimates were log odds. If they were, and if there were multiplicative interaction between e-cigarette use and outcome in causing selection, exponentiated log odds were potentially biased, otherwise no bias.

Continued on next page

Table 2.4: Assessment of approaches to address selection bias from potential differential loss to follow-up in 17 included studies (cont.)

Study	Approach to address potential differential loss to follow-up	Assessment
Ortega	None	If exposure groups were retained differently by outcome status, or there was multiplicative interaction between e-cigarette use and outcome in causing selection, RR^h_1s potentially biased, otherwise no bias.
Seidel	None	High degree of loss to follow-up (40%). If exposure groups were retained differently by outcome status, or there was multiplicative interaction between e-cigarette use and outcome in causing selection, RR^h_1s potentially biased, otherwise no bias.

Continued on next page

Table 2.4: Assessment of approaches to address selection bias from potential differential loss to follow-up in 17 included studies (cont.)

Study	Approach to address potential differential loss to follow-up	Assessment
Silveira	Moderate loss to follow-up from Wave 1 to Wave 2 (12% to 16%). PATH[d] survey weights were used to compensate for variable probabilities of selection and differential nonresponse rates. The weights addressed sampling design and loss to follow-up due to nonresponse.	Unlikely that there is selection bias.

[a] OR: odds ratio
[b] RR: Risk ratio
[c] MAR: Missing at random
[d] PATH: Population Assessment of Tobacco and Health

2.4.4 Effect modification and mediation

Two studies found evidence of three effect modifiers of nicotine vaping's longitudinal association with cannabis initiation (Table 2.5). First, the association appeared stronger for younger adolescents aged 12 to 14 years old (29.2% vs. 5.5%, adjusted OR=2.7) than for older adolescents aged 15 to 17 years old (25.3% vs. 10.6%, adjusted OR=1.6) (Table 2.5).[26] Second, the association appeared stronger among adolescents with low versus high sensation-seeking scores based on questions about engagement in dangerous activities; baseline risks were not reported.[11] The association was also stronger among adolescents without (23.3% vs. 9.6%) versus with a smoking history (30.8% vs. 26.4%) (Table 2.5).[11] There appeared to be no difference in associations by gender (Table 2.5).[12]

Some authors discussed different ways by which effects may operate, such as through familiarity with vaping devices, peer groups, and central nervous system changes, but none tested them.[8–10,26,27]

Table 2.5: Assessment of effect modification in 17 included studies

Study	Exposure	Effect modifier	Outcome	Assessment
Dai	Ever e-cigarette use	Age group (12-14 versus 15-17 years old)	Past-12-month marijuana use	Authors' premise for studying age group's modifying role was that early onset nicotine vaping increased the risk of cannabis initiation to a greater extent than late onset nicotine vaping. Participants were asked whether they ever vaped nicotine without specifying when. Nicotine vaping onset possibly occurred when older adolescents were younger, but the authors only considered current age (not the age at which nicotine vaping onset occurred) when comparing associations by age group.
Seidel	Lifetime e-cigarette use	High sensation-seeking	Cannabis initiation during the next 18 months	Authors hypothesized e-cigarette use conferred a higher risk of initiating cannabis for low than high sensation-seekers because sensation-seeking might raise the threshold to the point that e-cigarette use doesn't add much more risk. Authors found statistical evidence for their hypothesis, and if exchangeability held, could constitute evidence of causal multiplicative interaction.

Continued on next page

Table 2.5: Assessment of effect modification in 17 included studies (cont.)

Study	Exposure	Effect modifier	Outcome	Assessment
Seidel	Lifetime e-cigarette use	Baseline cigarette use	Cannabis initiation during the next 18 months	Authors hypothesized e-cigarette use conferred a higher risk of initiating cannabis for cigarette nonsmokers than smokers because smoking might raise the threshold to the point that e-cigarette use doesn't add much more risk. Authors found statistical evidence for their hypothesis, and if exchangeability held, could constitute evidence of causal multiplicative interaction.

Continued on next page

Table 2.5: Assessment of effect modification in 17 included studies (cont.)

Study	Exposure	Effect modifier	Outcome	Assessment
Silveira	Ever e-cigarette use	Gender	Past-12-month marijuana use at Wave 2 Past-12-month non-prescribed use of Ritalin or Adderall at Wave 2 Past-12-month non-prescribed use painkillers, sedatives, or tranquilizers at Wave 2 Past-12-month use of cocaine or crack, or other stimulants (methamphetamine or speed), heroin, inhalants, solvents, or hallucinogens at Wave 2	Authors tested an association between outcomes and the multiplicative interaction between gender and e-cigarette use without hypothesizing which gender they thought was at higher risk. They found no evidence that e-cigarette use's longitudinal association with any particular outcome substance was moderated by gender. If exchangeability held, there would not be evidence of causal multiplicative interaction.

2.5 Discussion

In this systematic review of 17 studies on longitudinal associations between nicotine vaping and cannabis and other substance use, I assessed the degree to which different sources of bias could alternatively explain associations and whether effect modification and mediation were evaluated. Nicotine vaping appears to positively affect cannabis and other substance use, but effects are likely smaller than associations (ORs around 2) reported in even the best conducted studies due to positive unmeasured confounding. A mixture of positive and negative measurement error and negative selection bias were less likely to meaningfully to change conclusions than positive unmeasured confounding. Three studies assessed effect modification by age group, sensation-seeking proclivity, cigarette smoking, and gender, which provide context for nicotine vaping's possible effects on cannabis initiation. How any effects occur is still unknown since no study evaluated mechanisms.

Forty of 44 "main effect" associations from 16 methodologically heterogeneous studies were positive. Fourteen studies reported ORs (or regression coefficients to calculate them), two reported RRs, and one reported Probit regression coefficients, which I dropped from my analyses. Under the assumption of no bias, RRs calculated from ORs and their baseline risks (among the studies that reported them) were not high enough for ORs to exaggerate RRs to a point that would prevent me from using ORs in quantitative bias analyses. Larger studies' associations were generally smaller than those of smaller studies. Small positive effects could be more susceptible to positive unmeasured confounding and measurement errors that bias away from the null than larger ones.

Unmeasured confounding offered a plausible alternative explanation for nearly half of the associations because studies often lacked control for important confounders of nicotine vaping and outcomes. For instance, only 10 of 17 studies controlled for substance use history,

which I identified as the most important of four groupings of potentially unmeasured confounders based on a review of substance use theory and prior research;[32,46–62] fewer than half of studies controlled for variables from the three other groupings. Moreover, effects of potential confounders on all kinds of substance use found in prior research are larger than most of the E-values I calculated in Table 2.2, which indicates susceptibility to unmeasured confounding.[51,53,56,58,91] For instance, severe externalizing behaviors, measures of psychopathology, increase nicotine vaping and cannabis use risk,[52,91,92] and parents' cannabis use, a measure of family and friends' influences, increases the odds of adolescent cigarette and cannabis use 2.7- and 3.1-fold, respectively.[53] Assuming *distributional monotonicity* and no other bias, my analysis suggests that nicotine vaping affects cannabis and other substance use, but effects are probably not as strong as reported in studies with even the strongest confounding control (ORs range from 1.37 to 2.16 in Table 2.3).[10,70,72]

When considered separately, different types of measurement error offer possible alternative explanations for associations. Independent nondifferential misclassification is a threat to validity in any study and generally biases towards the null when exposures and outcomes are binary, which was the case in all included studies. "Correcting" for this bias helps establish a revised association from which the impact of other biases can be assessed. Other research suggests my corrections may be accurate. In a simulation study, Levy *et al.* estimated the sensitivity of self-reported cannabis use to be about 78% in national data based on a small Chicago study (and assumed 100% specificity),[34,65] which is similar to the highest nicotine product sensitivity I could find (80%).[66] Dependent nondifferential misclassification may have the reverse effect. Specifically, social desirability, the tendency for study participants to provide inaccurate information to please interviewers, could cause bias away from the null.[88–90] Bias away from the null can occur due to the correlation of measurement errors,[86,87] which can occur if adolescents underreport all kinds of substance use for similar reasons (e.g., substance use is prohibited for adolescent or young adults, stigma associated with their

use).[88,93] Simulations correcting for both types of measurement error suggest that they would cancel out; therefore, measurement errors are unlikely to alternatively explain associations.

Bias from the expected process of differential loss to follow-up would not alternatively explain associations. The expected process based on prior research was that individuals reporting nicotine vaping and cannabis use are more likely than individuals reporting use of on substance (nicotine or cannabis use) or no use to be lost to follow-up.[35–38] Additionally, individuals reporting use of just one substance are expected to be lost more often than individuals reporting no substance use.[35–38] If differential loss to follow-up followed this process in the included studies, which would result in bias towards the null, the effect estimates appear to bear it out. The median ORs in studies that did and did not analyze the potential for differential loss to follow-up or preempt it were 1.97 and 1.20, respectively. The latter OR could reflect bias due to differential loss to follow-up.

I identified two studies that reported effect modification between nicotine vaping and cannabis initiation.[11,26] Dai *et al.* estimated stronger associations (ORs) in younger compared with older adolescents, and Seidel *et al.* estimated stronger associations (RRs) in adolescents with low sensation-seeking proclivity and non-smoking status than in their counterparts. All of these results were surprising, but Seidel *et al.* explained that adolescents with high sensation-seeking and a history of smoking may have been at a "risk ceiling,"[11] evidence of which exists in other substance use studies.[94,95] Dai *et al.*'s result may also be due to a risk ceiling.[26] Older adolescents, whose baseline risk was higher than that of younger adolescents,[26] are often at greater risk of engaging in many risky behaviors than their counterparts.[96] These results add context to the main effects if biases in strata of potential effect modifiers operate similarly to those in the full sample.

No study tested a mediation hypothesis, indicating that if there were a causal effect, the

process by which it occurs remains unexplained. Authors posited potential mechanisms, such as that e-cigarette experimentation increases youths' curiosity about cannabis use, reduces perceived harm of cannabis use, and increases social access to cannabis from peers and friends.[8,26] Some suggested that youth who experiment with e-cigarettes may use the same or newer devices to vape cannabis.[5,26,28,73] Biological mechanisms, such as that nicotine-induced changes to the central nervous system may reinforce pleasure from use of other substances and lead to dependence, were also discussed.[9,10,26,27] There is evidence that peers and peer networks are avenues connecting different kinds of substance use,[97,98] but the evidence is lacking in vaping literature. I know of no evidence that suggests familiarity with vaping equipment leads individuals from nicotine vaping to cannabis vaping or that nicotine induces central nervous system changes that lead to other substance use in humans (only animals).[99,100] None can be ruled out since the hypotheses have not been tested. Investigation of mechanisms would test the validity of total effects.

This review has some limitations. First, another reviewer assisted with screening articles and confirming eligibility, but only one reviewer conducted data extraction and possibly made errors. Second, I assumed that potentially unmeasured confounders from four groupings were the universe of variables necessary to control. Third, my assumption of distributional monotonicity nicotine vaping's potential effect on and cannabis and other substance use could be wrong. Fourth, I could be wrong about the extent of measurement error and/or the expected processes of loss to follow-up. Finally, I pooled all exposure measures together across analyses (e.g., ever, past year, past month), which likely have heterogeneous effects on the outcomes; however, my goal was to assess the qualitative effect and not to estimate a pooled effect. Despite these limitations, this review advanced an update of the body of literature on longitudinal associations between nicotine vaping and cannabis and other substance use.

2.5.1 Conclusions

To summarize, I reported effect estimates and their directions, argued why they were possibly biased towards or away from the null, and identified which effect modifiers were studied. Unmeasured positive confounding offered the best alternative explanation for associations in the included studies. Other biases were less likely to alternatively explain associations. Together, the few studies with strong positive estimates and comprehensive confounding control make up the most reliable body of evidence supporting nicotine vaping's effect on cannabis and other substance use.[10,70,72] As all the identified studies focused on effects among adolescents or adults over one to two years, future research should evaluate effects over many years from adolescence into young adulthood. Also, no studies evaluated how effects occur, so studies evaluating possible mechanisms are needed.

Chapter 3: Does nicotine vaping affect cannabis and other substance initiation from adolescence into young adulthood? Results from Wave 1 (2013-14) to Wave 5 (2018-19) of the Population Assessment of Tobacco and Health Study

3.1 Abstract

Background: Previous studies estimated longitudinal associations between nicotine vaping and subsequent initiation of cannabis and other substances (e.g., cocaine, heroin); however, their inferences may be limited by known unmeasured confounding and short follow-up periods. I examined whether there was evidence supporting effects of nicotine vaping on initiation of cannabis and other substances using a longer follow-up and stronger confounder control.

Methods: Youth aged 12 to 16 years old who never used cannabis or other substances, i.e., methamphetamine, speed, heroin, inhalants, solvents, hallucinogens, cocaine, crack, or used prescription substances nonmedically (N=9,571), were drawn from Wave 1 (baseline, 2013-14) of the Population Assessment of Tobacco and Health Study (hereon, PATH). Because the positivity assumption for causal identification was violated, I estimated incremental propensity score effects, which do not require the positivity assumption. This means I estimated effects of various shifts in the propensity of nicotine vaping (10% to 90% lower odds) on the cumulative risks of initiating cannabis and initiating other substances by Waves 2 (2014-15) through 5 (2018-19). I comprehensively adjusted estimates for known (baseline and time-varying) confounders that were unmeasured in prior studies in all analyses based on substance use

theories and prior research, and I calculated E-values (measures of whether unmeasured confounding may fully explain away effect estimates) as a sensitivity analysis.

Results: Cannabis and other substance initiation risks by the end of the study period (Wave 5) were 11.2 percentage points (95%CI: 9.5:12.8) and 10.1 (95%CI: 5.2:15.1) percentage points higher, respectively, than they would have been had nicotine vaping odds been 90% lower than observed in the prior wave. E-values suggested that known unmeasured confounders of associations between nicotine vaping and each outcome were unlikely to explain away most associations from Wave 3 forward.

Conclusions: I found evidence that nicotine vaping affected cannabis initiation and initiation of other substances over several years. These effects suggest that a population strategy that lowers nicotine vaping prevalence may help lower cannabis initiation and initiation of other substances.

3.2 Introduction

Nicotine vaping prevalence increased among adolescents and young adults during the 2010s as a novel way to consume nicotine and mimic cigarette smoking. In 2020, it decreased for the first time since national surveillance began;[2–4,101] however, 11% to 25% of students in grades eight through 12 and 20% to 22% of young adults aged 18 to 22 years old used e-cigarettes in the past month.[23,101] Research published since 2016 documents nicotine vaping's longitudinal associations with cannabis use and use of other substances, such as cocaine and heroin.[5–12] Regular use of these substances may cause psychosis and behavioral and cognition problems,[13] physiological dependence,[14] infectious disease,[15,16] sleep problems,[17] worsening interpersonal relationships,[18] poor job and school performance,[19] and unemployment.[20] Experimentation with the substances may indirectly lead to these problems through increasing the probability of regular use.[21]

Nearly all reported associations between nicotine vaping and cannabis and other substance use are positive,[25,102] but studies suffer from sources of bias expected to overestimate effects.[102] Unmeasured confounding is likely the most pernicious quantifiable source of bias across studies estimating these associations.[102] Fewer than half of studies controlled for variables from three or more groupings of known confounders based on a review of substance use theory, conceptual models, and prior research (e.g., substance use history, friends' and family's influences on substance use, psychopathology, and work or school factors).[32,46–61,102] Consequently, about half the effect estimates were likely biased in the positive direction.[102] More rigorous handling of confounding may result in more valid estimates of nicotine vaping's effects on cannabis initiation and other substance initiation. This includes motivating confounder selection with substance use theory, conceptual models, or research so that relevant confounders are controlled. Another approach is controlling for time-varying confounders when there are multiple time-points; prior studies' control for time-varying confounders may have been

inadequate.[6,10,103,104]

Evidence supporting the effects of nicotine vaping on initiation of cannabis and other substances is limited to studies with follow-up periods of less than three years among only adolescents or young adults;[10,70,102] however, potential long-term effects from adolescence into adulthood are unknown. Following adolescents into adulthood is important because, although most substance initiation occurs in adolescence,[105] a substantial minority occurs in young adulthood,[23] which would be missed if studying only the former period. Longer-term longitudinal studies may also help address potential unmeasured confounding because adult studies rarely consider events in adolescence, such as parental divorce, as formative to adult substance use risks.[7,28,70,71,73,106]

Long-term studies encompassing adolescence and young adulthood can also reveal more about potential effect modification by age. A longitudinal study from 2013 to 2015 found that nicotine vaping's effect on cannabis initiation was stronger for younger adolescents than for older adolescents.[26] Whether the stronger effect of nicotine vaping on cannabis initiation in younger versus older adolescents was constant or increased over several years is unknown. The latter would indicate the emergence of a cohort effect, evidence for which would be effect modification by both age and time. Therefore, investigating effect modification by age and by time can help bring attention to particular ages or cohorts at greater risk of nicotine vaping's potentially harmful effects.

Finally, the role of prior cigarette smoking and other non-vaping tobacco product use in nicotine vaping's effects during the transition from adolescence to adulthood is understudied. Smoking appears to have a stronger harmful effect on cannabis initiation than vaping (possibly because cannabis is also usually smoked);[102,107] however, two studies reported that vaping's effect on cannabis use was weaker among individuals who co-used tobacco products than individuals who did not.[11,73] The findings could be evidence of a ceiling effect; that is, the

baseline risk of cannabis use among people who with a history of using non-vaping tobacco products could be too high such that vaping adds little risk.[11] The studies were limited by small convenience samples; therefore, studying effect heterogeneity by tobacco product co-use in a larger representative sample could help clarify the group at greatest increased risk and most in need of public health attention.

In this nationally representative study, I:

1. examined whether the evidence supported nicotine vaping's effects on cannabis initiation and initiation of other substances among a cohort of adolescents as they aged into adulthood from 2013-14 to 2018-19,
2. examined whether potential time-specific effects differ by age,
3. examined whether potential age-specific effects differ across time,
4. examined whether potential effects differ by history of non-vaping tobacco product use,
5. and examined whether unmeasured confounding could alternatively explain effect estimates.

3.3 Methods

3.3.1 Population Assessment of Tobacco and Health Study (PATH)

Data come from Waves 1 (baseline, 2013-14) to Wave 5 (2018-19) of the ongoing PATH.[108] PATH was developed following the 2009 Family Smoking Prevention and Tobacco Control Act and is a collaboration between the National Institute on Drug Abuse at the National Institutes of Health, and the Center for Tobacco Products at the Food and Drug Administration.[108] PATH is a nationally representative cohort study of US adolescents and adults with the general purpose of evaluating tobacco use and tobacco-related health problems (e.g., respiratory, dental).[108] PATH's youth participants complete an annual survey about themselves and

a parent or guardian is asked to complete a survey mostly about their child. I was granted access to the restricted PATH data files to conduct the current study,[109] and the protocol was approved by the Columbia University Irving Medical Center Institutional Review Board.

3.3.2 Sampling design

PATH's sampling strategy was a four-stage stratified area probability sample.[108,109] A county or group of contiguous counties were selected as primary sampling units in the first stage based on a stratified random sample, which oversampled communities with high proportions of young adults, Black or African American adults, and people who use tobacco. In the second stage, primary sampling units were further stratified based on these demographic and behavioral characteristics into smaller geographic segments. At the third stage, residential addresses were sampled within segments. At the fourth stage, residents at those addresses were sampled. One adult residing in each household completed a household primary screener. Based on that screener and predetermined sampling rates of key subgroups (e.g., young adults, tobacco users, etc.), the screened adult and/or another adult were sampled to participate in a second screener and possibly a full interview. At most, two adults could be selected from each household for participation in the second screener. No more than two youths (aged 12 to 17 years old) were randomly sampled from each household in which an adult (parent or guardian) was sampled.[108] In other words, youth sampling required adult sampling from the same household, while adult sampling did not necessitate youth sampling.

3.3.3 Data collection

Data were collected annually (between 9 and 14 months) over four waves (Appendix B.1): Wave 1 (2013-14), Wave 2 (2014-15), Wave 3 (2015-16), and Wave 4 (2016-18).[108] Wave 5 (2018-19) data were collected about two years after Wave 4 data. Youth and adults completed specific instruments about themselves, including demographics, health, and substance use behaviors. Additionally, parents or guardians of youths completed an instrument mostly about their child.

Youth who reached 18 years old completed an adult interview.

3.3.4 Study population

The study population consisted of 9,571 youth aged 12 to 16 years old at baseline who reported never using cannabis products or other substances, i.e., methamphetamine, speed, heroin, inhalants, solvents, hallucinogens, cocaine, crack, or nonmedical use of prescription drugs, such as Ritalin, Adderall, painkillers, sedatives, and tranquilizers. Baseline 17-year-olds were excluded from the study population because most would have been adults (18 years old) in Wave 2, and past-year nicotine vaping was not measured among adults in that wave.

3.3.5 Estimand

My research question was “Does nicotine vaping increase the risk of initiating cannabis and initiating other substances over five years from adolescence into adulthood?"

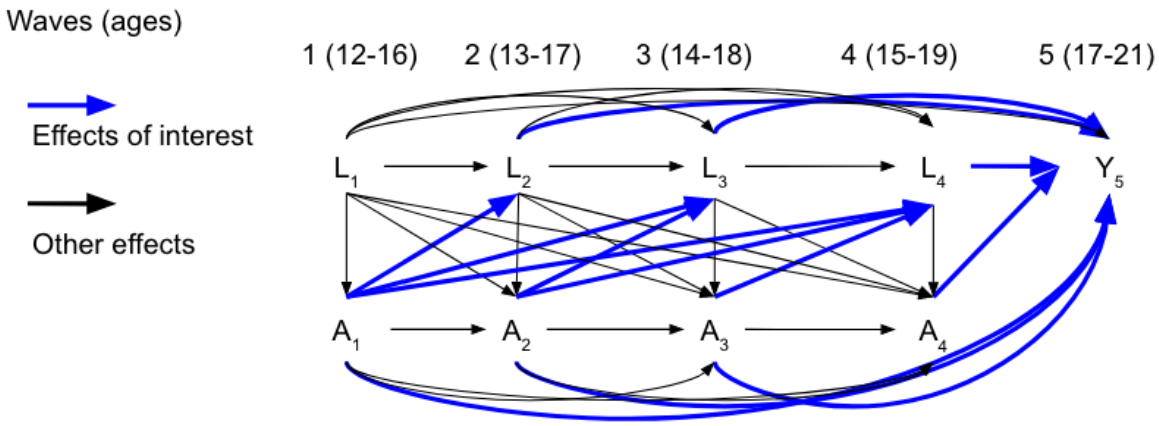

Figure 3.1: Causal diagram depicting assumed causal structure, where L_1 is a set of baseline confounders, A_t is time-varying nicotine vaping at Waves 1 through 4, nodes L_2, L_3, and L_4 comprise sets of time-varying confounders W_t and observed outcomes Y_t at Waves 2 through 4, and Y_5 is the observed outcome at Wave 5, and blue and black lines depict exposure and other effects, respectively

The diagram depicted in Figure 3.1 depicts the assumed causal structure for answering the research question. In the diagram, L_1 is a set of baseline confounders, A_t is nicotine vaping

at Waves 1 through 4, nodes L_2, L_3, and L_4 comprise sets of time-varying confounders W_t and observed outcomes Y_t at Waves 2 through 4, and Y_5 is the observed outcome at Wave 5. One way to answer the research question is to estimate longitudinal average treatment effects, which compare outcome risks given two counterfactual scenarios: had everyone vaped nicotine at every wave compared with everyone had they not vaped nicotine at every wave $\mathbb{E}\left(Y_t^{\bar{a}_{t-1}=1} - Y_t^{\bar{a}_{t-1}=0}\right)$, where Y_t denotes the binary outcome at time t, $\bar{a}_{t-1}$ denotes the history of binary nicotine vaping prior to time t, and $Y_t^{\bar{a}_{t-1}=1}$ and $Y_t^{\bar{a}_{t-1}=0}$ denote potential outcomes at time t had everyone vaped and not vaped nicotine, respectively, at every time point prior to time t. Identifying the longitudinal average treatment effect from observed data requires several causal identification assumptions. One verifiable assumption is positivity, the nonzero probability of being exposed at each time the exposure is measured, conditional on confounders (i.e., nonzero propensity scores). Near positivity violations (when propensity scores are near 0), also known as practical positivity violations, may result in finite sample bias and incorrect inference.[110] In a prior study of PATH data,[111] I found near positivity violations (Appendix Figure B.1).[111] For example, nearly a quarter of propensity scores were below 0.01.[111] Consequently, the longitudinal average treatment effect was not chosen as this study's estimand.

Approaches to restore positivity include reducing the number of sparse strata by reducing the number of confounders, truncating propensity scores, and trimming the sample to areas of overlap between propensity scores for exposed and unexposed groups;[110] however, restoring positivity with any of these approaches would compromise valid inference in the target population (all baseline youth aged 12 to 16 years old). Not controlling for confounders could bias effects. Bias from truncating propensity scores is usually tolerable if it is done for a few observations; however, a quarter of propensity scores would need to be truncated in my sample, which would result in model misspecification. Finally, I would remove a quarter of my sample if I were to trim propensity scores, and the resulting sample would no longer represent

the target population.

Another approach is to estimate a parameter that is identifiable without assuming positivity and which would still address my research question. The incremental propensity score effect is one such estimand.[112] It is the observed population average risk of an outcome compared with the population average risk of an outcome had the odds of exposure been shifted across a range of values.[112]

$$\mathbb{E}\left(Y_t - Y_t^{\delta \in (0.1,0.9)}\right) \tag{3.1}$$

In Expression 3.1, δ is a ratio of the shifted to unshifted odds of nicotine vaping $\left(\frac{\frac{q(l)}{1-q(l)}}{\frac{\pi(l)}{1-\pi(l)}}\right)$ where $q(l)$ and $\pi(l)$ are the shifted and unshifted propensity scores, respectively. In this study, $\delta < 1$, meaning $q(l) < \pi(l)$. Additionally, Y_t and $Y_t^{\delta \in (0.1,0.9)}$ in Expression 3.1 are observed and potential cannabis and other substance initiation outcomes, respectively. Specifically, $Y_t^{\delta \in (0.1,0.9)}$ are the outcomes that would have occurred had everyone's nicotine vaping odds been lower at the previous time point. Expression 3.1 is therefore the average difference by Wave $t = \{2,3,4,5\}$ between observed cannabis and other substance initiation risks $\mathbb{E}\left(Y_t\right)$ and cannabis and other substance initiation risks had nicotine vaping odds been 0.1 to 0.9 times the observed odds in the prior wave $\mathbb{E}\left(Y_t^{\delta \in (0.1,0.9)}\right)$ (90% to 10% lower).

Positivity can fail for the average treatment effect and holds by definition for the incremental propensity score effect. For simplicity, I will explain why using notation for the single timepoint setting. Estimates for parameters of the average treatment effect can be written as $\mathbb{E}\left(Y|L = l, A = a\right) = \mu(l,a)$, which are undefined if $\pi(l) = 0$ or $1 - \pi(l) = 0$ because of conditioning on a zero-probability event (i.e., $P(A = a|l) = 0$, positivity fails). When $\pi(l) \approx 0$ or $\pi(l) \approx 0$, estimates of $\mu(l,a)$ are defined but could be heavily biased. To show that positivity holds by definition when estimating incremental propensity score effects one can write the identifying expression of $\mathbb{E}(Y^{\delta})$ in Expression 3.1 as

$$\psi(\delta) = \mathbb{E}\left[\frac{\delta\pi(\mathbf{L})\mu(\mathbf{L},1) + 1 - \pi(\mathbf{L})\mu(\mathbf{L},0)}{\delta\pi(\mathbf{L}) + 1 - \pi(\mathbf{L})}\right] \tag{3.2}$$

where $\mu(L,1)$ and $\mu(L,0)$ are the average of predicted outcomes if everyone had A and did not have A, respectively, conditional on covariates. In Equation 3.2, $\psi(\delta)$ is defined because any subjects for whom $\pi(l) = 0$ or $1 - \pi(l) = 0$ receive weights of $q(l) = 0$ or $1 - q(l) = 0$, respectively, and subjects for whom $\pi(l) \approx 0$ or $1 - \pi(l) \approx 0$ receive very small weights when the expectation over the covariates is taken (i.e., positvity holds). In other words, $q(l) = \frac{\delta\pi(l)}{\delta\pi(l)+1-\pi(l)}$ and $1 - q$ weight predicted outcomes $\mu(L,1)$ and $\mu(L,0)$, respectively, when $\pi(l) \neq 0$ or $1 - \pi(l) \neq 0$. When propensity scores are farther from 0 or 1, $q(l)$ and $1 - q(l)$ weight the predicted outcomes more heavily.

When $\pi = 0$ or $1 - \pi = 0$, predicted outcomes remain unweighted. That is, participants who have $\pi = 0$ will have their predicted outcome $(\hat{Y}|a = 0)$, and participants who have $\pi = 1$ will have their predicted outcome $(\hat{Y}|a = 1)$ regardless of δ. Predicted outcomes for these individuals would cancel out in Expression 3.1.[112,113]

Another way of seeing that positivity always holds is by writing $\mathbb{E}(Y^{\delta})$ in Expression 3.1 as

$$\psi(\delta) = \mathbb{E}\{\mu(\mathbf{L}, A_{\delta})\} \tag{3.3}$$

where A_{δ} is a simulated version of incremental exposure $(A_{\delta}|\mathbf{L} = \mathbf{l})$ ~Bernoulli$\{q(l)\}$); In other words, A_{δ} are shifted propensity scores. Expression 3.3 is therefore the average outcome when setting the value of δ, conditional on confounders; it is always defined because A_{δ} is always defined for $p(l) > 0 \implies (A = A_{\delta}|l) > 0$ (i.e., positivity holds because we always condition on a nonzero-probability event).

Positivity holds but effects are unstable if δ gets too close to 0 or too large. The former would

have happened if I were to compare observed risks with risks if $\delta < 0.1$ (i.e., >90% lower odds); therefore, $\delta = 0.1$ was the smallest comparator δ when estimating main effects, and $\delta = 0.2$ was the smallest comparator in effect modification analyses.

3.3.6 Identification

I assumed no unmeasured time-invariant and time-varying confounding of nicotine vaping and outcomes to identify the estimand in Expression 3.1 (i.e., effects by Waves 2, 3, 4, and 5). No unmeasured time-varying confounding means that potential outcomes by each wave are independent of the exposure at the prior wave, conditional on histories of measured confounders, nicotine vaping, and outcomes.

3.3.7 Measures

Outcomes

There were two binary outcomes, each assessed in the past 12 months by each follow-up wave: 1) cannabis initiation was defined as use of THC, grass, pot, or weed, and 2) other substance initiation was defined as nonmedical use of prescription drugs, such as Ritalin, Adderall, painkillers, sedatives, or tranquilizers, or use of methamphetamine, speed, heroin, inhalants, solvents, hallucinogens, cocaine or crack. Once participants responded positively for either outcome, they were considered no longer at risk for the duration of follow-up. Despite alcohol being psychoactive, addictive, and prohibited for anyone under 21 years old, alcohol initiation was not considered as an outcome. Alcohol initiation usually precedes nicotine vaping initiation, or both are initiated concurrently,[41–43] which, as others have found,[44] makes estimating nicotine vaping's effects on alcohol initiation challenging.

Exposure

At Wave 1, participants answered whether they used e-cigarettes in the past 12 months. E-cigarettes were described as devices look like regular cigarettes, but are battery-powered and produce

vapor instead of smoke. Some common brands include NJOY, Blu, and Smoking Everywhere'. At Waves 2 through 4, they answered whether they used electronic nicotine products in the past 12 months. Examples of electronic nicotine products were described as as e-cigarettes, e-cigars, e-pipes, e-hookahs, and personal vaporizers, as well as vape pens and hookah pens that are battery-powered, use nicotine fluid rather than tobacco leaves, and produce vapor instead of smoke. Some common brands include Fin, NJOY, Blu, e-Go and Vuse'. Participants were shown generic pictures of the product at all three waves. I refer to affirmative responses to these questions as nicotine vaping, a binary variable, throughout.

Confounders

I considered several substance use theories and prior research to help guide confounder selection.[47–50] The theories I considered were problem behavior theory,[46] interactive theory,[50] and common liability theory.[114] I also considered the context of substance use using the drug, set, and setting framework because the aforementioned theories are context-agnostic.[115] Based on this process, I considered the most important confounders to be one's own substance use history, family and friends' influences on substance use, psychopathology, and school or employment factors.[102] For instance, I controlled for tobacco product use other than nicotine vaping, alcohol and other substance use, whether the youth lives with both parents, whether the parent(s) allows tobacco in the home, sensation-seeking tendencies (e.g., I like to do frightening things), internalizing and externalizing problems, school performance, and others in outcome models; all confounders in outcome models were included in propensity score models. In Appendix B.1, I describe my approach to confounding in greater detail. In Appendix Table B.2, I show the selected confounders and how they were measured.

Effect modifiers

I considered age and time to be potential positive effect modifiers such that increasing age and the passage of time would increase effects of nicotine vaping on the outcomes. Age was

self-reported at baseline. I calculated age at subsequent waves as age at baseline plus the number of years between baseline and the subsequent wave. Time was the difference in years between PATH waves. From Waves 1 to 4, there were one-year increments, and Wave 5 occurred about two years after Wave 4.

3.3.8 Statistical analysis

Missing data

Missing exposure data increased from 0.5% in Wave 1 to 27.7% in Wave 4, and missing data for both outcomes increased from about 13% in Wave 2 to about 28% in Wave 5 (Tables B.3-B.5). I estimated missing data with multiple imputation by chained equations using the R package **mice**.[116] Each of the 15 imputed dataframes were iterated 25 times using predictive mean matching. Estimates were combined using Rubin's rules with the R package **Amelia** for both descriptive and causal analyses described below.[117]

Descriptive analysis

I estimated the survey-weighted distributions of baseline confounders, both in the overall sample and stratified by baseline nicotine vaping status. I also estimated the survey-weighted prevalence of nicotine vaping in Waves 1 through 4.

Causal analysis

I estimated observed survey-weighted cannabis and other substance initiation risks by Waves 2 through 5, controlling for measured baseline confounding of nicotine vaping and the outcomes at all waves and measured baseline and time-varying confounding when estimating risks by Waves 3 through 5. I also estimated survey-weighted outcome risks by Waves 2 through 5 had the population's nicotine vaping odds been incrementally lower in the prior wave, controlling for measured confounding. Incremental propensity score effects by Waves 2 through 5 were differences in observed outcome risks and the outcome risks had nicotine vaping odds been

lower than observed. I used the R package **ipsi** and the SuperLearner machine learning algorithm to estimate propensity scores, outcome risks, risk differences, and 95% confidence intervals for risks and risk differences.[112] SuperLearner weights candidate algorithms' predicted outcomes by how much they minimize cross-validated loss functions (e.g., mean squared error in linear regression models).[118–120] Cross-validation is used to avoid overfitting. In cross-validation, data are split into k folds (i.e., equal splits of the data) using $k-1$ folds to train the data and the remaining "hold-out" fold to predict outcomes.[120,121] In my study, $k=5$, so each algorithm predicted five average outcomes and five corresponding loss functions, one for each hold-out fold.[120,121] SuperLearner predicts a single weighted average outcome based on each algorithm's performance (i.e., loss minimization).[120,121] I specified the following candidates in the SuperLearner library (`SL.Lib` argument): sample mean (`SL.mean`), generalized linear models (`SL.glm`), and multivariate adaptive regression splines (`SL.earth`). Standard errors were estimated by taking the square root of the sample variance of the influence function. To ensure that estimates were nationally representative, I modified **ipsi**'s code to accommodate survey sampling weights.[112] Additionally, I modified **ipsi**'s code to ensure that anyone who initiated cannabis or other substances was no longer at risk for the remainder of the study period. See Appendix B.2 for revised **ipsi** code.[112]

I also estimated possible effect modification across age categories within each PATH wave, across waves within each age category, and by non-vaping tobacco product co-use by subtracting risk difference estimates (to calculate differences in risk differences). I calculated standard errors for differences in risk differences by taking the square root of the sum of squared standard errors of each risk difference $\left(\sqrt{RD_1^2 + RD_2^2}\right)$.[122] If the contrasted risk differences are independent, the influence curve-based standard errors would be accurate; however, if the estimates were not independent, the standard errors would be too narrow.[122]

In a sensitivity analysis, I estimated E-values on the risk ratio scale and their lower 95% confidence

limits. An E-value is the smallest effect that a set of unmeasured confounders must have on an exposure and on an outcome to fully explain away a specific exposure-outcome association, conditional on measured confounders.[63]

3.4 Results

3.4.1 Distributions of exposure and confounders at baseline

There were 9,571 eligible baseline participants aged 12 to 16 years old included in the sample. Only 3.4% of them vaped nicotine products in the past year (Table 3.1) at baseline (2013-14), which may reflect removing anyone with prevalent cannabis or other substance use at Wave 1 from the analysis. The sample was three-quarters white and included slightly more boys (52%) than girls (47%). Two-thirds of parents were currently married. About 40% of youth reported that a parent or guardian lived separately from them. About 20% to 32% of youth reported mild behavioral or mood problems, while slightly less reported a willingness to break rules and do frightening things. Over 30% of respondents reported that tobacco is used by someone other than the youth in the home and that tobacco is not allowed in the home.

Table 3.1: Distribution of baseline confounders (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health)

			95% Confidence Interval[d]	
	N[a]	**Proportion (%)[d]**	**Lower (%)**	**Upper (%)**
Electronic nicotine product use/nicotine vaping				
No past-year or none	9,218	96.59	96.11	97.07
Past-year	302	3.41	2.93	3.89
Racialized groups				
White	6,751	74.38	71.89	76.88
Black	1,437	14.43	12.48	16.37
Asian and Other, including multi-racial	309	11.19	9.80	12.58
Sex[b]				
Male	4,979	52.03	50.95	53.11
Female	4,592	47.97	46.89	49.05

Continued on next page

Table 3.1: Distribution of baseline confounders (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) (cont.)

			95% Confidence Interval[d]	
	N[a]	Proportion (%)[d]	Lower (%)	Upper (%)
Household income				
Up to $50,000	3,671	40.68	38.28	43.08
More than $50,000	3,696	59.32	56.92	61.72
Parent education				
Up to some college but no degree	5,753	52.49	50.14	54.84
Post-high school degrees or graduate/professional education	3,760	47.51	45.16	49.86
Parent currently married				
Currently married	6,216	66.17	64.43	67.90
Not currently married	3,315	33.83	32.10	35.57
A parent lives elsewhere / youth sees them				
No	5,850	61.68	59.89	63.47
Yes, child sees them	1,962	22.36	20.98	23.75
Yes, child never sees them	1,714	15.96	14.80	17.12
Doctor told child they had ADHD[c](parent report)				
Never	8,301	86.31	85.25	87.36
Ever	1,228	13.69	12.64	14.75
Youth has curfew				
No	1,041	11.61	10.62	12.6
Yes	8,495	88.39	87.4	89.38
Parent's reaction if youth used tobacco				
Very upset	9,151	96.29	95.79	96.78
Not too upset or no reaction	364	3.71	3.22	4.21
Parent talked to child about not using tobacco in past 12 months				
Yes	4,880	50.55	49.38	51.73
No	4,590	49.45	48.27	50.62
Youth school performance (grades)				
Mostly B's or worse	3,548	34.36	32.97	35.76
Mostly A's and B's or better	5,979	65.64	64.24	67.03
Youth felt depressed at least twice in past year				
No	7,319	77.81	76.72	78.90
Yes	2,129	22.19	21.10	23.28
Youth had trouble sleeping at least twice in past year				
No	6,408	67.90	66.72	69.08
Yes	3,076	32.10	30.92	33.28

Continued on next page

Table 3.1: Distribution of baseline confounders (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) (cont.)

			95% Confidence Interval[d]	
	N[a]	Proportion (%)[d]	Lower (%)	Upper (%)
Youth felt anxious at least twice in past year				
No	6,549	69.13	68.06	70.21
Yes	2,923	30.87	29.79	31.94
Youth felt distressed about past events at least twice in past year				
No	7,310	77.64	76.58	78.70
Yes	2,154	22.36	21.30	23.42
Youth lied or con to get something at least twice in past year				
No	7,144	76.56	75.47	77.65
Yes	2,221	23.44	22.35	24.53
Youth had trouble paying attention at school/work/home at least twice in past year				
No	6,456	67.83	66.45	69.22
Yes	3,020	32.17	30.78	33.55
Youth had trouble listening to instructions at school/work/home at least twice in past year				
No	7,155	74.98	73.81	76.16
Yes	2,329	25.02	23.84	26.19
Youth bullied or threatened people at least twice in past year				
No	9,089	96.09	95.67	96.51
Yes	388	3.91	3.49	4.33
Youth started physical fights at least twice in past year				
No	9,222	97.17	96.72	97.62
Yes	284	2.83	2.38	3.28
Youth likes to do frightening things				
No	8103	85.14	84.25	86.04
Yes	1,368	14.86	13.96	15.75
Youth likes new and exciting experiences, even if they have to break rules				
No	7,876	82.43	81.44	83.43
Yes	1,630	17.57	16.57	18.56

Continued on next page

Table 3.1: Distribution of baseline confounders (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) (cont.)

			95% Confidence Interval[d]	
	N[a]	Proportion (%)[d]	Lower (%)	Upper (%)
Youth prefers friends who are exciting and unpredictable				
No	6,707	71.63	70.41	72.85
Yes	2,742	28.37	27.15	29.59
Youth tobacco use in past month[d]				
No	8,928	98.15	97.73	98.57
Yes	161	1.85	1.43	2.27
Youth sexual orientation				
Straight	4,876	94.92	94.19	95.66
Not straight	274	5.08	4.34	5.81
Youth's overall health (parent report)				
Good, very good, or excellent	8,298	89.70	88.78	90.62
Fair or poor	1,238	10.30	9.38	11.22
Someone other than youth uses tobacco in the home				
No	6,375	68.96	66.72	71.20
Yes	3,104	31.04	28.80	33.28
Tobacco prohibited in home				
Yes	6,692	69.39	67.48	71.30
No	2,833	30.61	28.70	32.52
Youth ever used alcohol				
Never	6,976	71.52	70.04	73.01
Ever	2,539	28.48	26.99	29.96

[a] N = Unweighted frequency

[b] Responses imputed from valid responses were male, female, don't know, and refused. Participants could have also requested that their response be removed.

[c] ADHD = Attention deficit hyperactivity disorder

[d] All proportions are weighted to reflect the complex survey design of the PATH and are a weighted average based on 15 imputed datasets to account for missingness. Frequency sums are different and do not add to the baseline cohort total of 9,571 for some variables because they are extracted from the original baseline wave that has missingness. Proportions are weighted by the baseline probability of being sampled.

[d] Tobacco was defined as any cigarettes, any cigar, cigarillos, filtered cigars, pipes, hookah, or smokeless tobacco in Wave 1.

3.4.2 Distributions of baseline confounders stratified by baseline nicotine vaping status

Nicotine-vaping youth were more likely to be white, male, and from lower income households than non-nicotine vaping youth (Table 3.2). They were also more likely to have parents who were unmarried, had less educational attainment, and would not be upset if the youth used

tobacco. Nicotine-vaping youth were more likely to report some types of behavioral or mood problems and sensation-seeking predisposition, tobacco use in the past month, residing with someone who used tobacco in the home, living in a home in which tobacco was not allowed, and lifetime alcohol use.

Table 3.2: Distribution of baseline confounders by baseline nicotine vaping status (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health)

	No past-year nicotine vaping				Past-year nicotine vaping			
			95% Confidence Interval[d]				95% Confidence Interval[d]	
Proportion (%)[d]	**Lower (%)**	**Upper (%)**	**N[a]**	**Proportion (%)**	**Lower (%)**	**Upper (%)**		
Racialized groups								
White	6,478	74.16	71.63	76.69	236	80.58	75.90	85.26
Black	1,401	14.56	12.58	16.54	29	10.59	7.14	14.03
Asian and Other, including multi-racial	305	11.28	9.86	12.69	34	8.83	5.14	12.52
Sex[b]								
Male	4,768	51.69	50.63	52.75	183	61.57	55.32	67.82
Female	4,450	48.31	47.25	49.37	119	38.43	32.18	44.68
Household income								
Up to $50,000	3,527	40.37	37.97	42.77	120	49.40	41.64	57.17
More than $50,000	3,578	59.63	57.23	62.03	101	50.60	42.83	58.36
Parent education								
Up to a bachelor's degree	5,518	52.24	49.95	54.54	195	59.42	51.06	67.78
Some graduate or professional education or degree	3,644	47.76	45.46	50.05	105	40.58	32.22	48.94
Parent currently married								
Currently married	6,009	66.44	64.68	68.20	178	58.50	51.51	65.50
Not currently married	3,171	33.56	31.80	35.32	122	41.50	34.50	48.49
A parent lives elsewhere / youth sees them								
No	5,657	61.96	60.17	63.75	163	53.71	46.68	60.73

Continued on next page

Table 3.2: Distribution of baseline confounders by baseline nicotine vaping status (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) (cont.).

	No past-year nicotine vaping				Past-year nicotine vaping			
			95% Confidence Interval[d]				95% Confidence Interval[d]	
	N[a]	Proportion (%)[d]	Lower (%)	Upper (%)	N[a]	Proportion (%)	Lower (%)	Upper (%)
Yes, child sees them	1,867	22.12	20.72	23.51	83	29.40	23.94	34.86
Yes, child never sees them	1,651	15.92	14.75	17.10	54	16.90	12.01	21.78
Doctor told child they had ADHD[c](parent report)								
Never	8,004	86.38	85.29	87.47	256	84.27	79.86	88.69
Ever	1174	13.62	12.53	14.71	44	15.73	11.31	20.14
Youth has curfew								
No	1,012	11.71	10.68	12.74	24	8.73	4.94	12.53
Yes	8,173	88.29	87.26	89.32	276	91.27	87.47	95.06
Parent's reaction if youth used tobacco								
Very upset	8,866	96.93	96.51	97.36	241	77.91	71.35	84.48
Not too upset or no reaction	302	3.07	2.64	3.49	58	22.09	15.52	28.65
Parent talked to child about not using tobacco in past 12 months								
Yes	4,680	50.21	49.00	51.42	175	60.14	54.06	66.21
No	4441	49.79	48.58	51.00	126	39.86	33.79	45.94
Youth school performance (grades)								
Mostly B's or worse	3,366	33.84	32.49	35.20	155	49.05	41.73	56.36
Mostly A's and B's or better	5,811	66.16	64.80	67.51	144	50.95	43.64	58.27
Youth felt depressed at least twice in past year								

Continued on next page

Table 3.2: Distribution of baseline confounders by baseline nicotine vaping status (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) (cont.).

	No past-year nicotine vaping				Past-year nicotine vaping			
			95% Confidence Interval[d]				95% Confidence Interval[d]	
	N[a]	Proportion (%)[d]	Lower (%)	Upper (%)	N[a]	Proportion (%)	Lower (%)	Upper (%)
No	7,082	78.24	77.18	79.29	196	65.72	60.17	71.27
Yes	2,020	21.76	20.71	22.82	103	34.28	28.73	39.83
Youth had trouble sleeping at least twice in past year								
No	6,219	68.43	67.28	69.58	154	53.02	47.32	58.72
Yes	2,919	31.57	30.42	32.72	147	46.98	41.28	52.68
Youth felt anxious at least twice in past year								
No	6,349	69.55	68.48	70.63	163	57.23	51.04	63.42
Yes	2,777	30.45	29.37	31.52	137	42.77	36.58	48.96
Youth felt distressed about past events at least twice in past year								
No	7,091	78.29	77.30	79.27	177	59.27	53.5	65.03
Yes	2,029	21.71	20.73	22.70	121	40.73	34.97	46.50
Youth lied or con to get something at least twice in past year								
No	6,927	77.20	76.13	78.27	177	58.42	51.73	65.10
Yes	2,097	22.80	21.73	23.87	121	41.58	34.90	48.27
Youth had trouble paying attention at school/work/home at least twice in past year								

Continued on next page

Table 3.2: Distribution of baseline confounders by baseline nicotine vaping status (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) (cont.).

	No past-year nicotine vaping				Past-year nicotine vaping			
			95% Confidence Interval[d]				95% Confidence Interval[d]	
	N[a]	Proportion (%)[d]	Lower (%)	Upper (%)	N[a]	Proportion (%)	Lower (%)	Upper (%)
No	6,254	68.29	66.86	69.71	165	55	47.97	62.03
Yes	2,875	31.71	30.29	33.14	136	45	37.97	52.03
Youth had trouble listening to instructions at school/work/home at least twice in past year								
No	6,926	75.37	74.19	76.55	191	63.96	57.48	70.44
Yes	2,213	24.63	23.45	25.81	107	36.04	29.56	42.52
Youth bullied or threatened people at least twice in past year								
No	8,770	96.25	95.84	96.67	275	91.43	88.09	94.77
Yes	362	3.75	3.33	4.16	25	8.57	5.23	11.91
Youth started physical fights at least twice in past year								
No	8,894	97.31	96.86	97.76	282	93.14	89.91	96.37
Yes	265	2.69	2.24	3.14	18	6.86	3.63	10.09
Youth likes to do frightening things								
No	7,848	85.62	84.70	86.53	210	71.59	66.02	77.16
Yes	1,272	14.38	13.47	15.3	91	28.41	22.84	33.98

Continued on next page

Table 3.2: Distribution of baseline confounders by baseline nicotine vaping status (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) (cont.).

	No past-year nicotine vaping				Past-year nicotine vaping			
			95% Confidence Interval[d]				95% Confidence Interval[d]	
	N[a]	Proportion (%)[d]	Lower (%)	Upper (%)	N[a]	Proportion (%)	Lower (%)	Upper (%)
Youth likes new and exciting experiences, even if they have to break rules								
No	7,665	83.41	82.44	84.38	165	54.87	48.35	61.39
Yes	1,491	16.59	15.62	17.56	136	45.13	38.61	51.65
Youth prefers friends who are exciting and unpredictable								
No	6,523	72.42	71.18	73.66	144	49.11	43.06	55.15
Yes	2,578	27.58	26.34	28.82	157	50.89	44.85	56.94
Youth tobacco use in past month[d]								
No	8,715	99.00	98.72	99.28	213	74.19	67.68	80.69
Yes	87	1.00	0.72	1.28	74	25.81	19.31	32.32
Youth sexual orientation								
Straight	4,640	95.02	94.25	95.79	214	92.17	88.12	96.22
Not straight	256	4.98	4.21	5.75	17	7.83	3.78	11.88
Youth's overall health (parent report)								
Good, very good, or excellent	7,998	89.69	88.75	90.63	260	89.88	86.08	93.69
Fair or poor	1,187	10.31	9.37	11.25	40	10.12	6.31	13.92
Someone other than youth uses tobacco in the home								

Continued on next page

Table 3.2: Distribution of baseline confounders by baseline nicotine vaping status (9,571 baseline 12- to 16-year-old youth meeting cohort inclusion criteria, Population Assessment of Tobacco and Health) (cont.).

	No past-year nicotine vaping				Past-year nicotine vaping			
			95% Confidence Interval[d]				95% Confidence Interval[d]	
	N[a]	Proportion (%)[d]	Lower (%)	Upper (%)	N[a]	Proportion (%)	Lower (%)	Upper (%)
No	6,227	69.90	67.71	72.08	117	42.48	35.16	49.80
Yes	2,903	30.10	27.92	32.29	184	57.52	50.20	64.84
Tobacco prohibited in home								
No	6,496	69.83	67.96	71.70	164	57.05	49.42	64.69
Yes	2,678	30.17	28.30	32.04	136	42.95	35.31	50.58
Youth ever used alcohol								
Never	6,830	72.84	71.40	74.29	105	34.21	28.28	40.15
Ever	2,333	27.16	25.71	28.60	197	65.79	59.85	71.72

[a] N = Unweighted frequency

[b] Responses imputed from valid responses were male, female, don't know, and refused. Participants could have also requested that their response be removed.

[c] ADHD = Attention deficit hyperactivity disorder

[d] All proportions are weighted to reflect the complex survey design of the PATH and are a weighted average based on 15 imputed datasets to account for missingness. Frequency sums are different and do not add to the baseline cohort total of 9,571 for some variables because they are extracted from the original baseline wave that has missingness. Proportions are weighted at the baseline probability of being sampled.

[d] Tobacco was defined as any cigarettes, cigars, cigarillos, filtered cigars, pipes, hookah, or smokeless tobacco in Wave 1.

3.4.3 Nicotine vaping prevalence over time

Past-year nicotine vaping prevalence increased from 3.4% in Wave 1 to 10% in Wave 2. The prevalence was 14% in Wave 3 and 15% in Wave 4 (Figure 3.2).

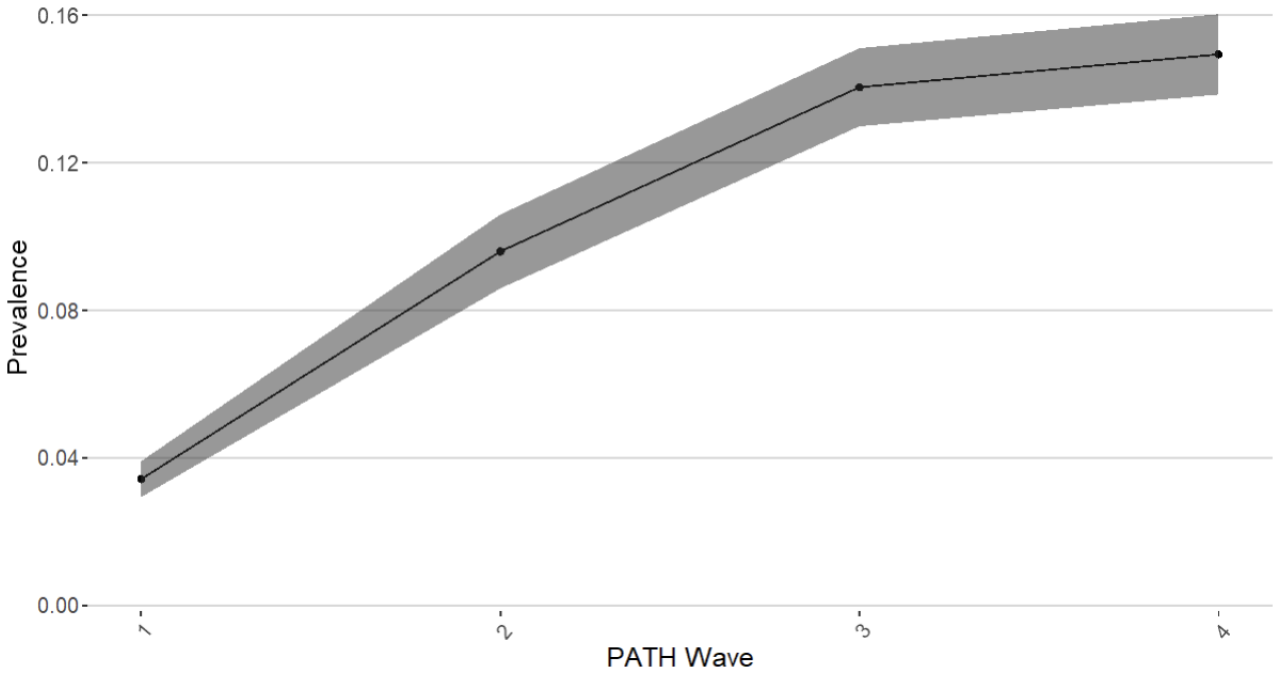

Figure 3.2: Nicotine vaping prevalence across PATH Waves 1 through 4 (N=9,571 baseline participants 12 to 16 years old)

3.4.4 Risks of cannabis initiation by Waves 2 through 5

Cannabis initiation risks at all follow-up waves for values of δ ranging from 0.1 to 1 in 0.1-unit increments are presented in Figure 3.3 (left- and right-most points on each plot, respectively). The following are the observed risks by Waves 2 (2014-15) through 5 (2018-19): 7.7% (95%CI: 7.1%:8.3%), 15.1% (95%CI: 14.3%:15.9%), 25.2% (95%CI: 24.2%:26.1%), and 39.7% (95% confidence interval (CI): 38.6%:40.8%). Had nicotine vaping odds been 90% lower than observed ($\delta = 0.1$), the risks by Waves 2 through 5 would have been: 7.2% (95%CI: 6.5%:7.8%), 11.6% (95%CI: 10.6%:12.6%), 17.2% (95%CI: 15.5%:18.9%), and 28.5% (95%CI: 26.6%:30.4%).

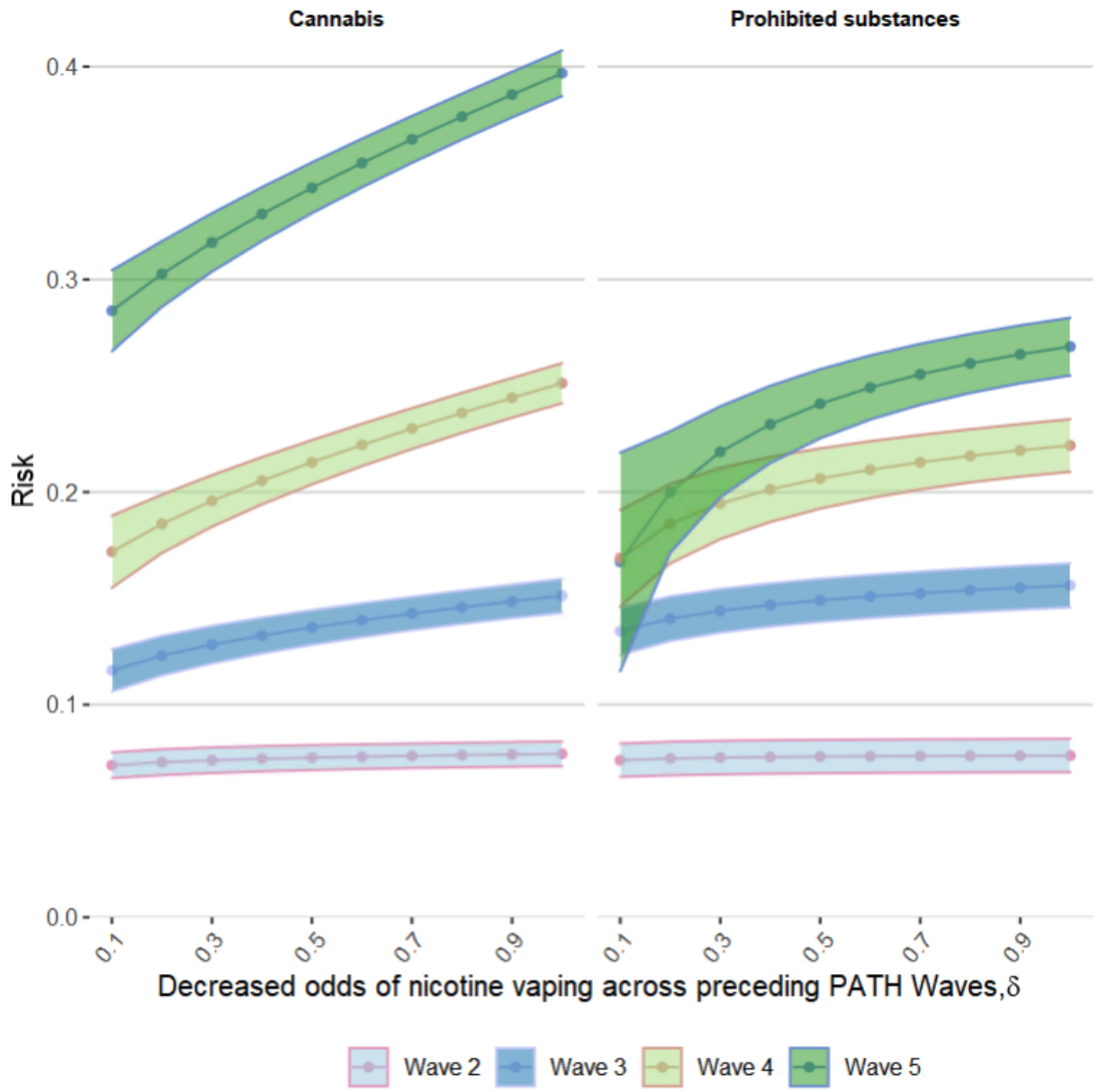

Figure 3.3: Past-year observed cannabis initiation risks and past-year observed other substance initiation risks and 95% confidence intervals (right-most dots) by each follow-up wave and those risks had nicotine vaping odds been 10% to 90% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.1$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old

3.4.5 Risks of other substance initiation by Waves 2 through 5

Other substance initiation risks for all values of δ ranging from 0.1 to 1 are also presented in Figure 3.3 (left- and right-most points, respectively). The following are the observed risks by Waves 2 through 5: 7.6% (95%CI: 6.8%:8.4%), 15.6% (95%CI: 14.6%:16.6%), 22.2% (95%CI: 21.0%:23.4%), and 26.8% (95%CI: 25.5%:28.2%). Had nicotine vaping odds been 90% lower than observed ($\delta = 0.1$), the risks by Waves 2 through 5 would have been: 7.3% (95%CI: 6.6%:8.1%), 13.5% (95%CI: 12.3%:14.6%), 16.9% (95%CI: 14.6%:19.1%), and 16.7% (95%CI: 11.6%:21.9%).

3.4.6 Incremental propensity score effects on cannabis initiation by Waves 2 through 5

Incremental propensity score effects on cannabis initiation by each follow-up wave (i.e., observed risk compared with risk if $\delta = 0.1$ to $\delta = 0.9$) are shown in the left panel of Figure 3.4 (left- and right-most points respectively). By Wave 2, the observed cannabis initiation risk was 0.5 percentage points (95%CI: 0.2:0.8) higher than it would have been had nicotine vaping odds been 90% lower in Wave 1 ($\delta = 0.1$). By Waves 3 and 4, the observed risk was 3.5 (95%CI: 2.7:4.3) and 7.9 (95%CI: 6.4:9.4) percentage points higher than it would have been had nicotine vaping odds been 90% lower in Waves 2 and 3, respectively. By Wave 5, the observed risk was 11.2 percentage points (95%CI: 9.5:12.8) higher than it would have been had nicotine vaping odds been 90% lower in Wave 4.

3.4.7 Incremental propensity score effects on other substance initiation by Waves 2 through 5

Incremental propensity score effects on other substance initiation by each follow-up wave (i.e., observed risk compared with risk if $\delta = 0.1$ to $\delta = 0.9$) are shown in the right panel of Figure 3.4 (left- and right-most points, respectively). By Wave 2, the observed other substance initiation risk was 0.2 percentage points (95%CI: -0.01:0.4) higher than it would have been had nicotine vaping odds been 90% lower in Wave 1 ($\delta = 0.1$). By Waves 3 and 4, the observed risk was 2.2 percentage points (95%CI: 1.5:2.8) and 5.3 percentage points (95%CI: 3.4:7.2) higher than it

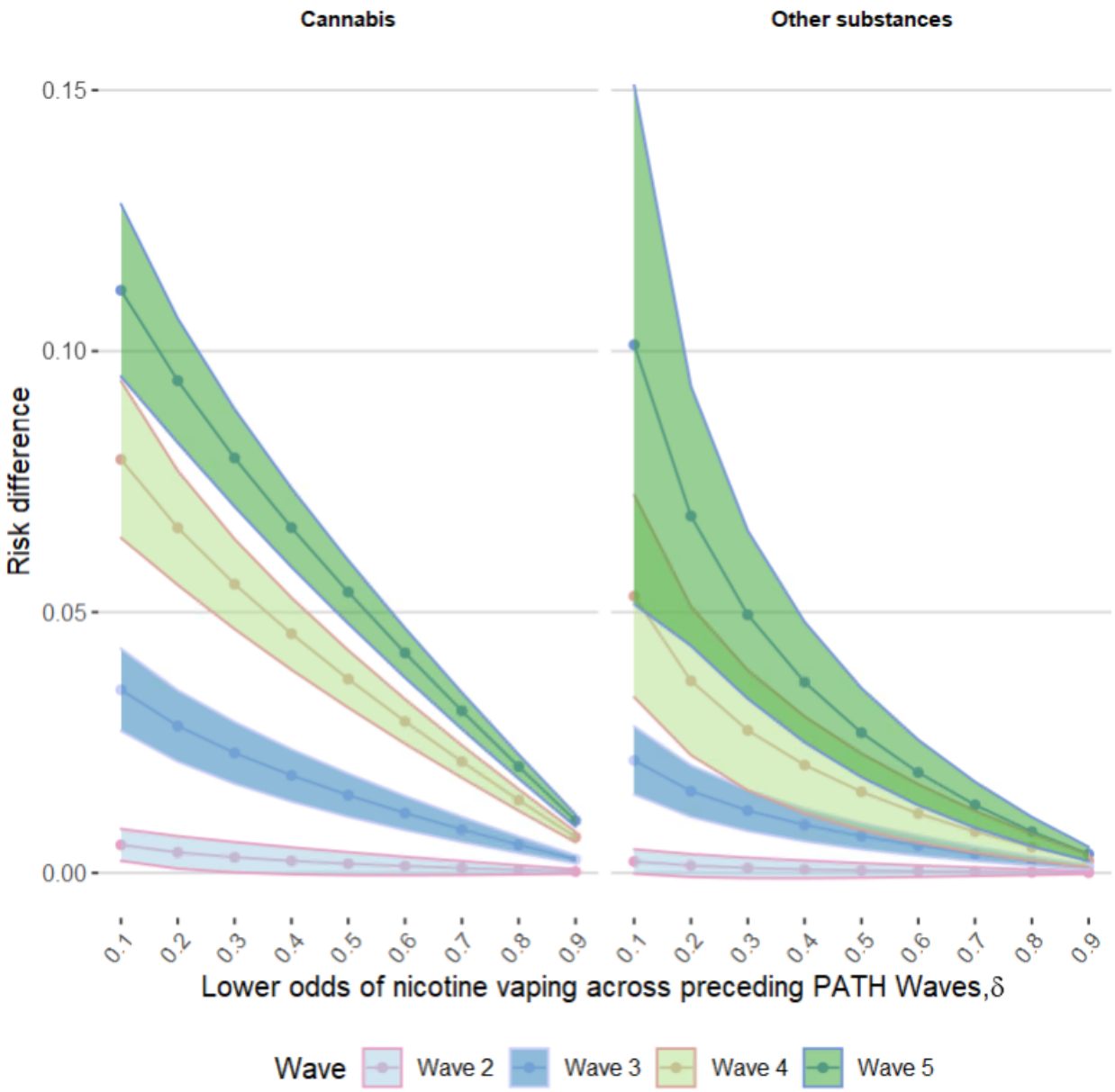

Figure 3.4: Risk differences and 95% confidence intervals comparing past-year observed cannabis initiation risks and past-year observed other substance initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 90% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.1$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old.
Example: where $\delta = 0.1$ on plot, risk difference shown is $\mathbb{E}\left(Y_t - Y_t^{\delta=0.1}\right)$

would have been had nicotine vaping odds been 90% lower in the Waves 2 and 3, respectively.

By Wave 5, observed risk was 10.1 percentage points (95%CI: 5.2:15.1) higher than it would

have been had nicotine vaping odds been 90% lower in the Wave 4.

3.4.8 Potential effect modification by age

Cannabis initiation

Age comparisons of incremental propensity score effects on cannabis initiation (i.e., observed risk compared with risk if $\delta = 0.1$ to $\delta = 0.8$) within each follow-up wave are shown in Figure 3.5 (left- and right-most points, respectively). By Wave 2, no differences in cannabis initiation risk by age appeared (Figure 3.5). By Wave 3, the risk difference comparing observed with 80%[1] lower nicotine vaping odds appeared 2.5 percentage points (95%CI: 1.0:4.1) higher for 18-year-olds than for 14-year-olds. By Wave 4, no differences appeared in cannabis initiation risk by age (Figure 3.5). By Wave 5, effects comparing observed with 80% lower nicotine vaping odds ($\delta = 0.2$) appeared 1.6 percentage points (95%CI: 0.01:3.0) higher for 18-year-olds than for 21-year-olds.

Other substance initiation

Age comparisons of incremental propensity score effects on other substance initiation (i.e., observed risk compared with risk if $\delta = 0.1$ to $\delta = 0.8$) within each follow-up wave are shown in Figure 3.6 (left- and right-most points, respectively). There appeared to be no effect modification by age in any wave (Figure 3.6).

[1] $\delta = 0.2$ exposure odds was the minimum δ with which I compared observed risks because comparisons with $\delta = 0.1$ were unstable.

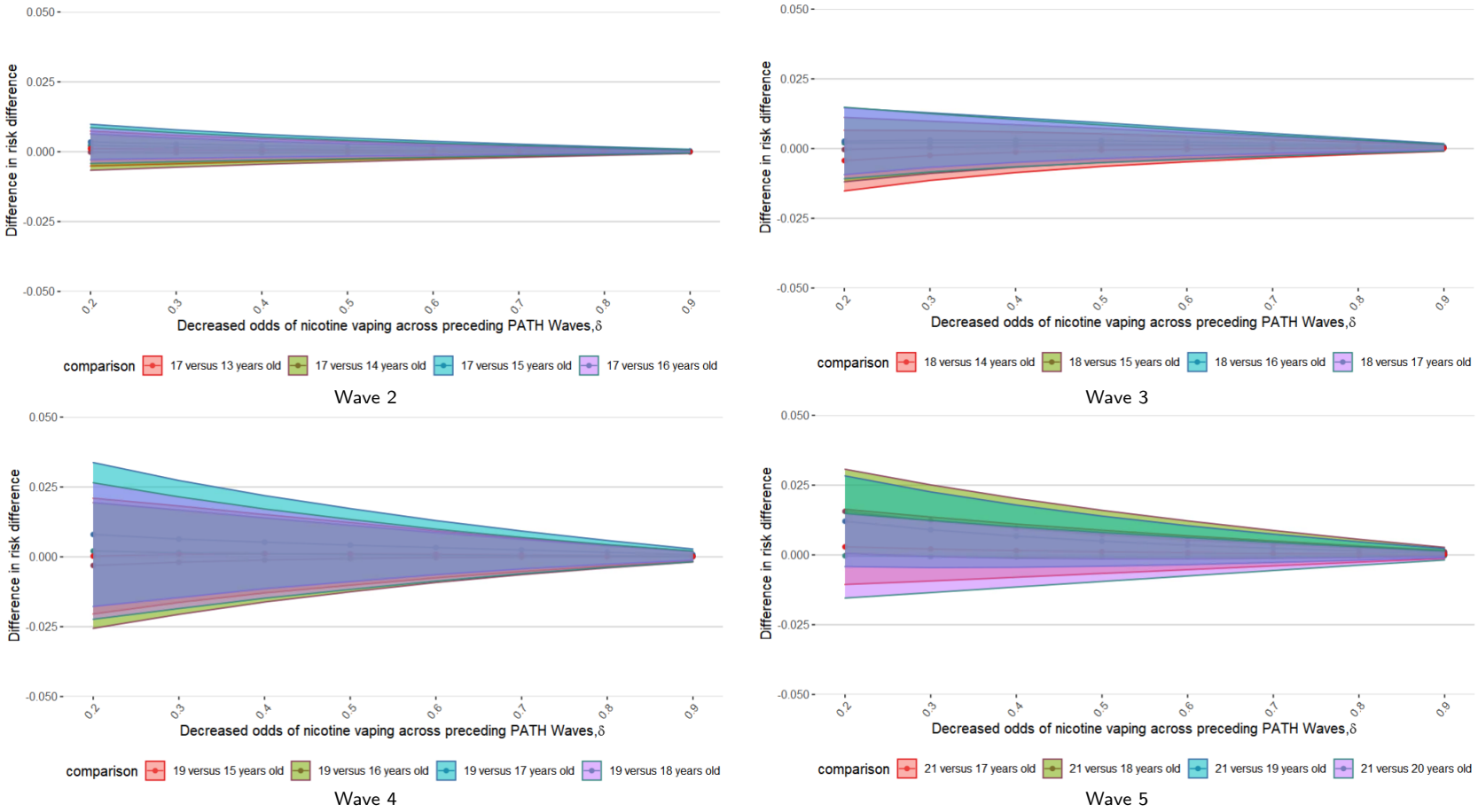

Figure 3.5: Age differences in risk differences comparing observed past-year cannabis initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 80% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.2$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old

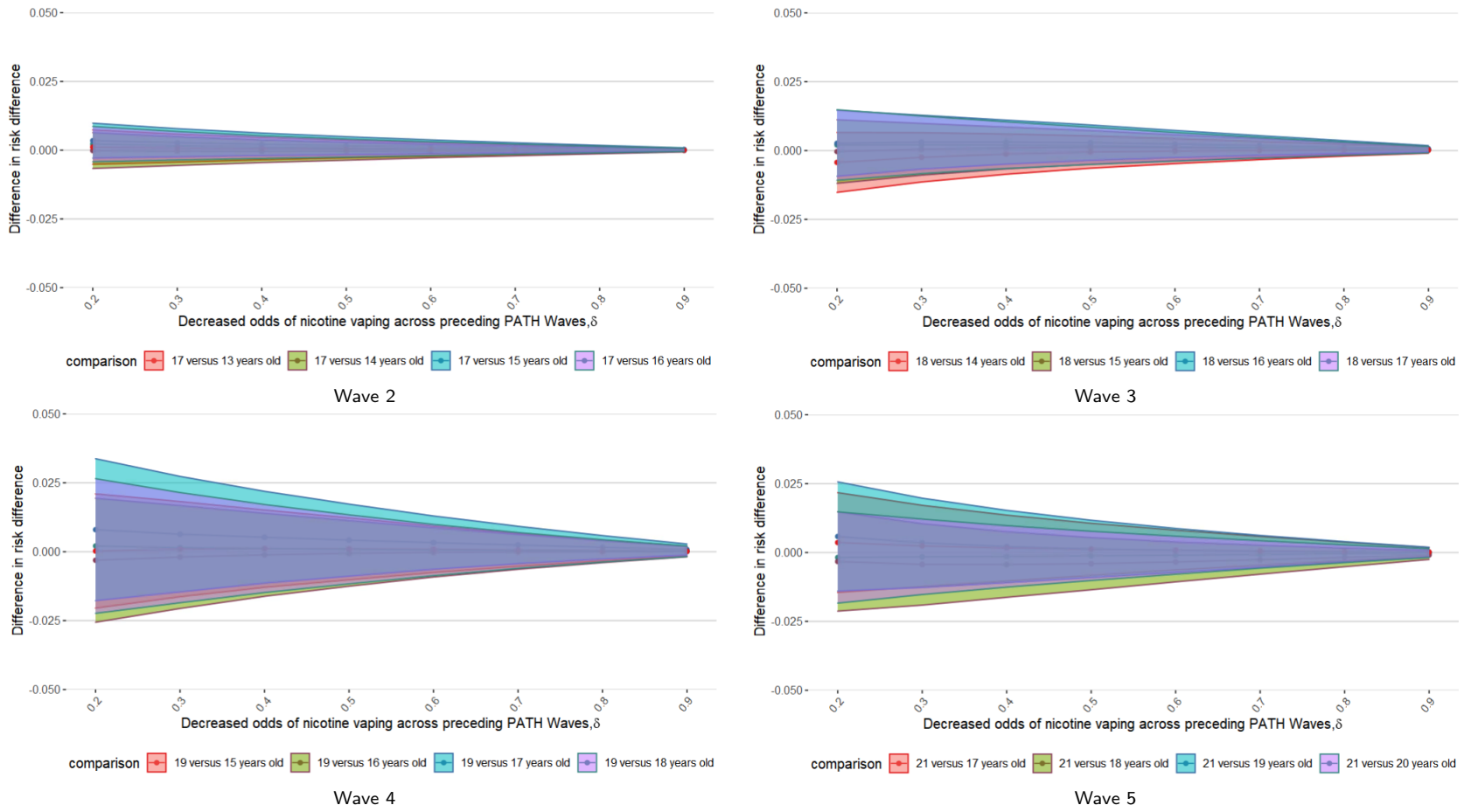

Figure 3.6: Age differences in risk differences comparing observed past-year other substance initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 80% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.2$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old

3.4.9 Potential effect modification by time

Cannabis initiation

Stratifying by age, some incremental propensity score effects appeared larger comparing later waves with earlier waves. Effects comparing observed with 80% lower nicotine vaping odds appeared 1.9 (95%CI: 1.0:2.7), 2.6 (95%CI: 1.7:3.4), and 1.8 (95%CI: 0.4:3.1) percentage points higher for Wave 3 15-, 16-, and 17-year-olds, respectively, than for their Wave 2 counterparts ($\delta = 0.2$, left-most points in all panels of Figure 3.7). An even smaller difference appeared comparing Wave 3 14-year-olds with their Wave 2 counterparts (0.65 (95%CI: 0.05:1.3)). Effects comparing observed with 80% lower nicotine vaping odds appeared 2.6 (95%CI: 1.8:3.4), 3.6 (95%CI: 2.6:4.5), and 4.9 (95%CI: 3.6:6.1) percentage points higher for Wave 4 15-, 16-, and 17-year-olds, respectively, than for their Wave 2 counterparts. Effects comparing observed with 80% lower nicotine vaping odds appeared 3.2 percentage points (95%CI: 1.9:4.4) higher for Wave 5 17-year-olds than for their Wave 2 counterparts.

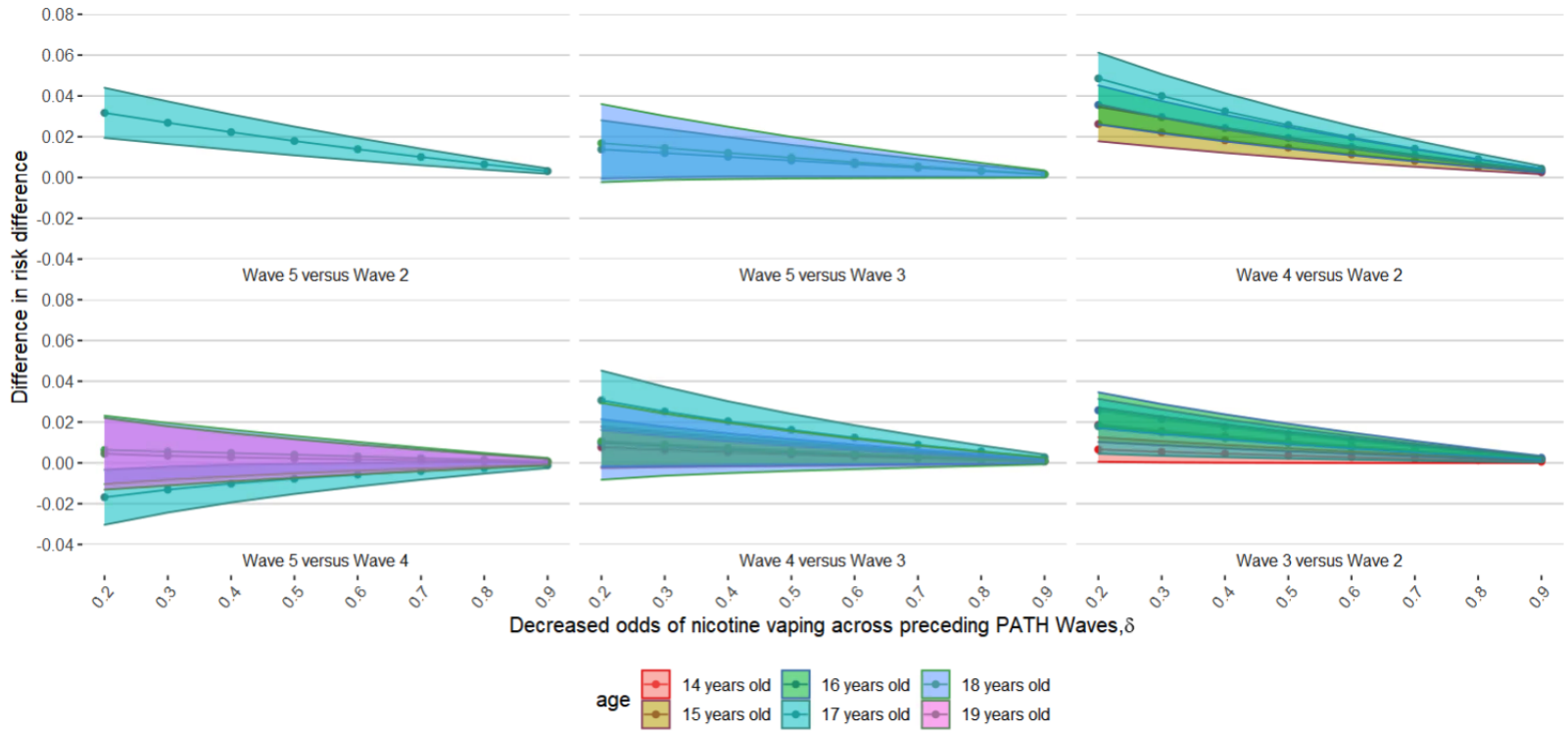

Figure 3.7: Wave differences in risk differences comparing observed past-year cannabis initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 80% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.2$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old

Other substance initiation

Stratifying by age, some incremental propensity score effects on other substance initiation appeared larger comparing later waves with earlier waves. Effects comparing observed with 80% lower nicotine vaping odds appeared 1.0 (95%CI: 0.1:1.9), 1.4 (95%CI: 0.5:2.3), and 1.6 (95%CI: 0.5:2.7) percentage points higher for Wave 3 15-, 16-, 17-year-olds, respectively than for their Wave 2 counterparts ($\delta = 0.2$, left-most points in all panels of Figure 3.8). A smaller difference appeared among 14-year-olds. Effects comparing observed with 80% lower nicotine vaping odds appeared 2.0 (95%CI: 0.9:3.1) and 2.3 (95%CI: 0.4:4.3) percentage points higher for Wave 4 15- and 17-year-olds than for their Wave 2 counterparts. Effects comparing observed with 80% lower nicotine vaping odds appeared 2.7 (95%CI: 1.2:4.1) percentage points higher for Wave 5 17-year-olds than for their Wave 2 counterparts.

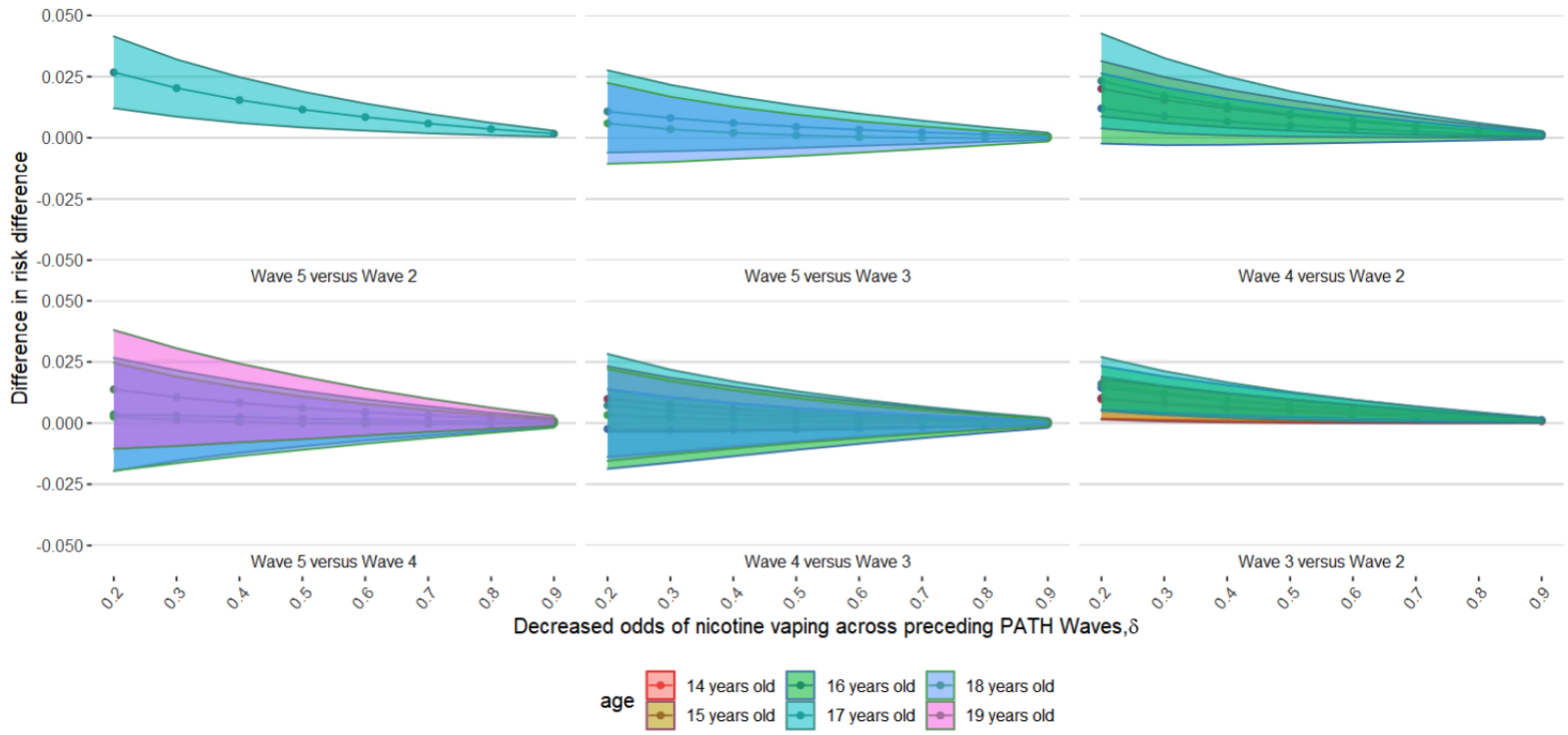

Figure 3.8: Wave differences in risk differences comparing observed past-year other substance initiation risks by each follow-up wave with those risks had nicotine vaping odds been 10% to 80% lower in the prior wave ($\delta = 0.9$ to $\delta = 0.2$, respectively), controlling for baseline confounding (effects at all waves) and baseline and time-varying confounding (effects by Waves 3 through 5), N=9,571 baseline participants 12 to 16 years old

3.4.10 Potential effect modification by tobacco product co-use

Cannabis initiation

The incremental propensity score effect on cannabis initiation by Wave 5 appeared larger comparing history of non-vaping tobacco product use with no history of non-vaping tobacco product use. Among individuals with a history of tobacco product use, the observed Wave 5 risk was 6.1 (95%CI: 4.6:7.6) percentage points greater than the risk had nicotine vaping odds been 80% lower (60.2% versus 54.1%, left panel of Figure 3.9). Among individuals with no history of tobacco product use, the observed Wave 5 risk was 4.1 (95%CI: 3.5:4.6) percentage points greater than the risk had nicotine vaping odds been 80% lower (31.9% versus 28.4%, left panel of Figure 3.9).

Other substance initiation

The incremental propensity score effect on other substance initiation by Wave 5 appeared larger comparing history of non-vaping tobacco product use with no history of non-vaping tobacco product use. Among individuals with a history of tobacco product use, the observed Wave 5 risk was 4.8 (95%CI: 3.5:6.2) percentage points greater than the risk had nicotine vaping odds been 80% lower (32.2% versus 27.4%, right panel of Figure 3.9). Among individuals with no history of tobacco product use, the observed Wave 5 risk was 1.7 (95%CI: 1.3:2.1) percentage points greater than the risk had nicotine vaping odds been 80% lower (22.4% versus 21.0%, right panel of Figure 3.9).

3.4.11 Sensitivity analyses

Despite efforts to control for measured confounding, unmeasured confounding could have explained away many incremental propensity score effects. Particularly, Wave 2 E-values (assessed on the risk ratio scale) for the largest exposure comparison (observed risk versus risk if $\delta = 0.1$) were 1.4 and 1.2 for initiation of cannabis and other substances, respectively

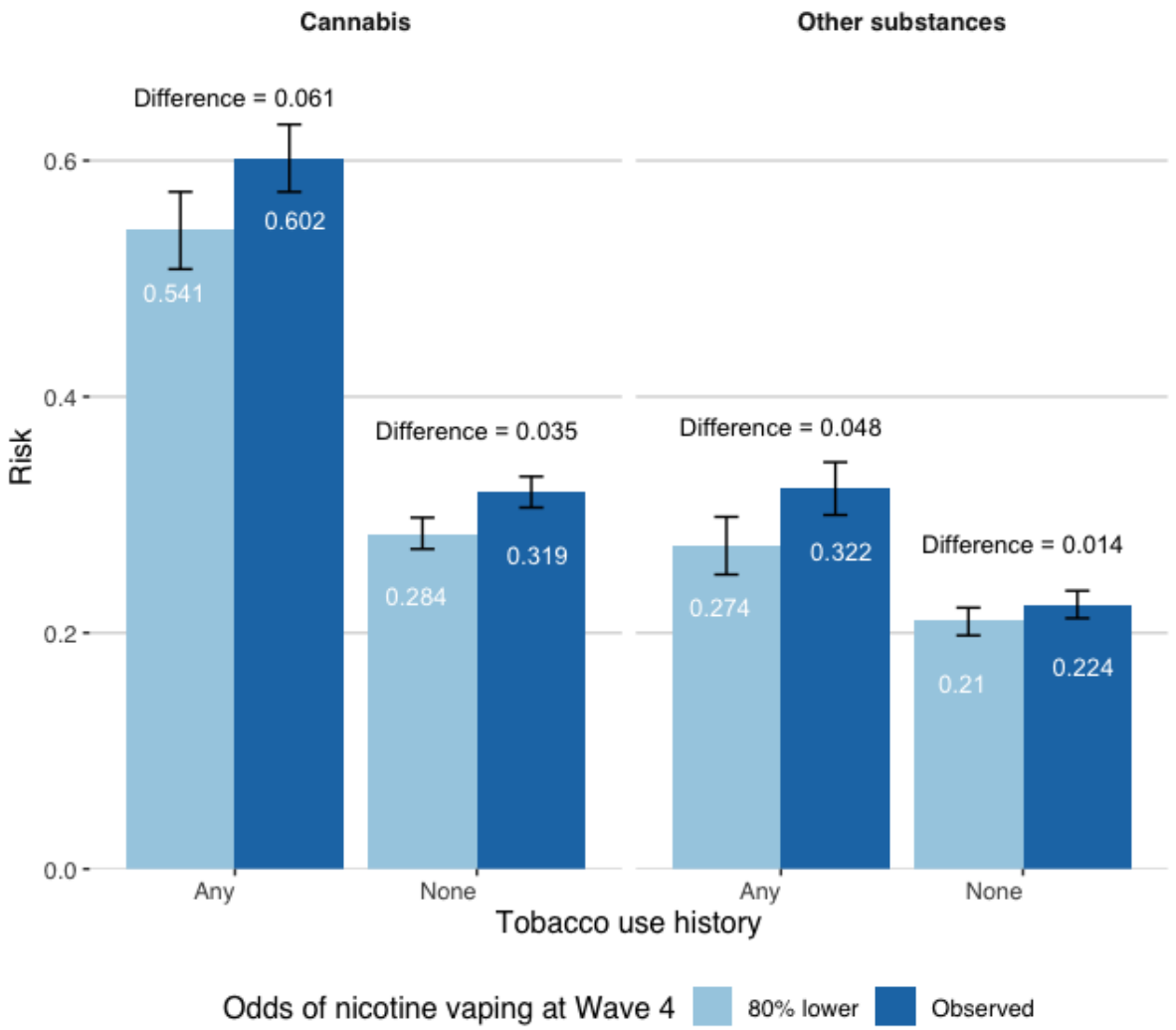

Figure 3.9: Observed Wave 5 risks and 95% confidence intervals of past-year observed cannabis initiation and past-year observed other substance initiation and those risks had nicotine vaping odds been 80% lower in Wave 4 and differences in those risks, all stratified by history of non-vaping tobacco product use at Waves 1 through 4, and controlling for baseline and time-varying confounding, N=9,571 baseline participants 12 to 16 years old

(Appendix Figure B.2). Estimates of lower CI bounds were ≤ 1 for Wave 2, so their E-values for those estimates were uninformative. By Wave 3, the smallest and largest E-values for effects on cannabis initiation were 1.15 and 1.92, respectively, while they were 1.09 and 1.59 for effects on other substance initiation. Lower CI bounds for effects on cannabis and other substance initiation were ≤ 1 when comparing observed risks with risks had nicotine vaping odds been 10% to 30% lower and 10% to 60% lower than observed, respectively, so E-values

were uninformative. E-values for lower CI bounds of effects on cannabis initiation ranged from 1.13 to 1.76 when comparing observed risks with risks had nicotine vaping odds been 40% to 90% lower than observed, respectively. E-values for lower CI bounds of effects on other substance initiation ranged from 1.14 to 1.41 when comparing observed risks with risks had nicotine vaping odds been 70% to 90% lower than observed, respectively. By Wave 4, incremental effects appeared even more robust to unmeasured confounding. E-values for the smallest and largest exposure comparisons for effects on cannabis initiation ranged from 1.91 to 2.28; E-values for lower CI bounds ranged from 1.56 to 2.16. For effects on other substance initiation, E-values ranged from 1.11 to 1.96, and E-values for lower CI bounds ranged from 1.00 to 1.82. E-values and their lower CI bounds at Waves 4 and 5 indicated an unlikely possibility that unmeasured confounding could alternatively explain effect estimates for the largest exposure comparisons.

3.5 Discussion

In this nationally representative study, I examined whether evidence supports effects of nicotine vaping on the initiation of cannabis and the initiation of other substances during a lengthy period of increasing nicotine vaping popularity among youth aging into young adulthood. Four observations emerged from my analysis. First, the cumulative risks of cannabis and other substance initiation were higher in the study's final wave (2018-19) than they would have been had nicotine vaping odds been 10% to 90% lower in the prior wave (2016-18). Second, these effects did not appear to differ by age, but third, they did appear to differ by time, both overall and within each age category. Last, most incremental effects (not effect modification estimates) appeared robust to potential unmeasured confounding starting at Wave 3 (2015-16).

There were more cases of cannabis and other substance initiation over time than there would have been had nicotine vaping's prevalence been up to 90% lower. My results are compatible

with others' findings that nicotine vaping affects initiation because the effects of shifting the odds of exposure on the outcomes depend on the exposure affecting the outcomes net of any biases.[10,26,70] I also expanded upon others' findings. My study period was three years longer than the next longest study, and it spanned the transitional period from adolescence into young adulthood for most participants. Additionally, I comprehensively controlled for time-invariant and time-varying confounders specified in substance use theory and prior studies, which likely affected nicotine vaping and the outcomes during the transitional period (e.g., living arrangements, academic performance).[32,46–62] Together, these findings are evidence that nicotine vaping may have long-term effects on cannabis and other substance initiation as adolescents age into adulthood. They are also evidence that a population strategy towards lowering nicotine vaping prevalence may have limited the number of new cannabis and other substance initiates.

I found no evidence for effect modification by age. Dai *et al.*, the only other study to examine possible effect modification by age, found that nicotine vaping multiplied cannabis initiation risk more for younger than older adolescents from baseline to Wave 2 in PATH. Their conclusion that nicotine vaping's effects differ by age may be incorrect because they likely violated the positivity assumption.[111] Dai *et al.* and my seemingly conflicting results do not mean that nicotine vaping's effects on cannabis initiation do not differ by adolescent age or age group. Rather, claims of effect modification by age currently lack support. Though I did not test the presence of cohort effects, the absence of age effects suggest the absence of cohort effects.

Positive incremental effects of shifting nicotine vaping's prevalence strengthened over time, which indicate that nicotine vaping's effects on the outcomes remained positive over the study period. If nicotine vaping's effects strengthened as the incremental effects did, a change in the prevalence of effect modifiers could be responsible. Internalizing symptoms are potential modifiers of nicotine vaping's effects because they are associated with (and possibly cause)

cannabis and other substance use,[72,123–125] and instances of internalizing symptoms (i.e., low self-esteem and high self-derogation) increased during the study period.[126] For internalizing symptoms to longitudinally modify nicotine vaping's effects in PATH, which I did not analyze, they must have increased in prevalence over time, and they must have affected the outcomes. Formally identifying potential effect modifiers that increase outcome risks and intervening to reduce their prevalence may help reduce the effect of nicotine vaping.

One effect modifier that poses a threat to young people's health is tobacco use. My results show that incremental propensity score effects of shifting nicotine vaping odds to be lower appear stronger among people who did versus did not have a history of using non-vaping tobacco products. These findings appear to contrast with prior research, which found that vaping had a larger effect on cannabis use among youth who did not smoke or consume tobacco products.[11,104] Their results are likely due to a high baseline risk of cannabis initiation in the tobacco use subgroups. That is, people who use tobacco products are very likely to use cannabis regardless of whether they vape, and vaping does not add much risk.[11] My findings appear to contrast with the prior studies because of my analytical approach. I modeled propensity scores of vaping, which were very small overall and were generally smaller among individuals who did not use tobacco compared with those who did. Consequently, changes in propensity scores were likely smaller in the tobacco non-use group such that the number of outcome cases would not have been too different compared with the tobacco use group. In other words, had the prevalence of nicotine vaping been lower, fewer outcomes would have occurred among individuals who used non-vaping tobacco products than among tobacco-naïve individuals, even though vaping has a stronger effect on the latter.

My study encountered some limitations. First, misclassification due to social desirability was likely because participants self-reported sensitive topics. In this study, social desirability can cause correlation between reasons for misreporting different substance use, such as stigma

and illegality of substance use for individuals under a particular age;[86] however, PATH used Audio Computer-Assisted Self-Interviewing and Computer-Assisted Personal Interviewing, and participants were assured that responses were confidential to avoid socially desirable responses. A second limitation is that unmeasured confounders, such as friends' substance use, parents' smoking during prenatal period or during childhood or adolescence, parental supervision, and neighborhood risks (e.g., density of vape shops), likely led to positive bias. Direct measures of these constructs were absent; however, I comprehensively controlled for indirect measures in my models (e.g., someone else uses tobacco in the home, house rules about tobacco in the home, would use e-cigarettes if a best friend offered, noticed e-cigarette advertising). In addition, E-values were greater than 2 for most effects starting at Wave 3, which indicated that unmeasured confounding was a possible but unlikely alternative explanation of effect estimates. Third, about 27% of individuals were missing exposure and outcome data at Waves 4 and 5, respectively; to avoid potential selection bias, I estimated missing data using multiple imputation by chained equations under the missing at random assumption, which was possibly inadequate.

In summary, I found evidence consistent with nicotine vaping's positive effects on cannabis and other substance initiation among adolescents as they aged into adulthood from 2013 to 2019. Public health interventions designed to decrease nicotine vaping prevalence include an increase in the age to legally purchase nicotine products,[127] bans on nicotine vaping flavors aimed at youth consumers,[128] the FDA ban of JUUL products,[129] and vaping restrictions where other tobacco products are not allowed. A comprehensive strategy of limiting cannabis and other substance initiates may include these interventions and other levers that blunt the effects of nicotine vaping on cannabis and other substance initiation; these levers may be identified through further investigation of potential effect modifiers and mediators.

Chapter 4: Do conduct problems mediate the effects of nicotine vaping on cannabis and other substance initiation? Results from Wave 1 (2013-14) to Wave 5 (2018-19) of the Population Assessment of Tobacco and Health Study

4.1 Abstract

Background: Previous studies estimated longitudinal effects between nicotine vaping and subsequent initiation of cannabis and other substances (e.g., cocaine, heroin) up to six years later; however, none evaluated mechanisms of these effects, which could be potential intervention targets. I examined whether there was evidence to support indirect effects of nicotine vaping on the initiation of cannabis and the initiation of other substances through conduct problems over a six-year period.

Methods: Youth aged 12 to 16 years old were drawn from Wave 1 (baseline, 2013-14) of the Population Assessment of Tobacco and Health Study (hereon, PATH) if they never had conduct problems by Wave 3 (2015-16) and never used cannabis (N = 2,949) or other substances, i.e., methamphetamine, speed, heroin, inhalants, solvents, hallucinogens, cocaine, crack, or used prescription substances nonmedically by Wave 4 (N = 2,995). I estimated stochastic interventional indirect effects, which were effects of various shifts in the propensity of nicotine vaping at Waves 1, 2, or 3 (10% to 90% lower odds) on the risks of initiating cannabis and initiating other substances at Wave 5 (2018-19) through potential shifts in the conditional joint distribution of three conduct problems at Wave 4 (2016-2018)—lying or conning to get something, bullying perpetration, and starting fights. I adjusted for baseline and intermediate confounding in all analyses based on substance use theories and prior research.

Results: Estimates of stochastic interventional indirect effects on both outcomes were null (risk difference = 0). Null results could be due to the low prevalence of some conduct problems (< 5%) resulting in inadequate power. Another possible explanation is the absence of associations between nicotine vaping and lying or between bullying and cannabis initiation, which suggest a low probability of mediation when considering the three posited mediators together.

Conclusions: I did not find evidence that three conduct problems considered together mediated nicotine vaping's effects on cannabis initiation or initiation of other substances. An intervention target for blunting nicotine vaping's negative effects after it has started remains unknown.

4.2 Introduction

Evidence that nicotine vaping increases the risks of initiating cannabis and other substances, such as cocaine, heroin, and methamphetamine, continues to accumulate.[8,10,72] Recent research shows that nicotine vaping's effects on these outcomes can last up to six years as adolescents become young adults.[130] Despite growing evidence about the effects, we have limited understanding about how they occur. Investigation of factors that mediate the effects may help test the validity of the total effects, specify potential intervention targets aimed at preventing nicotine vaping's effects on initiating cannabis and other substances, and add to the body of evidence upon which other researchers can build.

A few mechanisms of the effects have been posited. One is that familiarity with nicotine vaping devices leads to cannabis initiation because cannabis can also be vaped.[5,26,28,73] Another is that nicotine increases central nervous system sensitization to other drugs as well as pleasure and feelings of reward, which lead to use of those drugs.[8–10,26,27] A third hypothesis is that deviant peer affiliation mediates the effects because deviant peers may access and use cannabis and other substances.[26,28,69,72] This third hypothesis has some theoretical support. Positive effects of nicotine vaping on cannabis and other substance initiation through deviant peer affiliation would be consistent with differential association theory and social learning theory. Individuals learn antisocial values and deviant behaviors (e.g., substance use, conduct problems) through time spent in antisocial activities with peers.[131,132] When youth are rewarded from participating in antisocial activities with their peers through increased acceptance and/or loyalty, they are likely to continue associating and engaging in the antisocial activities with those peers.[133] The reverse process may also occur; behaviors are discouraged when peers do not reward them.[133] Differential association theory and social learning theory may predict that vaping, which may be considered a mild deviant behavior (as it is mostly experimental),[134] may increase the risk of affiliating with deviant peers. The theories could also predict that

continued deviant peer affiliation leads to cannabis and other substance initiation because deviant youth have access to these substances and can influence their friends to try them. In short, differential association theory and social learning theory provide the basis for deviant peer affiliation as a mechanism though which nicotine vaping's effects on deviant behaviors, including substance use, are promoted.

The deviant peer affiliation mediation hypothesis also has some empirical support. For instance, a nationwide study found that youth were at over two times the risk of cannabis vaping initiation if friends vaped nicotine,[72] which would suggest mediation via deviant peer affiliation if the youth already vaped nicotine. Other studies showed that social environments favorable to nicotine vaping, such as exposure to deviant peer behaviors (e.g., friends' substance use), deviant peer affiliation, and and positive attitudes towards nicotine vaping, increased adolescents' risk for substance use.[135–140] Despite empirical support for mediation through deviant peer affiliation, it has not been empirically tested. This may be because longitudinal data on youth's peers, their behaviors, and closeness of youths to their peers are typically unavailable. For instance, the Population Assessment of Tobacco and Health only questions whether youth would use tobacco products if a best friend offered. Monitoring the Future asks adolescents about their friends' drug use, but it only follows participants longitudinally starting at age 18, making mediation studies during adolescence when youth are most influenced by their friends challenging.[141,142]

If appropriate measures of deviant peer behaviors or affiliation are unavailable, alternative closely related mechanisms could be conduct problems because nicotine vaping predicts deviant peer affiliation, which predicts an array of deviant behaviors, including substance use and conduct problems.[143–145] Youth who vape nicotine are more likely than youth who do not vape nicotine to have friends who are truant, use cannabis, get drunk, and get bad grades in school.[146–151] They are also more likely to have respect for peer deviance and participate

in activities that may promote deviant acts with peers (e.g., staying out late with friends and no supervision, truancy).[146–151] Additionally, youth who vape and have friends who smoke cigarettes are more frequently delinquent in the past 12 months (e.g., engaged in theft, violence, selling drugs) than youth who vape and have no peers who smoke,[152] suggesting that youth who try vaping may fall into a crowd of deviant peers and consequently commit deviant acts (see conceptual model in Appendix Figure B.3). In other words, nicotine vaping may increase the risk of deviant peer affiliation, which in turn may increase the risk of conduct problems; therefore, conduct problems could be downstream mediators of deviant peer affiliation, the posited mediator.

If conduct problems are downstream mediators of nicotine vaping's effects on the initiation of cannabis and other substances, they must be associated with prior nicotine vaping and the subsequent outcomes. Nicotine vaping and conduct problems are associated, but the evidence is limited to cross-sectional studies. For instance, youth who vape nicotine are more likely than youth who do not vape nicotine to destroy property, lie, fight, sneak out of home, steal, be loud and belligerent, bully others and a have proclivity towards sensation-seeking and impulsivity.[151,153] Unlike associations between nicotine vaping and conduct problems, which appear temporally ambiguous, conduct problems consistently increase the risks for cannabis initiation and the initiation of other substances.[71,154,155] Initiating cannabis and other substances are deviant behaviors typical of individuals with conduct problems and requires planning of how to acquire and where to use the substances, which can mostly be done in groups of individuals with similar norm-breaking tendencies (Appendix Figure B.3).[156–158] In short, conduct problems satisfy the prerequisites for being mediators because of their associations with nicotine vaping, cannabis initiation, and other substance initiation.

In this study, I examined whether nicotine vaping was associated with three subsequent new onset conduct problems and whether the conduct problems were associated with subsequent

initiation of cannabis and other substances. I also formally examined whether new onset conduct problems mediated the effects of nicotine vaping on the initiation of cannabis and other substances from adolescence into adulthood.

4.3 Methods

4.3.1 Population Assessment of Tobacco and Health (PATH)

Data come from Wave 1 (baseline, 2013-14) to Wave 5 (2018-19) of the ongoing PATH Study (hereon, PATH).[108] PATH was developed following the 2009 Family Smoking Prevention and Tobacco Control Act and is a collaboration between the National Institute on Drug Abuse at the National Institutes of Health, and the Center for Tobacco Products at the Food and Drug Administration.[108] PATH is a nationally representative cohort study of adolescents and adults in the United States with the general purpose of evaluating tobacco use and tobacco-related health problems (e.g., respiratory, dental).[108] PATH's youth participants complete an annual survey about themselves and a parent or guardian is asked to complete a survey mostly about their child. I was granted access to the restricted PATH data files to conduct the current study,[109] and my protocol was approved by the Columbia University Irving Medical Center Institutional Review Board.

4.3.2 Sampling design

PATH's sampling strategy was a four-stage stratified area probability sample.[108,109] A county or group of contiguous counties were selected as primary sampling units in the first stage based on a stratified random sample, which oversampled communities with high proportions of young adults, Black or African American adults, and people who use tobacco. In the second stage, primary sampling units were further stratified based on these demographic and behavioral characteristics into smaller geographic segments. At the third stage, residential addresses were sampled within segments. At the fourth stage, residents at those addresses were sampled. One adult residing in each household completed a household primary screener. Based on that

screener and predetermined sampling rates of key subgroups (e.g., young adults, tobacco users, etc.), the screened adult and/or another adult were sampled to participate in a second screener and possibly a full interview. At most, two adults could be selected from each household for participation in the second screener. No more than two youths (aged 12 to 17 years old) were randomly sampled from each household in which an adult (parent or guardian) was sampled.[108] In other words, youth sampling required adult sampling from the same household, while adult sampling did not necessitate youth sampling.

4.3.3 Data collection

Data were collected annually (between 9 and 14 months) over four waves (Appendix Table B.1): Wave 1 (2013-14), Wave 2 (2014-15), Wave 3 (2015-16), and Wave 4 (2016-18).[108] Wave 5 (2018-19) data were collected about two years after Wave 4 data. Youth and adults completed specific instruments about themselves, including demographics, health, and substance use behaviors. Additionally, parents or guardians of youths completed an instrument mostly about their child. Youth who reached 18 years old completed an adult interview.

4.3.4 Study populations

This study included two analytic cohorts consisting of participants who were aged 12 to 16 years old at baseline. They were the

1. cannabis initiation cohort: 2,949 baseline participants who never had conduct problems through Wave 3 and who never used cannabis in any form through Wave 4.

2. other substance initiation cohort: 2,995 baseline participants who never had conduct problems through Wave 3 and who never used other substances (defined in section 4.3.7) through Wave 4.

Baseline 17-year-olds were excluded from the study population because most would have been adults (18 years old) in Wave 2 and past-year nicotine vaping was not measured among

adults in that wave.

4.3.5 Estimand

My research question was “Do conduct problems mediate nicotine vaping’s effects on cannabis initiation and initiation of other substances from adolescence to adulthood?"

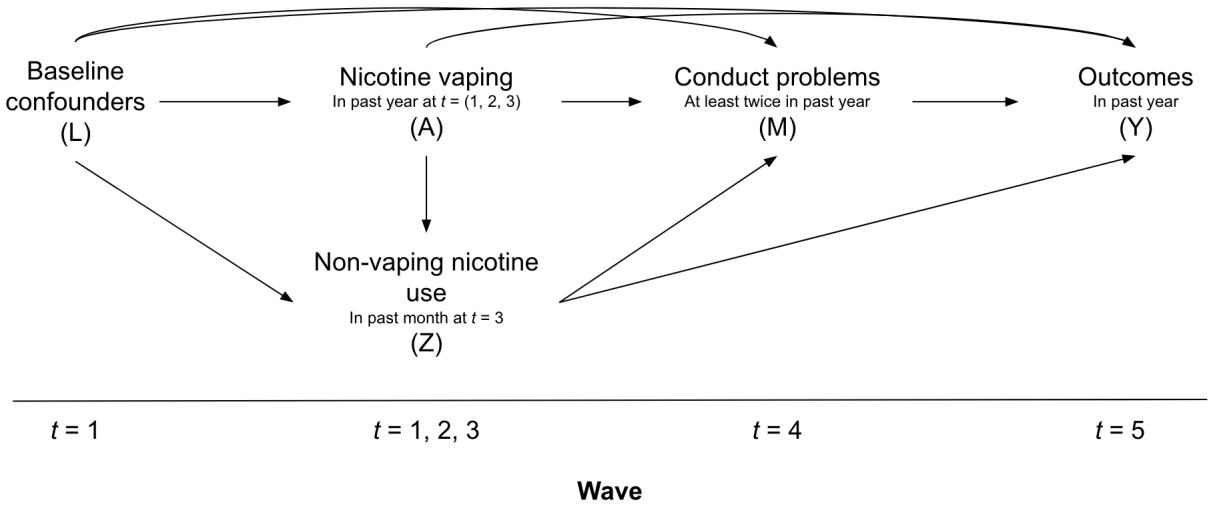

Figure 4.1: Causal diagram depicting assumed causal structure, where L is a set of baseline confounders, A is nicotine vaping at Waves 1, 2, or 3, Z is a non-vaping tobacco use at Wave 3, M is a set of mediators at Wave 4 (lying or conning to get something, bullying perpetration, starting fights), and Y is observed cannabis initiation or other substance initiation at Wave 5

The diagram depicted in Figure 4.1 depicts the assumed causal structure for answering this question. In the diagram, L is a set of baseline confounders, A is nicotine vaping at Waves 1, 2, or 3, Z is a non-vaping tobacco use at Wave 3, M is a set of mediators at Wave 4 (lying or conning to get something, bullying perpetration, starting fights), and Y is observed cannabis initiation or other substance initiation at Wave 5. The natural indirect effect is one estimand that can answer the research question and can be written as $\mathbb{E}(Y^{a^{'},M^{a^{'}}} - Y^{a^{'},M^{a^{*}}})$, where $Y^{a^{'}}$ are potential outcomes if everyone had exposure $a^{'}$, and $M^{a^{'}}$ and $M^{a^{*}}$ are potential mediators if everyone had exposure $a^{'}$ or a^{*}, respectively. Therefore, $Y^{a^{'},M^{a^{'}}}$ are potential outcomes if everyone was exposed to $a^{'}$ and had the mediator value they would have had if exposed to $a^{'}$, while $Y^{a^{'},M^{a^{*}}}$ are potential outcomes if everyone was exposed to $a^{'}$ and had the mediator

value they would have had if exposed to a^*. The natural indirect effect is therefore the effect of fixing the exposure at $a^{'}$ while contrasting $M^{a^{'}}$ and M^{a^*} (i.e., potential mediator values had different exposures happened). The natural indirect effect has an identifying assumption, $M^{a^{'}}$ is independent of $Y^{a^{'},M^{a^*}}$; however, the effect cannot be identified because of Z, an intermediate confounder (i.e., a common cause of M and Y caused by A), which creates a backdoor pathway between these counterfactuals.[39] Consequently, it was not chosen as this study's estimand.

Another estimand that could answer the research question and is identifiable in the presence of an intermediate confounder is the interventional indirect effect of a static exposure, which can be written as $\mathbb{E}(Y^{a^{'},G^{a^{'}}} - Y^{a^{'},G^{a^*}})$, where $Y^{a^{'}}$ are potential outcomes if everyone had exposure $a^{'}$, and $G^{a^{'}}$ and G^{a^*} are random draws from the conditional counterfactual joint distributions of mediators if everyone had exposure $a^{'}$ and a^*, respectively, conditional on L. Therefore, $Y^{a^{'},G^{a^{'}}}$ are potential outcomes if everyone was exposed to $a^{'}$ and had mediator values been stochastically drawn from the distribution of mediators that would have been observed if everyone was exposed to $a^{'}$, while $Y^{a^{'},G^{a^*}}$ are potential outcomes if everyone was exposed to $a^{'}$ and had mediator values been stochastically drawn from the conditional counterfactual distribution of mediators that would have been observed if everyone was exposed to a^*, conditional on L. It is the effect of fixing the exposure at $a^{'}$ while setting the mediator m to a random draw $G^{a^{'}}$ from individuals who have exposure $a^{'}$ versus a random draw G^{a^*} from individuals who have exposure a^*. The estimand is identified in the presence of intermediate confounding because the random draw $G^{a^{'}}$ breaks the dependence between $M^{a^{'}}$ and $Y^{a^{'},M^{a^*}}$. Identifying the estimand requires the verifiable assumption of exposure positivity, the nonzero probability of being exposed, conditional on confounders (i.e., nonzero propensity scores). Exposure positivity fails if propensity scores $\pi(l)$ are 0 or 1 because the outcome model would be conditioned on zero-probability events $(a^{'}|l)$ or $(a^*|l)$.[159] Near positivity violations (when propensity scores are near 0 or 1), also known as practical positivity

violations, may result in finite sample bias and incorrect inference..[110] I checked the positivity assumption and found near violations as shown in Appendix Figure B.4. For example, 20% of participants had propensity scores lower than 0.025 in the cohort where cannabis initiation was the outcome. Therefore, the static exposure interventional indirect effect was not chosen as this study's estimand.

Approaches to restore positivity include reducing the number of sparse strata by reducing the number of confounders, truncating propensity scores, and trimming the sample to areas of overlap between propensity scores for exposed and unexposed groups;[110] however, restoring positivity with any of these approaches would compromise valid inference in the target population (all baseline youth aged 12 to 16 years old). Not controlling for confounders could bias effects. Bias from truncating propensity scores is usually tolerable if it is done for a few observations; however, 20% of propensity scores would need to be truncated in my sample, which would result in model misspecification. Finally, I would remove 20% of my sample if I were to trim propensity scores, and the resulting sample would no longer represent the target population.

Another approach is to estimate a parameter that is identifiable without assuming positivity and which would still address my research question. The stochastic interventional indirect effect shown in Expression 4.1, which is effectively an indirect incremental propensity score intervention effect, is one such estimand.[160] It can be identified in the presence of intermediate confounding as it is the stochastic analog to the static interventional indirect effect. The stochastic interventional indirect effect can be expressed as the effect of a shift in the odds of nicotine vaping on cannabis initiation or the initiation of other substances had the shift in nicotine vaping odds changed only the conditional joint distribution of conduct problems (hereon, conditional distribution).[160]

$$\mathbb{E}\left(Y^{A_\delta,G} - Y^{A_\delta,G_\delta}\right) \tag{4.1}$$

In Expression 4.1, δ is a ratio of the shifted to unshifted odds of nicotine vaping $\left(\frac{\frac{q(l)}{1-q(l)}}{\frac{\pi(l)}{1-\pi(l)}}\right)$ where $q(l)$ and $\pi(l)$ are the shifted and unshifted propensity scores, respectively. In this study, $\delta < 1$, meaning $q(l) < \pi(l)$. Applying similar notation from the static interventional indirect effect described earlier, Y^{A_δ} in Expression 4.1 are potential cannabis and other substance initiation outcomes that would have been observed had everyone's nicotine vaping odds been lower, G is a random draw from the distribution of conduct problems M_A that would have been observed had everyone's nicotine vaping odds not changed, conditional on L (i.e., the observed conditional distribution), and G_δ is a random draw from the distribution of conduct problems M^{A_δ} that would have been observed had nicotine vaping odds been lower, conditional on L; therefore, $Y^{A_\delta,G}$ and Y^{A_δ,G_δ} are potential outcomes that would have been observed had nicotine vaping odds been lower and had values of conduct problems been stochastically drawn from the observed and counterfactual conditional distributions of conduct problems, respectively.

To explain guaranteed positivity, I rewrote components of Expression 4.1 below. Specifically, if identified, $\mathbb{E}(Y^{A_\delta,G})$ can be written as

$$\mathbb{E}\left(\hat{Y}|G = m, A_\delta = a, \mathbf{L}\right) \tag{4.2}$$

and $\mathbb{E}(Y^{A_\delta,G_\delta})$ can be written as

$$\mathbb{E}\left(\hat{Y}|G_\delta = m, A_\delta = a, \mathbf{L}\right) \tag{4.3}$$

where $\hat{Y}$ are predicted outcomes, G and G_δ denote random draws of the mediators m from observed and counterfactual conditional distributions of conduct problems M_A and M^{A_δ}, respectively, A_δ is a simulated version of incremental exposure (random variable indicating

possibly shifted nicotine vaping odds), a is a realization of A_δ, and $\mathbf{L}$ is a vector of baseline confounders. In Expressions 4.2-4.3, A_δ is always defined as $p(l) > 0 \implies (A = A_\delta | l) > 0$ (i.e., having any propensity score). Exposure positivity always holds because Y is conditioned on the always-defined A_δ. For an in-depth discussion of theoretical issues with interventional estimands, see Miles.[161]

4.3.6 Identification assumptions

Stochastic interventional indirect effects still require assumptions despite their desirable properties of identifiability in the presence of intermediate confounding and guaranteed exposure positivity. One of these assumptions is no unmeasured confounding as defined below:

1. potential outcomes are independent of nicotine vaping given baseline confounders,
2. potential outcomes are independent of conduct problems given baseline confounders, nicotine vaping, and tobacco use (not including nicotine vaping),
3. potential mediators are independent of nicotine vaping given baseline confounders.

Mediator positivity (i.e., $\mathrm{p}(m|a,l) > 0$ and $\mathrm{p}(m|a,z,l) > 0$) is also required for identification. I will explain why this is true using the identification result from Hejazi *et al.* for $\mathbb{E}\left(Y^{A_\delta, G}\right)$ in Expression 4.1.:[160]

$$\int \mathbb{E}(Y|m,a,z,l)\mathrm{p}(m|l)\mathrm{P}(z|a,l)\mathrm{g}_\delta(a|l)\mathrm{p}(l)\,d\nu(y,m,a,z,l) \tag{4.4}$$

We can rewrite

$$\mathbb{E}(Y|m,a,z,l) = \frac{\mathrm{YP}(y,m,a,z,l)}{\mathrm{p}(m,a,z,l)} \tag{4.5}$$

and then factorize its denominator so that Expression 4.4 can be rewritten as

$$\int \frac{\mathrm{YP}(y,m,a,z,l)}{\mathrm{p}(m|a,z,l)\mathrm{P}(z|a,l)\mathrm{g}_\delta(a|l)\mathrm{p}(l)} \times \mathrm{p}(m|l)\mathrm{P}(z|a,l)\mathrm{g}_\delta(a|l)\mathrm{p}(l) \tag{4.6}$$

In Expression 4.6, one can see that density of baseline confounders $p(l)$ and conditional probability of the intermediate confounder $P(z|a,l)$ can be 0 only if $l = 0$, and if so, each of these terms would cancel out in the numerator and denominator. As previously stated, $a \neq 0$ in all terms because $p(l) > 0 \implies \mathrm{g}_\delta(a|l) > 0$. In other words, everyone has a shifted propensity score value (even if it were unshifted, i.e., $\pi(l) = 0 \implies q(l) = 0$ or $\pi(l) = 1 \implies q(l) = 1$). The only term in the denominator that would not cancel out if it were 0 is the conditional density of the mediators $\mathrm{p}(m|a,z,l)$ because it is not in the numerator, so it must be positive wherever $\mathrm{p}(m|l) > 0$; therefore, $\mathrm{p}(m|l) > 0 \implies \mathrm{p}(m|a,z,l) > 0$. One can apply the same intuition to see that mediator positivity is required to identify $\mathbb{E}\left(Y^{A_\delta, G_\delta}\right)$.

I empirically evaluated positivity of mediators altogether and separately. Positivity was met when all conduct problems were considered simultaneously or when lying or conning to get something was considered alone (Appendix Figure B.5). Positivity was tenuously met when considering bullying perpetration alone or starting fights alone.

4.3.7 Measures

Outcomes

There were two binary outcomes, each assessed in the past 12 months at Wave 5 among those with no prior use: 1) cannabis initiation, defined as use of THC, grass, pot, or weed, and 2) other substance initiation, defined as nonmedical use of prescription drugs, such as Ritalin, Adderall, painkillers, sedatives, and tranquilizers, or use of methamphetamine, speed, heroin, inhalants, solvents, hallucinogens, cocaine, or crack. Despite alcohol being a psychoactive, addictive, and a prohibited substance for anyone under 21 years old, alcohol initiation was not considered as an outcome. Alcohol initiation usually precedes nicotine vaping initiation, or

both are initiated concurrently,[41–43] which, as others have found,[44] makes estimating nicotine vaping's effects on alcohol initiation challenging.

Exposure

At Wave 1, participants answered whether they used e-cigarettes in the past 12 months. E-cigarettes were described as devices look like regular cigarettes, but are battery-powered and produce vapor instead of smoke. Some common brands include NJOY, Blu, and Smoking Everywhere'. At Waves 2 and 3, they answered whether they used electronic nicotine products in the past 12 months. Examples of electronic nicotine products were described as as e-cigarettes, e-cigars, e-pipes, e-hookahs, and personal vaporizers, as well as vape pens and hookah pens that are battery-powered, use nicotine fluid rather than tobacco leaves, and produce vapor instead of smoke. Some common brands include Fin, NJOY, Blu, e-Go and Vuse'. Participants were shown generic pictures of the product at all three waves. I refer to affirmative responses to these questions as nicotine vaping, a binary variable, throughout.

Mediators

I posited three binary mediators, which I collectively term new onset conduct problems: 1) lying or conning to get something (or get one's way), 2) bullying perpetration, and 3) starting fights. They had to occur at least twice in the past 12 months at Wave 4 among those with no prior history. I operationalized the possible mediators this way because I tried to approximate conduct disorder based on criteria in the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5). DSM-5 conduct disorder is "a repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rules are violated, as manifested by the presence of three (or more) of 15 criteria in the past 12 months, with at least one criterion present in the past 6 months."[162] Conduct disorder is a strong measure because it is validated, widely understood by clinical and public health audiences, and taps into numerous escalatory problem behaviors. I could not estimate indirect effects through

DSM-5 conduct disorder because the full 15-symptom battery was unavailable in PATH.[163] Nevertheless, my operationalization reflects a "repetitive and persistent pattern of [problem] behavior" in that the problems had to occur two or more times in the past 12 months. My measure also reflects the past-year timing of symptoms in the conduct disorder definition.

I did not create an aggregate variable from the three conduct problems in PATH although I wanted to approximate DSM-5 conduct disorder, an aggregate variable, because such a measure would not be validated and would introduce measurement error. An advantage of keeping the conduct problems separate is being able to examine reasons for observing associations of different sizes. For instance, a null indirect effect through the three conduct problems could be due to a balancing of indirect effects through the individual mediators, which I could not detect if I estimated a null effect through a single composite variable.

As a reminder, the three binary conduct problems are treated as either a single observed conditional distribution M_A or a single counterfactual conditional distribution M^{A_δ} (i.e., conditional on A or A_δ, respectively, and L). Random draws from M_A and M^{A_δ} are denoted G and G_δ, respectively. For example, one observation's G from M_A could be $(1,1,0)$, and the same observation's G_δ from M^{A_δ} could be $(0,1,0)$.

Confounders

I considered several substance use theories and prior research to help guide confounder selection.[47–50] The theories I considered were problem behavior theory,[46] interactive theory,[50] and common liability theory.[114] I also used the drug, set, and setting framework to consider the context of substance use because the aforementioned theories are context-agnostic.[115] Based on this process, I considered the most important confounders of nicotine vaping and the outcomes, nicotine vaping and conduct problems, and conduct problems and the outcomes to be one's own substance use history, family and friends' influences on substance

use, psychopathology, and school or employment factors.[102] For instance, I controlled for nicotine use other than nicotine vaping, alcohol and other substance use, whether the youth lives with both parents, whether the parent(s) allows tobacco in the home, sensation-seeking tendencies (e.g., I like to do frightening things), internalizing and externalizing problems, school performance, and others. I also controlled for whether adolescents would use e-cigarettes in Wave 1 as an indicator of deviant peer influence, which I posited is partially why youth engage in deviant behaviors, including conduct problems and substance use; this should have partially deconfounded my estimates. I also considered tobacco use (not including nicotine vaping) to be an intermediate confounder (i.e., caused by the exposure and a cause of the mediator and outcome). There is consistent evidence that nicotine vaping increases the risk of smoking.[32,164] as well as that smoking increases the risks of conduct disorders and cannabis and other substance use.[165–167] The intermediate confounder was assessed at Wave 3. In Appendix B.1, I describe my approach to confounding in greater detail. In Appendix Table B.2, I show the selected confounders and how they were measured.

4.3.8 Statistical analysis

Missing data

Missing exposure data in Waves 1 through 3 was about 32% (Appendix Tables B.6-B.7). Missing mediator data in Wave 4 was about 30% (Appendix Tables B.6-B.7) Missing outcome data in Wave 5 was about 37% (Appendix Tables B.6-B.7) I estimated missing data with multiple imputation by chained equations using the R package **mice**.[116] Each of the 15 imputed dataframes were iterated 25 times using predictive mean matching. Estimates were combined using Rubin's rules with the R package **Amelia** for both descriptive and causal analyses described below.[117]

Descriptive analysis

I first described the baseline characteristics of each analytic cohort. Next, I estimated the proportions of individuals reporting past-year nicotine vaping at Waves 1 through 3 and lying or conning to get something, bullying perpetration, or starting fights two or more times in the past year at Wave 4 stratified by each outcome. Then, I estimated the proportions of individuals reporting past-year nicotine vaping at Waves 1 through 3 stratified by the posited mediators at Wave 4 (lying or conning to get something, bullying perpetration, or starting fights two or more times in the past year). In another analysis, I estimated the proportion of baseline characteristics stratified by past-year nicotine vaping at Waves 1 through 3 (Appendix Table B.9). This analysis showed differences in baseline confounders by nicotine vaping status, which was motivation to control for those confounders.

Causal analysis

I estimated stochastic interventional indirect effects of shifting the odds of nicotine vaping at Waves 1, 2, or 3 on outcomes at Wave 5 through possible shifts in the conditional distribution of conduct problems. Specifically, these effects indicated whether lower than observed nicotine vaping odds A_δ distinguished the observed conditional distribution of three conduct problems at Wave 4 (i.e., lying or conning to get something, bullying perpetration, starting fights) M_A from the counterfactual conditional distribution of these conduct problems at Wave 4 M^{A_δ}, and if so, whether a difference in these distributions would have changed the outcome. There was only one estimate of each stochastic interventional indirect effect despite there being three posited mediators because the three posited mediators were considered together as a single distribution.

I used the R package **pub_medshift_interv_biostats** to estimate these effects via the SuperLearner machine learning algorithm.[160] SuperLearner weights candidate algorithms' predicted outcomes by how much they minimize cross-validated loss functions (e.g. mean squared error in linear

regression models).[118–120] Cross-validation is used to avoid overfitting. In cross-validation, data are split into k folds (i.e. equal splits of the data) using $k-1$ folds to train the data and the remaining "hold-out" fold to predict outcomes.[120,121] In my study $k=5$, so each estimator predicted five average outcomes and five corresponding loss functions one for each hold-out fold.[120,121] SuperLearner predicts a single weighted average outcome based on each estimator's performance (i.e. loss minimization).[120,121] I specified sample mean (`SL.mean`), generalized linear models (`SL.myglm`), and gradient boosting trees (`SL.xgboost`) as candidate algorithms in my SuperLearner library. Standard errors were estimated by taking the square root of the sample variance of the influence function.See Appendix B.3 for **pub_medshift_interv_biostats** analysis code.[160]

4.4 Results

4.4.1 Distributions of baseline confounders

In Appendix Table B.8, I describe each analytic sample's baseline characteristics. There were 2,949 and 2,995 baseline participants aged 12 to 16 years old in the cannabis initiation and other substance initiation cohorts, respectively. Only 1.3% of participants vaped nicotine in the past year at Waves 1, 2, or 3 in the former, while 3.4% did in the latter. In general, the samples had healthier behaviors or characteristics associated with better health, such lower parental tolerance of substance use, than in a prior PATH study with less restrictive inclusion criteria.[130] For instance, about 16% to 27% of youth from each cohort reported types of behavioral or mood problems, compared with 20% to 32% in the prior study.[130] Nearly 70% of youth from each cohort had parents who were currently married, and about 35% of youth in each cohort lived separately from at least one parent. In the prior study, 66% of parents were currently married, and 38% of youth lived separately from a parent.[130] About 26% to 27% of respondents from each cohort reported that tobacco is used by someone other than the youth in the home and that tobacco is prohibited in the home. In the other study, 30% to 31% reported the same characteristics.[130] Finally, only 16% and 19% in the cannabis and other substance initiation

cohorts ever used alcohol at baseline, respectively, while over 28% had ever used alcohol in the prior study.[130] Unlike the aforementioned characteristics, higher parental education is a characteristic normally associated with better health but was lower in this study's samples; just over 45% of parents had graduate or professional education in these cohorts, which was the case for 48% of parents in the prior study.[130]

4.4.2 Distributions of nicotine vaping and posited mediators stratified by outcome status

In Table 4.1, I show the distributions of nicotine vaping status and posited mediators stratified by the presence and absence of the outcomes. Individuals who initiated cannabis (N = 474) appeared more likely to vape nicotine at Waves 1, 2, or 3 (26.1% versus 13.7%), lie or con to get something (28.6% versus 16.4%), and start fights at Wave 4 (3.2% versus 2.0%).

Individuals who initiated other substances (N = 154) appeared more likely to vape nicotine at Waves 1, 2, or 3 (27.1% versus 11.1%), lie or con to get something (33.4% versus 18.2%), bully others (4.1% versus 2.9%), and start fights at Wave 4 (3.9% versus 2.3%).

Variance was high because there were few participants who vaped nicotine and bullied others or started fights. Based on comparing point estimates only, there appeared to be associations between cannabis initiation and nicotine vaping, lying or conning, and starting fights, while other substance initiation appeared to be associated with the exposure and all posited mediators.

Table 4.1: Table assessing Wave 5 outcomes' associations with nicotine vaping at Waves 1, 2, or 3 and conduct problems at Wave 4 (i.e., posited mediators) in cannabis initiation cohort (N = 2,949) and other substance initiation cohort (N = 2,995), Population Assessment of Tobacco and Health

	Cannabis/Other substance initiation absent[a]				Cannabis/Other substance initiation present[a]			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[b]	Weighted %[c]	Lower (%)	Upper (%)	N[b]	Weighted %[c]	Lower (%)	Upper (%)
Cannabis initiation cohort								
Nicotine vaping	311	13.67	11.76	15.57	120	26.06	21.06	31.06
Lying or conning to get something	406	16.39	14.29	18.50	134	28.54	21.80	35.28
Bullying perpetration	87	3.40	2.09	4.70	19	3.41	1.06	5.76
Starting fights	57	1.96	1.11	2.81	14	3.23	0.75	5.71
Other substance initiation cohort								
Nicotine vaping	292	11.14	9.54	12.74	50	27.14	14.83	39.54
Lying or conning to get something	516	18.19	16.79	21.03	51	33.42	21.97	44.88
Bullying perpetration	76	2.89	1.91	3.87	12	4.14	0	10.89
Starting fights	78	2.25	1.41	3.09	10	3.90	0	9.58

[a] There were 474 and 154 participants who initiated cannabis and other substances, respectively. Subtract these numbers from cohort size for participants without outcomes.

[b] N = Average unweighted frequency across 15 imputed datasets.

[c] All proportions are weighted to reflect the complex survey design of the PATH and are a weighted average based on 15 imputed datasets to account for missingness. Proportions are weighted by the baseline probability of being sampled.

4.4.3 Distribution of nicotine vaping stratified by posited mediators

Table 4.2 shows the distributions of nicotine vaping stratified by posited mediators in each outcome cohort. In the cannabis initiation cohort (N = 2,949), there were 540 participants who lied or conned to get something, 106 who bullied others, and 71 who started fights at Wave 4. There were 198 participants who vaped nicotine at Waves 1, 2, or 3. About 6% who lied and 7% of participants who did not lie vaped nicotine, while 14.9% who bullied others and 20.2% who started fights each vaped nicotine. These proportions appeared higher than the proportions of participants who vaped nicotine among those who did not bully others or start fights (about 6.5% each).

In the other substance initiation cohort (N = 2,995), there were 567 participants who lied or conned to get something, 88 who bullied others, and 88 who started fights at Wave 4. Three hundred forty-two participants vaped nicotine at Waves 1, 2, or 3. About 12% who lied and 13% of participants who did not lie each vaped nicotine, while 20.6% who bullied others and 25.6% who started fights each vaped nicotine. These proportions appeared higher than the similar proportions of participants who vaped nicotine among those who did not bully others or start fights (about 11.7% each).

Like earlier descriptive results, variance was high because there were few participants who vaped nicotine and bullied others or started fights. Based on comparing point estimates only, there appeared to be associations between nicotine vaping and both bullying perpetration and starting fights in each cohort.

Table 4.2: Table assessing associations between three conduct problems at Wave 4 (i.e., posited mediators) and nicotine vaping at Waves 1, 2, or 3 in cannabis initiation cohort (N = 2,949) and other substance initiation cohort (N = 2,995), Population Assessment of Tobacco and Health

	Posited mediator absent[a]				Posited mediator present[a]			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[b]	Weighted %[c]	Lower (%)	Upper (%)	N[b]	Weighted %[c]	Lower (%)	Upper (%)
Cannabis initiation cohort								
Nicotine vaping stratified by posited mediators:								
Lying or conning to get something	155	6.63	5.09	8.18	43	6.38	2.71	10.05
Bullying perpetration	186	6.30	4.84	7.75	12	14.88	3.96	25.80
Starting fights	186	6.29	4.86	7.72	12	20.20	3.88	36.52
Other substance initiation cohort								
Nicotine vaping stratified by posited mediators:								
Lying or conning to get something	279	11.83	10.27	13.40	63	12.77	9.00	16.54
Bullying perpetration	326	11.76	10.32	13.19	16	20.59	8.97	32.21
Starting fights	318	11.62	10.14	13.10	24	25.58	13.44	43.73

[a] In the cannabis initiation and other substance initiation cohorts, there were 540 and 567 participants who lied, respectively, 105 and 88 who bullied others, respectively, and 71 and 88 who started fights, respectively. Subtract these numbers from cohort size for participants without posited mediators.

[b] N = Average unweighted frequency across 15 imputed datasets.

[c] All proportions are weighted to reflect the complex survey design of the PATH and are a weighted average based on 15 imputed datasets to account for missingness. Proportions are weighted by the baseline probability of being sampled.

4.4.4 Stochastic interventional indirect effects

In Figure 4.2, I plotted estimates of stochastic interventional indirect effects of lower nicotine vaping odds on the initiation of cannabis and the initiation of other substances if lower nicotine vaping odds changed only the conditional distribution of three conduct problems. There were no meaningful effects on either outcome. For example, the right-most points on each plot indicate that if nicotine vaping odds were 10% lower than observed and had mediator values been stochastically drawn from observed versus counterfactual conditional distribution of conduct problems, there would have been only eight more cannabis initiates per 10,000 youth (95%CI -129:135) and only four more initiates of other substances per 10,000 youth (95%CI -72:82). There were still no meaningful risk differences even if nicotine vaping odds were 90% lower. The left-most points on each plot indicate that 90% lower nicotine vaping odds would have resulted in only one fewer cannabis initiate per 10,000 youth (95%CI -19:16) and only four more initiates of other substances per 100,000 youth (95%CI -263:264) if 90% lower nicotine vaping odds changed only the distribution of conduct problems. The plots also show that there were no meaningful effects of shifts in nicotine vaping odds between $\delta = 0.9$ (10% lower odds) and $\delta = 0.1$ (90% lower odds). Therefore, the data were inconsistent with mediation via the three conduct problems.

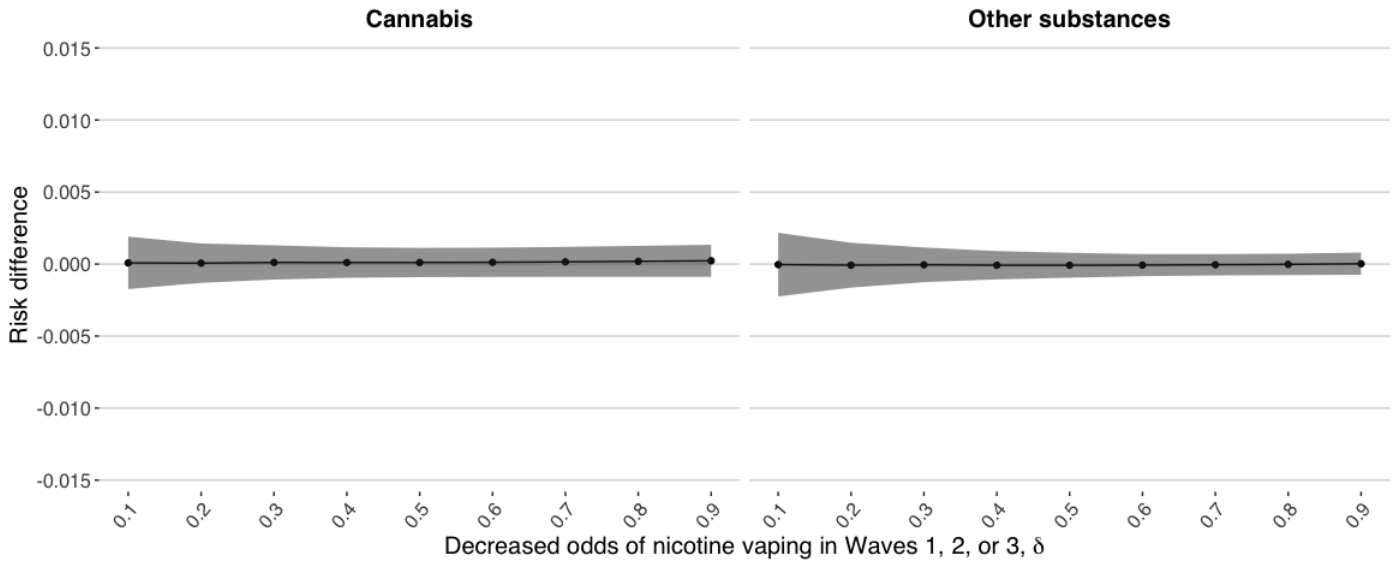

Figure 4.2: Stochastic interventional indirect effects: effects of shifting nicotine vaping odds on cannabis initiation (left) and the initiation of other substances (right) if lower nicotine vaping odds changed only the conditional distribution of three conduct problems (lying or conning to get something, bullying perpetration, and starting fights at least twice in the past 12 months).

4.5 Discussion

This was the first study to systematically examine possible mechanisms through which nicotine vaping affects the initiation of cannabis and the initiation of other substances. Specifically, I posited that lying or conning to get something, bullying perpetration, and starting fights (considered together) mediated the effects. Contrary to expectations, I found no evidence that these conduct problems mediated the effects together.

For mediation through the three conduct problems to be present, I would expect associations between nicotine vaping and the conduct problems and associations between the conduct problems and each of the outcomes. I possibly estimated null indirect effects because there was no association between nicotine vaping and lying or conning, and there was no association between bullying perpetration and cannabis initiation; this evidence would suggest a low probability of mediation when considering indirect effects through the three posited mediators together. These results of the mediation analyses were surprising because in prior research nicotine vaping was associated with a greater frequency of conduct problems, including lying, bullying perpetration, and starting fights,[151] and because bullying perpetration was prospectively associated with cannabis use in the first three waves of PATH.[168] Though associations necessary for mediation through all three conduct problems were null, I still estimated indirect effects because I considered the mediators together (as a conditional distribution). Even if a change in nicotine vaping changed bullying and fighting (but not lying), the potentially shifted conditional distribution of the three posited mediators could have still changed the outcome. In other words, positive indirect effects would have still indicated mediation through the three conduct problems together.[169]

Another reason my results are incompatible with my mediation hypothesis could be selection bias arising from strict eligibility criteria. Specifically, eligible participants had to be free

of a posited mediator through Wave 3, conditional on being free of the other two posited mediators through Wave 3, and participants had to be outcome-free through Wave 4, conditional on being free of all posited mediators through Wave 3. Practically, the youngest participants had to be about 14 years old without ever having conduct problems at Wave 3; however, the median age of onset of conduct disorder is 11 years old, and most youth are diagnosed with conduct disorder between five to 15 years old.[170,171] In other words, it is possible that participants from the original PATH cohort (N = 13,651) for whom there were indirect effects were ineligible for the current study because they already had the mediators (and outcomes); however, selection bias possibly resulted in a sample in which there were no indirect effects. Another possible consequence of strict eligibility criteria was inadequate power to detect indirect effects. That is, there potentially were too few participants at risk of and positive for the conduct problems and the outcomes. For example, bullying and fighting were rare in these analytic cohorts (<4%), as were the outcome risks compared with prior research. In this study, the observed cannabis initiation risk was 17% and the observed risk of other substance initiation was 6%, while the risks in Chapter 3 were 45% and 26%, respectively.[130] Future studies may consider expanding eligibility to younger at-risk participants when estimating indirect effects to avoid selection bias and inadequate power.

Another potential reason for the null indirect effects is that nicotine vaping causes sustained use (but not initiation) of outcome substances through deviant peer affiliation and conduct problems. Others have found that nicotine may lead to sustained use of cocaine through neurobehavioral mechanisms like compromised inhibitory control, although this research was conducted in rodents, and it is unknown if the same effects occur in humans.[99,100] Future studies may consider estimating nicotine vaping's indirect effects on sustained use of outcome substances through deviant peer affiliation, conduct problems, or constructs associated with them.

There are a few additional study artifacts that may be useful in explaining the results. First, (dependent nondifferential) misclassification due to social desirability was likely because participants self-reported sensitive topics. Participants may provide socially desirable (i.e., inaccurate) responses to questions about stigmatized or illicit behaviors, such as substance use under the legal age, which likely resulted in underreporting and bias away from the null;[86,102] however, PATH used Audio Computer-Assisted Self-Interviewing and Computer-Assisted Personal Interviewing, and participants were assured that responses were confidential to avoid socially desirable responses. There was also likely independent nondifferential misclassification of some variables, which was expected to bias estimates towards the null. For instance, it may be difficult for participants to remember the last time they used a substance or whether they lied or conned to get something two or more times. Another manifestation of independent nondifferential misclassification that could explain the null results is poor construct validity of DSM-5 conduct disorder. DSM-5 conduct disorder requires at least three of 15 symptoms to be present in the past year with at least one present in the past six months; however, only three of the 15 symptoms were measured either in the past year or in the past month (not in the past six months per DSM-5). Had DSM-5 conduct disorder been measured, I may have estimated positive indirect effects through it. A prior study suggested that bias away from and towards the null due to dependent and independent nondifferential misclassification, respectively, could balance each other out.[130] A related issue with my measures and even DSM-5 conduct disorder is pathologizing behaviors that may not be pathological in some contexts. For instance, adolescents may steal food repeatedly if their families are experiencing an extended period of poverty or hunger. In my study, adolescents who never start fights could all of a sudden start fights with someone who repeatedly bullies them. That is, "conduct problems" may be a suboptimal description of the mediators posited in this study yet consistent with prior work. Categorizing truly unproblematic behaviors as conduct problems may be another instance of independent nondifferential misclassification expected to bias estimates towards the null. A second limitation was unmeasured confounding due to friends' substance

use, parents' smoking during prenatal period or during childhood or adolescence, parental supervision, and neighborhood risks (e.g., density of vape shops). Direct measures of these constructs were absent; however, I comprehensively controlled for indirect measures in my analyses (e.g., someone else uses tobacco in the home, house rules about tobacco in the home, noticed e-cigarette advertising). Importantly, I controlled for deviant peer influence measured as whether adolescents would use e-cigarettes if a best friend offered, which may have partially deconfounded the exposure-mediator, exposure-outcome, and mediator-outcome effects. Confounding is expected to lead to positive bias in this study, so successfully controlling it could attenuate effects to the null. Third, methods were not yet available at the time of analysis to validly estimate stochastic interventional mediation effects when there were simultaneously multiple or continuous intermediate confounders and multiple or continuous mediators. Since I posited three mediators, my analysis had one composite intermediate confounder variable, non-vaping tobacco use, because it is a known outcome of nicotine vaping and is known to cause both outcomes.[32,164,166,167,172,173] In other words, there was a risk for unmeasured intermediate confounding.[102] Last, about 29% to 32% of individuals were missing exposure data at Waves 1, 2, or 3 and mediator data at Wave 4, while 36% to 37% were missing and outcome data at Wave 5. To avoid potential selection bias, I estimated missing data using multiple imputation by chained equations under the missing at random assumption. This assumption may have been inadequate.

Despite the null results, this study has important takeaways. First, the results do not mean that lying, bullying, and fighting do not mediate the effects; however, they are a piece of evidence against mediation through these factors. Second, conduct problems (and deviant peer affiliation) are likely still threats to adolescent health. Third, my study does not validate the total effects of nicotine vaping on the initiation of cannabis and other substances estimated in prior research.[10,26,130] Other researchers should study mediation to help validate the total effects. Finally, my study shows that researchers who use PATH to estimate nicotine vaping's

indirect effects should carefully consider addressing exposure positivity violations in the design and/or analysis.

4.5.1 Conclusions

In summary, I found no evidence supporting indirect effects of nicotine vaping on cannabis and prohibited substance initiation through conduct problems among adolescents as they aged into adulthood from 2013 to 2019. Public health interventions designed to prevent people from vaping nicotine already exist. They include the increase in the age to legally purchase nicotine products,[127] bans on nicotine vaping flavors aimed at youth consumers,[128] the FDA ban of JUUL products,[129] and vaping restrictions where other tobacco products are not allowed. While these interventions may prevent nicotine vaping and its sequelae, intervening on mediators may help prevent cannabis initiation and initiation of other substances among those who started nicotine vaping. Thus, it is imperative to continue this line of scientific inquiry.

Chapter 5: Conclusions

Vaping is among the highest priority adolescent health concerns in the United States. Its rapid increase among youth during the 2010s signified a threat to the decades of progress in reducing adolescent nicotine and tobacco use.[174–177] It also appears to affect cannabis use and the use of other substances, suggesting it could have even bigger implications for adolescent health than just increasing nicotine use.[25–28] This dissertation addressed the causal effects of nicotine vaping on cannabis use and the use of other substances in several ways.

First, I systematically reviewed the literature to evaluate whether the reported associations between nicotine vaping and cannabis use or other substance use were robust to potential bias and whether any study assessed effect measure modification and mediation. *Second*, I examined whether nicotine vaping affected the initiation of cannabis and the initiation of other substances from 2013 to 2019 among youth as they aged into adulthood, as well as whether any of these effects varied by age, and separately, whether any of these effects varied by time. *Third*, I examined whether three conduct problems together mediated nicotine vaping's effects on cannabis initiation and the initiation of other substances. The results of these studies are discussed below.

5.1 Summary of findings

In the systematic literature review (**Chapter 2**), I examined the extent to which confounding, measurement errors, and loss to follow-up may alternatively explain consistently positive longitudinal associations between nicotine vaping and cannabis use and between nicotine vaping and the use of other substances (excluding alcohol) found across 17 eligible studies.

First, I showed that studies often did not control for unmeasured confounders based on theory and prior research. *Second*, I showed that different kinds of measurement error likely biased results in different directions, possibly balancing each other out. *Third*, I showed that differential loss to follow-up likely biased associations towards the null; that is, observed positive associations were likely more positive than they would would have been without bias. Based on the strengths of the potential biases, I concluded that nicotine vaping likely has positive longitudinal effects on cannabis use and use of other substances over six to 24 months, but any positive effects are likely smaller than estimated.

Besides analyzing the extent of bias in associations, I also systematically reviewed whether any studies evaluated effect measure modification or mediation effects. Concerning the former, two studies found that nicotine vaping's effects on cannabis initiation were stronger among younger (versus older) adolescents and among adolescents who had a higher (versus lower) sensation-seeking proclivity or who smoked (versus did not smoke). These results were unexpected because the stronger effects were estimated in groups typically considered lower risk. The unexpected direction of effect measure modification could be because older adolescents and adolescents who have a higher sensation-seeking proclivity or who vape were at a risk ceiling. This means their baseline risks of cannabis initiation (i.e., among individuals who do not vape) may be higher than baseline risks of adolescents who were younger, had a lower sensation seeking proclivity, or did not smoke; higher baseline risks leave little room for additional risk attributable to nicotine vaping. Evidence of effect measure modification adds context to the effects. Finally, my review revealed that no study investigated mechanisms of the effects. Investigating mechanisms could help test the total effects' validity. Two observations from this review led directly to this dissertation's second and third aims. First, no study estimated effects lasting longer than 24 months; this led directly to Aim 2. Second, no study tested mediation hypotheses; this led directly to Aim 3.

In **Chapter 3**, I sought to estimate nicotine vaping's effects on cannabis initiation and the initiation of other substances lasting up to six years in response to my observations from **Chapter 2** that studies estimated effects lasting up to 24 months. Unfortunately, my data did not support identification of the longitudinal average treatment effect or other conventional static estimands because positivity, a necessary and verifiable causal inference assumption, was unsatisfied.[111] To overcome this issue, I estimated longitudinal incremental propensity score effects because they guarantee positivity. Incremental propensity score effects in **Chapter 3** compare observed outcome risks with outcome risks had nicotine vaping odds been lower. They guarantee positivity because outcomes are conditioned on non-zero events (i.e., propensity scores, which are defined for all observed covariates, and the observed covariates themselves). Incremental propensity score effects can be interpreted as the difference in the risks of initiating cannabis or other substances over a six-year period and those risks had, counterfactually, the conditional probability of nicotine vaping (i.e., propensity score) been incrementally lower than observed.

The estimated incremental effects in my study were increasingly positive because observed risks were higher than risks that would have occurred had nicotine vaping odds been incrementally lower. These effects suggest that nicotine vaping affected the initiation of cannabis and other substances up to six years later. They are concerning because nicotine vaping can set young people up for harms related to cannabis and other substances years later that otherwise may not have happened (e.g., mental health problems, substance dependence, infectious disease, etc.).[13–20] I also estimated whether effects differed by age within each time period and whether effects differed across time within each age grouping. This analysis showed that effects did not meaningfully differ by age within each wave, but they may meaningfully differ across time within each age grouping and overall. Though I did not test the presence of cohort effects, the absence of age effects suggest the absence of cohort effects. Finally, I estimated effects by non-vaping tobacco use history and found that effects were stronger among those with a

history.

In **Chapter 4**, I sought to estimate mechanisms through which nicotine vaping affected cannabis initiation and the initiation of other substances in response to my observation from **Chapter 2** that no evidence of mechanisms existed. The possible mechanism with the strongest support in the literature was deviant peer affiliation, but data measuring the construct were scant. I argued that conduct problems were possible mediators because nicotine vaping may cause deviant peer affiliation,[135–140] which predicts an array of conduct problems.[143–145] Ideally, I would have estimated natural indirect effects, but my hypothesized causal structure included an intermediate confounder (a common cause of the posited mediators and outcome, which is caused by the exposure). A workaround when intermediate confounding is present is estimating an interventional indirect effect of a static exposure; however, my data did not support identification of an effect of a static exposure because positivity was unsatisfied as in **Chapter 3**. The hypothesized presence of intermediate confounding and positivity violations necessitated estimating an analog of the interventional indirect effect of a static exposure called the stochastic interventional indirect effect. In **Chapter 4**, these effects were the differences in the potential risks of initiating cannabis or other substances had mediator values been stochastically drawn from observed versus counterfactual conditional distributions of conduct problems, and counterfactually, had the probability of nicotine vaping in both cases been incrementally lower than in the observed data. As with natural indirect effects, stochastic interventional indirect effects tell us that nicotine vaping affects the outcomes if nicotine vaping affects conduct problems (and includes interaction between nicotine vaping and the conduct problems); however, they do not quantify the magnitude of static nicotine vaping's indirect effects on the outcomes.

I posited three conduct problems—lying or conning to get something, bullying perpetration, and starting fights—as mediators because they may indicate escalation from minor deviant

behavior like nicotine vaping to deviant behaviors with potentially more serious consequences, such as cannabis and other substance use. Conduct problems are also consistently associated with deviant peer behavior and affiliation, which are some of the most reliable predictors of deviant behaviors in adolescence and young adulthood, including substance use and conduct problems.[71,146,148,151,154,155,157,178–180]

There were no meaningful stochastic intervention indirect effects of nicotine vaping odds on either outcome. There are several possible explanations for the absence of indirect effects. One is that nicotine vaping was not associated with lying or conning, and bullying was not associated with cannabis initiation. This evidence would suggest a low probability of mediation when considering stochastic interventional indirect effects through the three posited mediators together. I still estimated the indirect effects because positive stochastic interventional indirect effects could have still indicated mediation through the three conduct problems together.[169] Another possible reason for the null indirect effects was selection bias due to strict eligibility criteria (i.e., no cannabis initiation or initiation of other substances through Wave 4 and none of three conduct problems through Wave 3). That is, participants for whom there could have been individual indirect effects were ineligible because they already acquired the outcome(s) or had a conduct problem in an earlier wave. A third possible reason for the null indirect effects is that nicotine vaping causes sustained use (but not initiation) of outcome substances through conduct problems. Finally, the three conduct problems (available in PATH and) considered together in the analysis may have been poor measures of the ideal measure, DSM-5 conduct disorder. Had DSM-5 conduct disorder been available, I may have estimated positive indirect effects through it.

5.2 Reflection on the dissertation's results and achievements

5.2.1 Longitudinal effects

The goal of this dissertation was to advance evidence of nicotine vaping's potential harmful effects on youth and young adults, which could be used to support interventions aimed at reducing the burden of nicotine vaping's outcomes. To my knowledge, it is the first study to quantitatively examine the long-term effects of nicotine vaping on the initiation of cannabis and the initiation of other substances among adolescents as they age into adulthood, whether these effects differed by age within time periods, and whether these effects differed across time periods within age groupings. This evidence was needed because vaping's prevalence continued increasing during the second half of the 2010s,[23] and there was no prior evidence that effects lasted longer than 24 months. This dissertation shows that nicotine vaping may have persistent long-term effects on negative health outcomes and suggests that youth of different ages were similarly affected.[130] The results help justify the need for continued commitment towards universal youth nicotine vaping prevention, not just during middle and high school years, but even into early adulthood. Preventing youth nicotine vaping may help reduce the number of new initiates of cannabis and other substances.

My findings should be considered along with other evidence supporting different goals when making decisions about nicotine vaping products. For example, another goal could be advancing nicotine vaping's effectiveness at reducing smoking and smoking-related illness among nicotine-dependent people, although the research supporting this goal is mixed. In a longitudinal study using PATH data, Glasser *et al.* found that adults who smoked at baseline and used e-cigarettes daily for three years were two times more likely to quit smoking than individuals who never used e-cigarettes;[181] however, inconsistent e-cigarette use over the same period was associated with a lower likelihood of smoking cessation compared with non-use of e-cigarettes.[181] A different longitudinal study using more recent PATH data found that e-cigarette use was

less effective for smoking cessation than pharmaceutical aids and no aids;[182] Together, the evidence suggests that nicotine vaping could be effective smoking cessation devices if used regularly.

Nicotine vaping's more consistent effects on cannabis and other substance initiation among youth than on smoking cessation among adults might suggest that restricting nicotine vaping should be prioritized more than promoting it health benefits; however, policymakers should be careful when making decisions without enough evidence. For example, prioritizing the promotion of nicotine vaping over its restriction may make more sense if a tiny number of adolescents who vape nicotine heavily smoke or use cannabis by adulthood, and a large number of cigarette-dependent adults quit smoking by regularly vaping nicotine. There are also other considerations that may require nuance. Around 24% of adults smoke cigarettes and 14% of high school students vape in West Virginia, while 15.5% of adults smoke cigarettes and 15.7% of high school students vape in Utah;[183,184] reducing the harms of smoking may be a bigger priority in the former than in the latter owing to its larger smoking prevalence. There are also real-world examples of policy decisions made without adequate evidence possibly resulting in unrealized public health benefits. The FDA and multiple state agencies banned flavored vaping products because the products were thought to increase the risk of youth smoking despite evidence to the contrary[185,186] and an apparent benefit for adult smoking cessation.[181,187,188] Basically, there is no evidence that youth who vaped flavored nicotine were more likely to end up smoking than if they had vaped unflavored nicotine,[185,186] yet there is some evidence that adults who vaped flavored nicotine daily were more likely to quit smoking than adults who did not vape flavored nicotine daily.[181,187,188] This example shows us the importance of patience when it comes to making evidence-based decisions; in this case, adults can no longer legally access flavored vaping products to help them quit smoking because policymakers hastily made decisions without enough evidence.

5.2.2 Mechanisms

Consistent with my goal of advancing evidence of nicotine vaping's potential harmful effects on youth and young adults, this dissertation is also the first to quantitatively examine possible mechanisms of nicotine vaping's effects on cannabis initiation and the initiation of other substances. Testing mechanisms serves multiple purposes. One is that mechanisms could be intervention targets to reduce the burden of outcomes and further harms. For instance, if the three conduct problems considered together mediated the effects, intervening on them could reduce the risks of the outcomes. Some may question whether intervening on the conduct problems together would be easier than intervening on nicotine vaping since interventions on the latter already exist (e.g., raising the minimum age to purchase nicotine and tobacco products,[127] bans on flavored nicotine products aimed at youth consumers,[128] the FDA ban of JUUL products[129]). In other words, what did investigating conduct problems really offer for the purpose of potentially intervening? Had I shown that the conduct problems were mechanisms, it may not necessarily be easier to intervene on them than nicotine vaping; however, some interventions exist. For instance, supervised after-school activities (e.g., club and sports participation) could prevent peers from being deviant and engaging in conduct problems,[189] although the evidence is mixed.[190]

Testing mechanisms also helps build an edifice of knowledge about how effects occur. That is, knowing that one mechanism mediated the effects could lead to inquiry about subsequent mechanisms in the causal sequence on which interventions are more likely. For example, if evidence existed that conduct problems and deviant peer affiliation were mechanisms, others could investigate whether nicotine vaping causes neurocognitive deficits and whether the latter causes conduct problems and subsequent use of other substances. The relatively new Adolescent Brain and Cognitive Development Study could be useful for understanding potential neurocognitive pathways.

A third purpose of testing mechanisms is to build confidence that total effects are valid. That is, had the evidence supported my mediational theory, I would be more confident in the total effects. Not finding evidence for mediation does not mean I am less confident in the total effects because there are no alternative explanations that would render total effects null. Other tests of mechanisms could build on my study and help validate the total effects.

5.2.3 Methodological advancements

Besides the significance of investigating and finding long-term effects and investigating possible mechanisms, my studies are the only ones among those that asked similar research questions to exhaustively consider multiple conditions for identification of their causal estimands. *First*, I methodically specified potential measured and unmeasured time-invariant and time-varying confounders—and whether measured variables could be proxies for unmeasured confounders—based on a review of prior research and several theories about the development of substance use and problem behaviors. For instance, in **Chapter 4**, I specified deviant peer affiliation as an important common cause of substance use (exposure and outcome) and conduct problems (mediators); therefore, I controlled for a variable that measured whether youth would use e-cigarettes if their best friend offered them. This approach likely led to residual confounding but less than there would have been had I not thoroughly addressed confounding. I encourage readers to navigate to Appendix B.1 for a detailed description of my process of addressing potential measured and unmeasured confounding.

Second, I considered positivity, a necessary and verifiable causal inference assumption. I previously evaluated positivity in PATH data for the longitudinal average treatment effect and found that it was violated.[111] This suggests that positivity did not hold in prior studies that used the same data and estimated estimands of static exposures.[10,26] While these studies were otherwise well conducted—they comprehensively considered measured confounding and selection bias—not addressing positivity violations meant that authors made inferences

about a large proportion of individuals for whom comparisons were unavailable. For instance, Dai *et al.* estimated different positive effects of nicotine vaping on cannabis initiation among younger and older adolescents; however, without positivity, Dai *et al.* is not an optimal source of evidence for effect modification by age. Unlike Dai *et al.*, I estimated incremental propensity score effects by age in **Chapter 3**, which were methodologically appropriate because they guarantee positivity. I found no effect modification by age. My findings do not mean there was none, but I concluded that there was not enough evidence for effect modification by age. This example shows that my studies are among the most reliable sources of evidence on the effects because I used the most appropriate methods to answer my causal questions.

While our effect measure modification results were inconsistent, Dai *et al.* and I estimated consistent total effects. This begs the question, how important is positivity? Had I violated positivity and estimated a longitudinal average treatment effect instead of incremental propensity score effects in **Chapter 3**, my conclusions would have been identical since positive incremental propensity score effects tell us that there is a positive longitudinal average treatment effect. Violating positivity is illogical from a methodological perspective, which prioritizes satisfying all causal identification conditions; however, some may tolerate simply knowing the direction of the effect if causal identification is not a priority. For example, it may not matter to Dai *et al.* that they violated positivity and estimated an odds ratio of 1.9 if the causal odds ratio was 1.5 or 2.5. Dai *et al.*'s effect estimate and the hypothetical causal odds ratios all say the same thing: nicotine vaping increases the risk of cannabis initiation. I value causal identification, so I would have recommended truncating or trimming propensity scores close to zero if I were Dai *et al.*'s co-author.

It is important to keep the bigger picture in mind despite the problems that positivity violations pose. All data have problems, and methods are developed to help overcome them. The incremental propensity score effect was developed to overcome the problem of reporting an incorrect

(longitudinal) average treatment effect or other static exposure effect when positivity is violated. Its main disadvantage is that it does not tell us anything other than what is already known from estimating the average treatment effect (i.e., the direction). Bodies tasked with reducing youth nicotine vaping should equally consider incremental propensity score effects alongside average treatment effects (with positivity violations) given that each estimand provides similar information.

5.3 Reflection on the dissertation's challenges

Despite this dissertation's achievements, there were challenges along the way. *First*, the R function available to estimate stochastic interventional mediation effects at the time I conducted the mediation analyses in **Chapter 4** was not fully developed. Specifically, the method allows for either high-dimensional mediators or high-dimensional intermediate confounders, but the R function's capabilities allowed for only the former. This created tension because I would have preferred analyzing mediation through DSM-5 conduct disorder, a low-dimensional construct, with high-dimensional intermediate confounders. DSM-5 conduct disorder was the next best option for a mediator after deviant peer affiliation and consistent with my mediational theory; however, I could not analyze indirect effects through DSM-5 conduct disorder because a measure of it was unavailable in PATH. This was challenging to accept because one of the major achievements of **Chapter 3** was debiasing effect estimates through comprehensive confounding control, which I hoped to carry into **Chapter 4**; nevertheless, my approach of analyzing mediation through changes in the conditional distribution of the three conduct problems may be a biased partial assessment of mediation through DSM-5 conduct disorder.

Second, I found it difficult to design the study in **Chapter 4** with adequate power. This was because very few participants had the exposure in early waves, so I combined exposure data from Waves 1 to 3 to create the exposure measure. I also applied the following eligibility

criteria based on the timing of the exposure measure leaving about 3,000 participants in each cohort: no outcome prior to Wave 5 and no potential mediators prior to Wave 4. The eligibility criteria indicate delayed problem behavior and substance use, which are associated with salubrious behaviors and may be why indirect effects were not positive.[191,192] In other words, the eligibility criteria possibly resulted in a healthier study population than the target population, and there were possibly not enough mediator or outcome events in the study population to detect effects. The null findings do not mean that lying, bullying, and fighting do not mediate the effects together; however, they are a piece of evidence against mediation through these factors. Also, they could still be mediators in different study populations or at different times in this study population. I may have also estimated positive indirect effects had eligibility criteria been different or if the exposure, mediators, or outcomes were measured differently. Additionally, I could have estimated positive indirect effects through a different set of conduct problems. This initial foray into uncovering the mechanisms of nicotine vaping's effects on the outcomes could serve as a basis for future improvements in design and analysis.

Last, perhaps the biggest challenge was learning to interpret and describe my studies' causal estimands, which were unfamiliar to me at the start of the dissertation. Longitudinal incremental propensity score effects, which I estimated in **Chapter 3**, were more difficult for me to understand than conventional static estimands like the longitudinal average treatment effect. A longitudinal average treatment effect of 0.1 would indicate that for every 100 individuals, 10 more would have initiated an outcome substance if they vaped nicotine at all waves than if they did not vape nicotine at all waves. In contrast, the longitudinal incremental propensity score effects of 0.1 would indicate that there were 10 more new cases of the outcome than there would have been had nicotine vaping odds been lower than observed at the prior wave. I have issues with both estimands (which I will explain by simplifying to the single-time-point setting). With the average treatment effect, it is difficult for me to imagine how everyone in the population could come to vape or not vape. With the incremental propensity score effect, I found it difficult

to imagine how someone can have a probability of vaping while disregarding their observed vaping status. It was also difficult to understand how that probability might be different and what that change conceptually means. I talked myself into understanding the incremental effect by thinking about changing the prevalence of the exposure: the mean of everyone's observed vaping status (i.e., prevalence) equals the mean of everyone's probabilities of vaping; proportionally changing everyone's observed vaping status equals proportionally changing everyone's vaping probabilities. Therefore, incremental propensity score effects in my study tell us the difference between the observed risk and the risk that would have happened if the vaping prevalence were lower.

Fortunately, longitudinal incremental propensity score effects imply a longitudinal average treatment effect. To intuit why, imagine that nicotine vaping always resulted in cannabis initiation, but the prevalence of nicotine vaping was 0; nicotine vaping would appear to have no effect. However, if the prevalence of nicotine vaping were >0, nicotine vaping would always appear to have an effect. The causal pathway $\pi_\delta \rightarrow A \rightarrow Y$ clarifies this intuition; π_δ does not affect Y except through A (where π_δ is the lower propensity score, A is nicotine vaping, and Y is either outcome). Unfortunately, longitudinal incremental propensity score effects do not indicate the magnitude of nicotine vaping's effects on the outcomes over time, which I would have reported had longitudinal average treatment effects been identifiable.

Another identifiable estimand when positivity is violated is the longitudinal modified treatment policy. One possible longitudinal modified treatment policy is the effect of delaying nicotine vaping by X years and can be written as $\mathbb{E}\left(Y - Y_{\bar{d}}\right)$ where Y is the observed risk, and $Y_{\bar{d}}$ is the risk if everyone's nicotine vaping were delayed by X years;[193] however, like the longitudinal incremental propensity score effect, this estimand suffers a similar interpretational disadvantage in that it does not answer what the effect of nicotine vaping is; rather, it answers what the effect of hypothetically delaying nicotine vaping by X years is. Given the two options, I chose

to estimate longitudinal incremental propensity score effects because 1) they do not have a time component (i.e., delaying exposure) that would have further complicated interpretation of my effect measure modification analyses by age and time, and 2) longitudinal modified treatment policies appear (to me) to be more useful than longitudinal incremental propensity score effects for examining intervention/policy effects (e.g., delaying the implementation of Naloxone Access Laws on opioid overdose rates).[193]

Stochastic interventional indirect effects in **Chapter 4** are also more difficult to understand than conventional static mediation effects. Positive stochastic interventional indirect effects imply that nicotine vaping affects the outcomes because of nicotine vaping's effects on the mediator(s). As with **Chapter 3**'s estimand, a stochastic interventional indirect effect does not express the magnitude of nicotine vaping's effects on the outcomes (through nicotine vaping's effects on the mediators). Nevertheless, it is the only identifiable estimand when positivity is violated and intermediate confounding is posited.

Once I grasped interpretation of the estimands to the point I could conduct the analyses, the next part of the challenge was explaining the estimands to others. In doing so, I tried to adhere to a general roadmap for answering causal questions.[194] One of the roadmap's steps is to define the causal estimand, without simultaneously interpreting it.[194] This was challenging because I was not confident the approach would work unless one was already familiar with the estimand. As stated, the longitudinal average treatment effect is a widely understood estimand. Because it is familiar—its technical description and interpretation are similar—it requires less guidance than the ones I used. Without such familiarity, I felt the audience would have trouble understanding the meanings of incremental propensity score effects and stochastic interventional indirect effects until the discussion sections, where the roadmap prescribes interpretation.[194] In other words, I felt that strict adherence to the roadmap sometimes made it harder on the readers in my studies and overshadowed my

substantive contributions.

Wide adoption of the incremental propensity score effect and stochastic interventional mediation effects in substance use epidemiology is unlikely because their interpretations are challenging, texts describing them are esoteric, and they are consistent (in direction) with static exposure effects that violate positivity. I also do not see these estimands being adopted before additive effects of a static exposure like the average treatment effect, which few substance use epidemiologists estimate in my experience. Most substance use epidemiologists are not also causal inference methodologists, so it is a tall order to expect adoption when the former typically prioritizes interpretation over satisfying causal identification conditions. Substance use epidemiologists who prioritize satisfying causal inference conditions should estimate the novel estimands when positivity is violated.

Coincidentally, the authors who developed stochastic interventional indirect effect introduced the estimand with an example from substance use epidemiology.[160] Hejazi *et al.* estimated stochastic interventional indirect effects of universally increasing the odds of buprenorphine-naloxone dose schedule in which dose is increased early in opioid use disorder relapse treatment (medication for opioid use disorder) on the risk of relapse.[160] Their analysis suggests that pharmacoepidemiology may be a useful place for estimating incremental propensity score effects. That is, if we already know a drug to be effective (as we did for buprenorphine-naloxone), it may be useful to know how prevalent dose increases must be to prevent a particular amount of morbidity or mortality (whether or not positivity is violated).

5.4 Future directions

This dissertation is the first set of studies on nicotine vaping's long-term effects on cannabis initiation and the initiation of other substances as adolescents age into young adulthood. Future studies can advance this dissertation's goal. For example, we could further study

whether nicotine vaping increased the risk of initiating any particular "other substances" in **Chapters 3-4** (i.e., heroin, tranquilizers, cocaine, etc.). I discuss future directions pertaining to each chapter below.

In **Chapter 2**, I conducted a methodological systematic review, which included 17 studies that estimated longitudinal associations between nicotine vaping and both cannabis use and other substance use. In the review, I analyzed whether various sources of bias could alternatively explain apparent effects. I could have conducted a conventional systematic review and meta-analysis, which would have updated one published three years ago and which included only three longitudinal studies.[25] Conducting an updated systematic review and meta-analysis should be a future direction in this research. Additionally, future work could assess the frequency with which positivity is violated in this body of literature if all the data sources were accessible.

In **Chapter 3**, I investigated nicotine vaping's total effects on cannabis initiation and the initiation of other substances up to six years later among youth as they aged into young adulthood. My study restricted participants to the baseline sample, although shadow youth technically entered the PATH cohort at subsequent waves, and there was a replenishment sample at Wave 4. Future research could estimate effects in the dynamic cohort, which would allow for later entry and drop-out. Censoring weights could be used to correct for potential bias due to loss to follow-up instead of multiple imputation by chained equations. This approach could result in positivity (not being violated) if late unexposed entrants had higher conditional probabilities of nicotine vaping than the baseline cohort. If positivity were met, one could estimate an estimand of static longitudinal nicotine vaping.

In **Chapter 4**, I investigated whether new onset conduct problems mediated nicotine vaping's effects on cannabis initiation and the initiation of other substances. Many of the potential

future directions pertaining to **Chapter 3** are applicable to **Chapter 4**. One unique to **Chapter 4** is to continue researching possible mechanisms. One mechanism I could have investigated consistent with my mediational theory was whether participants would have used tobacco products (other than e-cigarettes) if a best friend offered. Offering tobacco to an underage friend is arguably deviant and the friend's closeness implies peer affiliation; therefore, the variable could be a decent measure of deviant peer affiliation. I chose not to operationalize my mediator with this variable because its prevalence was extremely low, meaning that mediator positivity probably would have failed (i.e., very few participants if any who did not vape nicotine would have responded positively). Others could expand upon my mediational theory about deviant peer affiliation and conduct problems or craft an entirely new one. I also previously suggested looking at potential cognitive pathways using data from the Adolescent Brain and Cognitive Development Study.

Generally, more studies are needed that explain the consistent causal effects of nicotine vaping on my dissertation's outcomes. This includes specifying subgroups in which effects are strongest and mechanisms of the effects, given the scarcity of this kind of research.

5.5 Final thoughts

When I began thinking about a foundations essay topic in 2019 after the doctoral methods exam, I knew I wanted to study vaping. The E-cigarette or Vaping Use-Associated Lung Injury (EVALI) crisis was sweeping the country. Several dozen people died from tainted vape cartridges. Very few faculty or students were studying vaping in mid-2019, so as a trainee in the T32 substance use epidemiology training program, I thought I could lead the way with my dissertation.

I first thought to study vaping's potential effects on a respiratory condition like EVALI, so I chose asthma. There were very few participants who developed incident asthma in subsequent waves of PATH data (the optimal longitudinal data set for studying vaping), so any analysis of

the effects would have been underpowered. My next idea was to study whether PATH data were more consistent with the gateway hypothesis or common liability theory. Estimates of total and indirect effects would be consistent with the gateway hypothesis, and no or greatly attenuated effects would be more consistent with common liability theory. I abandoned this plan when it was unclear what a logical literature review chapter (2) would be. Some of the questions I raised for the review chapter were: What theory is the right theory for understanding associations between nicotine vaping and cannabis use? Why are other theories wrong? If other studies do not motivate their study with theory, are their inferences wrong? These were interesting questions but possibly too ambitious to tackle in one dissertation chapter. Also, I could not easily connect them to other planned aims to write a coherent story. In other words, it was unclear why I needed to figure out which theory was better or right to achieve the goal of advancing evidence of nicotine vaping's potential harms.

I was advised to couch my argument in the research gaps instead of the utility of different theories, which led to the current formula for the dissertation. I was already set on studying mechanisms because I was perturbed by the lack of mediation evidence in the gateway hypothesis literature (other than indirect effects through different forms of substance use). As mechanisms require a total effect, I needed to know that some associations between nicotine vaping and my outcomes were likely to be effects; therefore, I planned a review assessing the extent to which prior studies were resistant to bias. Finding some, I could then specify the two gaps that led directly to **Chapters 3** and **4**. First, prior evidence of effects were limited to 24 months, which I sought to expand. Predictably, the second gap was that mechanisms of the effects were unknown.

Crafting the argument for studying mechanisms was just as zigzaggy of a process as arriving at the gaps. I originally argued that nicotine vaping could cause cognitive alterations that make vulnerable individuals more likely to have conduct problems and engage in other

forms of substance use. The hypothesis was partially based on observations from Kandel and Kandel in mice.[99] Part of my argument was that vulnerable individuals tend to form groups of deviant peers and engage in problem behaviors and substance use together.[97,158] Eventually, I decided to remove the piece about cognitive changes from my argument because it was more of a stretch to get from cognitive changes to conduct problems, the planned mediators. It was much easier to argue that nicotine vaping could introduce youth to deviant peers, and consequently the youth could be at an increased risk of conduct problems and the outcomes.

My journey taught me a couple of important lessons. First, it taught me that research is incremental. As described in the preceding paragraphs, I naïvely tried to piece together an incoherent story instead of one in which chapters build on each other. My story's incoherence led me to revise my arguments multiple times until the story was coherent. That I knew nothing about vaping epidemiology prior to starting my dissertation could be partially to blame. In the future, I plan to storyboard projects to avoid similar pitfalls.

I also learned that I and other epidemiologists may have different priorities concerning methodological issues, such as identification and measurement. Most substance use epidemiologists probably ignore positivity, and if not, they may prefer estimating an odds or risk ratio that violates positivity because their audiences understand those estimands better than the ones I estimated. That is, they may prioritize interpretation over identification. There were also competing views within my committee on how I should operationalize conduct problems in **Chapter 4**. One view was that I should operationalize them as a scale of increasing severity (e.g., having none, one, two, or all three), which could tap into DSM-5 conduct disorder. The other view was that I should operationalize them as one conditional joint distribution, which is more difficult to interpret, because a scale would not be validated and could contain measurement error. I valued both viewpoints and eventually went with the latter approach. These examples show that some epidemiologists may weight methods over interpretation

while others may weight the latter more than the former.

In conclusion, I achieved my dissertation's goal of advancing the science on nicotine vaping's harmful effects on youth substance use. In the process, I learned new estimands and estimation approaches and when they can be useful, and I became a better writer and presenter. I end the dissertation leaving the science about nicotine vaping's effects a tad richer than when I began and wish for others to enrich it further.

References

[1] R. Miech et al. “Trends in Adolescent Vaping, 2017–2019”. In: *New England Journal of Medicine* 381.15 (2019), pp. 1490–1491.

[2] A. M. Glasser et al. “Youth Vaping and Tobacco Use in Context in the United States: Results from the 2018 National Youth Tobacco Survey”. In: *Nicotine and Tobacco Research* 23.3 (2021), pp. 447–453.

[3] T. W. Wang et al. “E-cigarette Use Among Middle and High School Students — United States, 2020”. In: *MMWR. Morbidity and Mortality Weekly Report* 69.37 (2020), pp. 1310–1312.

[4] K. A. Cullen et al. “e-Cigarette Use among Youth in the United States, 2019”. In: *JAMA - Journal of the American Medical Association* 322.21 (2019), pp. 2095–2103.

[5] P. Lozano et al. “A longitudinal study of electronic cigarette use and onset of conventional cigarette smoking and marijuana use among Mexican adolescents”. In: *Drug and alcohol dependence* 180 (2017), pp. 427–430.

[6] E. Park et al. “Adolescent E-cigarette use trajectories and subsequent alcohol and marijuana use”. In: *Addictive Behaviors* 103 (2020), p. 106213.

[7] S. M. Rogers et al. “Multi-directional pathways of tobacco and marijuana use, including comorbid use, among a population-based cohort of adolescents in Texas: A six-wave cross-lagged model”. In: *Addictive behaviors* 115 (2021), p. 106771.

[8] R. J. Evans-Polce et al. “Prospective associations of e-cigarette use with cigarette, alcohol, marijuana, and nonmedical prescription drug use among US adolescents”. In: *Drug and alcohol dependence* 216 (2020), p. 108303.

[9] A. E. Butler et al. “Bidirectional associations between cannabis, e-cigarette, and cigarette use among Canadian youth: findings from the COMPASS Study”. In: *Drugs: Education, Prevention and Policy* (2021), pp. 1–9.

[10] K. Bentivegna et al. “Electronic cigarettes associated with incident and polysubstance use among youth”. In: *Journal of Adolescent Health* 68.1 (2021), pp. 123–129.

[11] A.-K. Seidel et al. “Use of electronic cigarettes as a predictor of cannabis experimentation: a longitudinal study among German youth”. In: *Nicotine and Tobacco Research* 24.3 (2022), pp. 366–371.

[12] M. L. Silveira et al. “Longitudinal associations between youth tobacco and substance use in waves 1 and 2 of the Population Assessment of Tobacco and Health (PATH) Study”. In: *Drug and Alcohol Dependence* 191 (2018), pp. 25–36.

[13] N. D. Volkow et al. “Effects of cannabis use on human behavior, including cognition,motivation, and psychosis: A review”. In: *JAMA Psychiatry* 73.3 (2016), pp. 292–297.

[14] D. S. Hasin et al. “Cocaine and heroin dependence compared in poly-drug abusers.” In: *American Journal of Public Health* 78.5 (May 1988), pp. 567–569.

[15] A. J. Rondinelli et al. “Young adult injection drug users in the United States continue to practice HIV risk behaviors”. In: *Drug and Alcohol Dependence* 104.1-2 (2009), pp. 167–174.

[16] W. M. Compton, C. M. Jones, and G. T. Baldwin. “Relationship between nonmedical prescription-opioid use and heroin use”. In: *New England Journal of Medicine* 374.2 (2016), pp. 154–163.

[17] X. Chen, B. Gelaye, and M. A. Williams. “Sleep characteristics and health-related quality of life among a national sample of American young adults: assessment of possible health disparities”. In: *Quality of Life Research* 23.2 (2014), pp. 613–625.

[18] C. N. Yen et al. “Quality of life and its correlates among heroin users in Taiwan”. In: *Kaohsiung Journal of Medical Sciences* 27.5 (2011), pp. 177–183.

[19] A. C. Paolini. “Heroin usage: Impact on student performance and truancy among high school students”. In: *Journal of Drug Abuse* 2 (2016), pp. 1–7.

[20] A. S. Perlmutter et al. “Is employment status in adults over 25 years old associated with nonmedical prescription opioid and stimulant use?” In: *Social psychiatry and psychiatric epidemiology* 52.3 (2017), pp. 291–298.

[21] K. W. Griffin and G. J. Botvin. “Evidence-based interventions for preventing substance use disorders in adolescents”. In: *Child and Adolescent Psychiatric Clinics* 19.3 (2010), pp. 505–526.

[22] R. A. Arrazola, S. R. Dube, and B. A. King. “Tobacco product use among middle and high school students—United States, 2011 and 2012”. In: *Morbidity and Mortality Weekly Report* 62.45 (2013), p. 893.

[23] R. A. Miech et al. “Monitoring the Future National Survey Results on Drug Use, 1975-2020. Volume I, Secondary School Students.” In: *Institute for Social Research* (2021).

[24] Surgeon General of the United States. *Surgeon General’s Advisory on E-cigarette Use Among Youth*. Tech. rep. 2018.

[25] N. Chadi et al. “Association between Electronic Cigarette Use and Marijuana Use among Adolescents and Young Adults: A Systematic Review and Meta-analysis”. In: *JAMA Pediatrics* 173.10 (2019), e192574.

[26] H. Dai et al. “Electronic cigarettes and future marijuana use: A longitudinal study”. In: *Pediatrics* 141.5 (2018), e20173787.

[27] J. Audrain-Mcgovern et al. “Adolescent e-cigarette, hookah, and conventional cigarette use and subsequent marijuana use”. In: *Pediatrics* 142.3 (2018), e20173616.

[28] J. B. Unger, D. W. Soto, and A. Leventhal. “E-cigarette use and subsequent cigarette and marijuana use among Hispanic young adults”. In: *Drug and Alcohol Dependence* 163 (2016), pp. 261–264.

[29] J. M. Castaldelli-Maia et al. “Exploring the latent trait of opioid use disorder criteria among frequent nonmedical prescription opioid users”. In: *Journal of psychiatric research* 80 (2016), pp. 79–86.

[30] N. Chadi, E. R. Weitzman, and S. Levy. “Understanding the Impact of National and State Medical Marijuana Policies on Adolescents”. In: *Current Addiction Reports* 5.2 (2018), pp. 93–101.

[31] R. J. Evans-Polce et al. “Initiation Patterns and Trends of E-Cigarette and Cigarette Use Among U.S. Adolescents”. In: *Journal of Adolescent Health* 66.1 (2020), pp. 27–33.

[32] G. C. K. Chan et al. “Gateway or common liability? A systematic review and meta-analysis of studies of adolescent e-cigarette use and future smoking initiation”. In: *Addiction* 116.4 (Apr. 2021), pp. 743–756.

[33] C. B. Clark et al. “The impact of non-concordant self-report of substance use in clinical trials research”. In: *Addictive Behaviors* 58 (2016), pp. 74–79.

[34] N. S. Levy et al. “What is the prevalence of drug use in the general population? Simulating underreported and unknown use for more accurate national estimates”. In: *Annals of Epidemiology* 68 (2022), pp. 45–53.

[35] R. C. Trenz et al. “Early onset of drug and polysubstance use as predictors of injection drug use among adult drug users”. In: *Addictive Behaviors* 37.4 (2012), pp. 367–372.

[36] D. L. Snow, J. K. Tebes, and M. W. Arthur. “Panel attrition and external validity in adolescent substance use research.” In: *Journal of Consulting and Clinical Psychology* 60.5 (1992), p. 804.

[37] O. Siddiqui, B. R. Flay, and F. B. Hu. “Factors affecting attrition in a longitudinal smoking prevention study”. In: *Preventive Medicine* 25.5 (1996), pp. 554–560.

[38] J. Powers and D. Loxton. “The Impact of Attrition in an 11-Year Prospective Longitudinal Study of Younger Women”. In: *Annals of Epidemiology* 20.4 (2010), pp. 318–321.

[39] T. VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2015. ISBN: 019932588X.

[40] D. Moher et al. “Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement”. In: *Systematic reviews* 4.1 (2015), pp. 1–9.

[41] A. Perlmutter. “Alcohol initiation usually occurs before initiation of nicotine vaping”. In: *In preparation* (2022).

[42] H. Shahid et al. “Differentiating Patterns of Vaping, Alcohol, and Cannabis Use Among Early Adolescents”. In: *The Journal of Early Adolescence* (Sept. 2022), p. 02724316221116043.

[43] S. E. McCabe, B. T. West, and V. V. McCabe. “Associations Between Early Onset of E-cigarette Use and Cigarette Smoking and Other Substance Use Among US Adolescents: A National Study”. In: *Nicotine & Tobacco Research* 20.8 (July 2018), pp. 923–930.

[44] A. N. Rothrock et al. “Association of E-cigarettes with adolescent alcohol use and binge drinking-drunkenness: A systematic review and meta-analysis”. In: *The American Journal of Drug and Alcohol Abuse* 46.6 (Nov. 2020), pp. 684–698.

[45] A. Persoskie and R. A. Ferrer. “A most odd ratio:: interpreting and describing odds ratios”. In: *American Journal of Preventive Medicine* 52.2 (2017), pp. 224–228.

[46] R. Jessor and S. L. Jessor. *Problem behavior and psychosocial development: A longitudinal study of youth*. New York: Academic Press, 1977. ISBN: 0123847508.

[47] J. D. Hawkins and J. G. Weis. “The social development model: An integrated approach to delinquency prevention”. In: *Developmental and life-course criminological theories*. Routledge, 2017, pp. 3–27. ISBN: 1315094908.

[48] R. F. Catalano et al. “Modeling the etiology of adolescent substance use: A test of the social development model”. In: *Journal of drug issues* 26.2 (1996), pp. 429–455.

[49] U. Bronfenbrenner. “Toward an experimental ecology of human development.” In: *American Psychologist* 32.7 (1977), pp. 513–531.

[50] G. J. Huba, J. A. Wingard, and P. M. Bentler. “Longitudinal analysis of the role of peer support, adult models, and peer subcultures in beginning adolescent substance use: An application of setwise canonical correlation methods”. In: *Multivariate Behavioral Research* 15.3 (1980), pp. 259–279.

[51] B. S. G. Molina et al. “Substance use through adolescence into early adulthood after childhood-diagnosed ADHD: findings from the MTA longitudinal study”. In: *Journal of Child Psychology and Psychiatry and Allied Disciplines* 59.6 (2018), pp. 692–702.

[52] M. J. Van Ryzin, G. M. Fosco, and T. J. Dishion. “Family and peer predictors of substance use from early adolescence to early adulthood: An 11-year prospective analysis”. In: *Addictive Behaviors* 37.12 (2012), pp. 1314–1324.

[53] C. Li, M. A. Pentz, and C. P. Chou. “Parental substance use as a modifier of adolescent substance use risk”. In: *Addiction* 97.12 (2002), pp. 1537–1550.

[54] S. M. King, W. G. Iacono, and M. McGue. “Childhood externalizing and internalizing psychopathology in the prediction of early substance use”. In: *Addiction* 99.12 (2004), pp. 1548–1559.

[55] T. D. Becker et al. “Systematic Review of Electronic Cigarette Use (Vaping) and Mental Health Comorbidity among Adolescents and Young Adults”. In: *Nicotine and Tobacco Research* 23.3 (2021), pp. 415–425.

[56] M. U. Pedersen et al. “Externalizing behavior problems are related to substance use in adolescents across six samples from Nordic countries”. In: *European Child & Adolescent Psychiatry* 27.12 (2018), pp. 1551–1561.

[57] A. M. Wasserman et al. “The externalizing and internalizing pathways to marijuana use initiation: Examining the synergistic effects of impulsiveness and sensation seeking.” In: *Developmental Psychology* 57.12 (2021), p. 2250.

[58] P. L. Ellickson, J. S. Tucker, and D. J. Klein. “High-risk behaviors associated with early smoking: results from a 5-year follow-up”. In: *Journal of adolescent health* 28.6 (2001), pp. 465–473.

[59] E. C. Hair et al. “Longitudinal tobacco use transitions among adolescents and young adults: 2014–2016”. In: *Nicotine and Tobacco Research* 21.4 (2019), pp. 458–468.

[60] A. Glasser et al. “Patterns of E-Cigarette Use among Youth and Young Adults: Review of the Impact of E-Cigarettes on Cigarette Smoking”. In: *Nicotine and Tobacco Research* 21.10 (2019), pp. 1320–1330.

[61] S. Soneji et al. “Association between initial use of e-cigarettes and subsequent cigarette smoking among adolescents and young adults: a systematic review and meta-analysis”. In: *JAMA pediatrics* 171.8 (2017), pp. 788–797.

[62] A. C. Villanti et al. “Association of flavored tobacco use with tobacco initiation and subsequent use among US youth and adults, 2013-2015”. In: *JAMA network open* 2.10 (2019), e1913804–e1913804.

[63] T. J. Van Der Weele and P. Ding. “Sensitivity analysis in observational research: Introducing the E-Value”. In: *Annals of Internal Medicine* 167.4 (2017), pp. 268–274.

[64] T. J. VanderWeele, M. A. Hernán, and J. M. Robins. “Causal directed acyclic graphs and the direction of unmeasured confounding bias”. In: *Epidemiology (Cambridge, Mass.)* 19.5 (2008), p. 720.

[65] M. Fendrich et al. “The utility of drug testing in epidemiological research: results from a general population survey”. In: *Addiction* 99.2 (2004), pp. 197–208.

[66] R. Boykan et al. “Self-Reported Use of Tobacco, E-cigarettes, and Marijuana Versus Urinary Biomarkers”. In: *Pediatrics* 143.5 (May 2019), e20183531.

[67] J. J. Palamar and A. Le. “Discordant reporting of nonmedical amphetamine use among Adderall-using high school seniors in the US”. In: *Drug and Alcohol Dependence* 181 (2017), pp. 208–212.

[68] E. Single, D. Kandel, and R. Faust. “Patterns of multiple drug use in high school.” In: *Journal of health and social behavior* 15.4 (1974), pp. 344–357.

[69] R. N. Cassidy et al. “Initiation of vaporizing cannabis: individual and social network predictors in a longitudinal study of young adults”. In: *Drug and alcohol dependence* 188 (2018), pp. 334–340.

[70] S. W. Wong et al. “Is E-cigarette use a gateway to marijuana use? Longitudinal examinations of initiation, reinitiation, and persistence of e-cigarette and marijuana use”. In: *Drug and Alcohol Dependence* 208 (2020), p. 107868.

[71] P. Pokhrel et al. “Predictors of marijuana vaping onset and escalation among young adults”. In: *Drug and Alcohol Dependence* 216 (2020), p. 108320.

[72] J. Lee et al. “Predictors of vaping marijuana initiation among US adolescents: Results from the Population Assessment of Tobacco and Health (PATH) study Wave 3 (2015–2016) and Wave 4 (2016–2018)”. In: *Drug and alcohol dependence* 226 (2021), p. 108905.

[73] A. J. Ksinan et al. “E-cigarette use is prospectively associated with initiation of cannabis among college students”. In: *Addictive Behaviors* 106 (2020), p. 106312.

[74] A. Ortega et al. “Longitudinal investigation of the bidirectional associations between initiation of e-cigarettes and other substances in adolescents”. In: *Journal of Substance Use* 26.1 (2021), pp. 40–47.

[75] P. Lozano et al. “A longitudinal study of electronic cigarette use and onset of conventional cigarette smoking and marijuana use among Mexican adolescents”. In: *Drug and Alcohol Dependence* 180 (2017), pp. 427–430.

[76] J. H. Hwang et al. “Underestimation of self-reported smoking prevalence in Korean adolescents: evidence from gold standard by combined method”. In: *International journal of environmental research and public health* 15.4 (2018), p. 689.

[77] M. E. Morean et al. “Querying About the Use of Specific E-Cigarette Devices May Enhance Accurate Measurement of E-Cigarette Prevalence Rates Among High School Students”. In: *Nicotine & Tobacco Research* 22.5 (Apr. 2020), pp. 833–837.

[78] R. C. Wiener, L. W. M. Swager, and S. Morgan. “Adolescent Self-Reported Smoking and Electronic Cigarette Use vs. Serum Cotinine Level, NHANES, 2015–16 and 2017–18”. In: *Journal of Child & Adolescent Substance Abuse* (Apr. 2022), pp. 1–9.

[79] T. Johnson and M. Fendrich. “Modeling sources of self-report bias in a survey of drug use epidemiology”. In: *Annals of Epidemiology* 15.5 (2005), pp. 381–389.

[80] T. P. Johnson and P. J. Bowman. “Cross-cultural sources of measurement error in substance use surveys”. In: *Substance Use & Misuse* 38.10 (2003), pp. 1447–1490.

[81] J. J. Palamar. “Barriers to accurately assessing prescription opioid misuse on surveys”. In: *The American journal of drug and alcohol abuse* 45.2 (2019), pp. 117–123.

[82] J. J. Palamar et al. “Hair testing to assess both known and unknown use of drugs amongst ecstasy users in the electronic dance music scene”. In: *International Journal of Drug Policy* 48 (2017), pp. 91–98.

[83] J. J. Palamar et al. “Detection of “bath salts” and other novel psychoactive substances in hair samples of ecstasy/MDMA/“Molly” users”. In: *Drug and alcohol dependence* 161 (2016), pp. 200–205.

[84] J. J. Palamar and P. Acosta. “Synthetic cannabinoid use in a nationally representative sample of US high school seniors”. In: *Drug and Alcohol Dependence* 149 (2015), pp. 194–202.

[85] P. Kristensen. “Bias from Nondifferential but Dependent Misclassification of Exposure and Outcome”. In: *Epidemiology* 3.3 (1992), pp. 210–215.

[86] E. Susser et al. *Psychiatric epidemiology: Searching for the causes of mental disorders*. Ed. by M. D. Begg, J. M. Gorman, and M.-C. King. New York, NY, US: Oxford University Press, 2006, pp. 516, xxii, 516–xxii. ISBN: 0-19-510181-2 (Hardcover); 978-0-19-510181-2 (Hardcover).

[87] D. P. Crowne and D. Marlowe. “A new scale of social desirability independent of psychopathology.” In: *Journal of Consulting Psychology* 24.4 (1960), pp. 349–354.

[88] R. Tourangeau et al. “Population Assessment of Tobacco and Health (PATH) reliability and validity study: Selected reliability and validity estimates”. In: *Tobacco Control* 28.6 (2019), pp. 663–668.

[89] E. Single, D. Kandel, and R. Faust. “Patterns of Multiple Drug Use in High School”. In: *Journal of Health and Social Behavior* 15.4 (1974), pp. 344–357.

[90] J. J. Palamar and A. Le. “Discordant report of vaping of cannabis among high school seniors in the United States”. In: *The American Journal of Drug and Alcohol Abuse* 48.2 (Mar. 2022), pp. 148–157.

[91] K. E. Riehm et al. “Mental health problems and initiation of e-cigarette and combustible cigarette use”. In: *Pediatrics* 144.1 (2019), e20182935.

[92] C. R. Colder et al. “Prospective associations of internalizing and externalizing problems and their co-occurrence with early adolescent substance use”. In: *Journal of Abnormal Child Psychology* 41.4 (2013), pp. 667–677.

[93] S. M. Murphy and R. Rosenman. “The “Real” Number of Washington State Adolescents Using Marijuana, and Why: A Misclassification Analysis”. In: *Substance Use and Misuse* 54.1 (2019), pp. 89–96.

[94] K. F. Romm et al. “Changes in young adult substance use during COVID-19 as a function of ACEs, depression, prior substance use and resilience”. In: *Substance Abuse* 43.1 (2022), pp. 212–221.

[95] L. D. Hawke, L. Wilkins, and J. Henderson. “Early cannabis initiation: Substance use and mental health profiles of service-seeking youth”. In: *Journal of Adolescence* 83 (2020), pp. 112–121.

[96] D. P. Farrington and B. C. Welsh. “Saving children from a life of crime: Early risk factors and effective interventions”. In: (2008).

[97] W. A. Mason, J. E. Hitchings, and R. L. Spoth. “The interaction of conduct problems and depressed mood in relation to adolescent substance involvement and peer substance use”. In: *Drug and Alcohol Dependence* 96.3 (2008), pp. 233–248.

[98] D. M. Fergusson, L. J. Woodward, and L. J. Horwood. “Maternal smoking during pregnancy and psychiatric adjustment in late adolescence”. In: *Archives of General Psychiatry* 55.8 (1998), pp. 721–727.

[99] D. Kandel and E. Kandel. “The Gateway Hypothesis of substance abuse: Developmental, biological and societal perspectives”. In: *Acta Paediatrica, International Journal of Paediatrics* 104.2 (2015), pp. 130–137.

[100] A. Levine et al. “Molecular mechanism for a gateway drug: Epigenetic changes initiated by nicotine prime gene expression by cocaine”. In: *Science Translational Medicine* 3.107 (2011), pp. 107–109.

[101] R. Miech et al. “Trends in Use and Perceptions of Nicotine Vaping among US Youth from 2017 to 2020”. In: *JAMA Pediatrics* 175.2 (2021), pp. 185–190.

[102] A. S. Perlmutter. “Does the current evidence support any effects of nicotine vaping on cannabis and prohibited substance use risks? A systematic review”. In: *In preparation* (2022).

[103] S. W. Wong et al. “Measuring characteristics of e-cigarette consumption among college students”. In: *Journal of American College Health* 67.4 (2019), pp. 338–347.

[104] A. J. Ksinan et al. “E-cigarette use is prospectively associated with initiation of cannabis among college students”. In: *Addictive Behaviors* 106 (2020), p. 106312.

[105] K. C. Alcover and C. L. Thompson. “Patterns of Mean Age at Drug Use Initiation among Adolescents and Emerging Adults, 2004-2017”. In: *JAMA Pediatrics* 174.7 (2020), pp. 725–727.

[106] F. Auersperg et al. “Long-term effects of parental divorce on mental health – A meta-analysis”. In: *Journal of Psychiatric Research* 119 (2019), pp. 107–115.

[107] T. Korhonen et al. “Externalizing Behavior Problems and Cigarette Smoking as Predictors of Cannabis Use: The TRAILS Study”. In: *Journal of the American Academy of Child & Adolescent Psychiatry* 49.1 (2010), pp. 61–69.

[108] Westat. *PATH Study Public Use Files User Guide*. Tech. rep. Ann Arbor, MI, 2019, pp. 1–214.

[109] A. Hyland et al. “Design and methods of the Population Assessment of Tobacco and Health (PATH) Study”. In: *Tobacco Control* 26.4 (2017), pp. 371–378.

[110] M. L. Petersen et al. “Diagnosing and responding to violations in the positivity assumption”. In: *Statistical Methods in Medical Research* 21.1 (Oct. 2010), pp. 31–54.

[111] A. S. Perlmutter. “Estimating the extent to which nicotine vaping increases cannabis initiation risk in the presence of positivity violations”. In: *Society for Epidemiologic Research*. Chicago, 2022.

[112] E. H. Kennedy. “Nonparametric Causal Effects Based on Incremental Propensity Score Interventions”. In: *Journal of the American Statistical Association* 114.526 (Apr. 2019), pp. 645–656.

[113] M. Bonvini et al. “Incremental causal effects: an introduction and review”. In: *arXiv preprint arXiv:2110.10532* (2021).

[114] M. M. Vanyukov et al. “Liability to substance use disorders: 2. A measurement approach”. In: *Neuroscience and Biobehavioral Reviews* 27.6 (2003), pp. 517–526.

[115] N. E. Zinberg. *Drug, Set, and Setting: The Basis for Controlled Intoxicant Use*. Yale University Press, 1986. ISBN: 9780300036343.

[116] S. van Buuren and K. Groothuis-Oudshoorn. “mice: Multivariate imputation by chained equations in R”. In: *Journal of Statistical Software* 45.3 (2011), pp. 1–67.

[117] J. Honaker, G. King, and M. Blackwell. “Amelia II: A Program for Missing Data”. In: *Journal of Statistical Software* 45.7 (Dec. 2011), pp. 1–47.

[118] S. E. Sinisi et al. “Super learning: An application to the prediction of HIV-1 drug resistance”. In: *Statistical Applications in Genetics and Molecular Biology* 6.1 (2007), pp. 1–24.

[119] K. E. Rudolph and M. J. van der Laan. “Robust estimation of encouragement design intervention effects transported across sites”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 79.5 (2017), pp. 1509–1525.

[120] M. J. Van Der Laan, E. C. Polley, and A. E. Hubbard. “Super learner”. In: *Statistical Applications in Genetics and Molecular Biology* 6.1 (2007).

[121] A. I. Naimi and L. B. Balzer. “Stacked generalization: an introduction to super learning”. In: *European Journal of Epidemiology* 33.5 (2018), pp. 459–464.

[122] D. G. Altman and J. M. Bland. “Interaction revisited: the difference between two estimates”. In: *BMJ* 326.7382 (Jan. 2003), p. 219.

[123] J. Girgis et al. “Cannabis Use and Internalizing/Externalizing Symptoms in Youth: A Canadian Population-Based Study”. In: *Journal of Adolescent Health* 67.1 (2020), pp. 26–32.

[124] B. T. Halbert, R. B. Davis, and C. C. Wee. “Disproportionate longer-term opioid use among U.S. adults with mood disorders”. In: *PAIN* 157.11 (2016).

[125] M. A. Davis et al. “Prescription Opioid Use among Adults with Mental Health Disorders in the United States”. In: *The Journal of the American Board of Family Medicine* 30.4 (July 2017), p. 407.

[126] M. S. Askari et al. “Structure and trends of externalizing and internalizing psychiatric symptoms and gender differences among adolescents in the US from 1991 to 2018”. In: *Social Psychiatry and Psychiatric Epidemiology* 57.4 (2022), pp. 737–748.

[127] 116th Congress. *Tobacco to 21 Act*. 2019.

[128] M. Siegel and A. Katchmar. “Effect of flavored E-cigarette bans in the United States: What does the evidence show?” In: *Preventive Medicine* 165 (Dec. 2022), p. 107063.

[129] *FDA Denies Authorization to Market JUUL Products Currently Marketed JUUL Products Must Be Removed from the US Market*. Tech. rep. Food and Drug Administration, June 2022.

[130] A. S. Perlmutter. “Does nicotine vaping affect cannabis and other prohibited substance initiation from adolescence into young adulthood? Results from Wave 1 (2013-14) to Wave 5 (2018-19) of the Population Assessment of Tobacco and Health (PATH)”. In: *In preparation* (2022).

[131] N. S. Timasheff, E. H. Sutherland, and D. R. Cressey. *Principles of Criminology*. Vol. 16. 2. Altamira Press, 1955, p. 143. ISBN: 0759117896.

[132] J. E. Grusec. *Social Learning Theory*. Vol. 1. Englewood cliffs Prentice Hall, 2020, pp. 221–228. ISBN: 9780128165119.

[133] M. Warr. *Companions in Crime*. Cambridge University Press, Feb. 2002. ISBN: 9780521810838.

[134] R. J. Evans-Polce et al. “Reasons for Vaping Among U.S. 12th Graders”. In: *Journal of Adolescent Health* 62.4 (Apr. 2018), pp. 457–462.

[135] J. L. Barrington-Trimis et al. “The E-cigarette Social Environment, E-cigarette Use, and Susceptibility to Cigarette Smoking”. In: *Journal of Adolescent Health* 59.1 (2016), pp. 75–80.

[136] J. K. Pepper et al. “How Do Adolescents Get Their E-Cigarettes and Other Electronic Vaping Devices?” In: *American Journal of Health Promotion* 33.3 (Aug. 2018), pp. 420–429.

[137] G. Kong et al. “Sources of electronic cigarette acquisition among adolescents in Connecticut”. In: *Tobacco regulatory science* 3.1 (2017), pp. 10–16.

[138] H. I. Lanza, G. Motlagh, and M. Orozco. “E-cigarette use among young adults: A latent class analysis examining co-use and correlates of nicotine vaping”. In: *Addictive Behaviors* 110 (Nov. 2020), p. 106528.

[139] G. C. Rocheleau, A. G. Vito, and J. Intravia. “Peers, Perceptions, and E-Cigarettes: A Social Learning Approach to Explaining E-Cigarette Use Among Youth”. In: *Journal of Drug Issues* 50.4 (June 2020), pp. 472–489.

[140] B. H. García et al. “Risk for substance use among adolescents at-risk for childhood victimization: The moderating role of ADHD.” In: *Child Abuse & Neglect* 114 (Apr. 2021), p. 104977.

[141] B. B. Brown and J. Larson. “Peer relationships in adolescence.” In: *Handbook of adolescent psychology: Contextual influences on adolescent development, Vol. 2, 3rd ed.* Hoboken, NJ, US: John Wiley & Sons, Inc., 2009, pp. 74–103. ISBN: 978-0-470-14922-5 (Hardcover).

[142] L. Steinberg and K. C. Monahan. “Age differences in resistance to peer influence.” In: *Developmental psychology* 43.6 (2007), p. 1531.

[143] T. E. Moffitt et al. “Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years”. In: *Development and Psychopathology* 14.1 (2002), pp. 179–207.

[144] C. L. Odgers et al. “Female and male antisocial trajectories: From childhood origins to adult outcomes”. In: *Development and psychopathology* 20.2 (2008), pp. 673–716.

[145] K. C. Monahan et al. “Adolescent Pathways to Co-Occurring Problem Behavior: The Effects of Peer Delinquency and Peer Substance Use”. In: *Journal of Research on Adolescence* 24.4 (Dec. 2014), pp. 630–645.

[146] A. L. Kristjansson et al. “Social Profile of Middle School-Aged Adolescents Who Use Electronic Cigarettes: Implications for Primary Prevention”. In: *Prevention Science* 19.6 (2018), pp. 805–812.

[147] J. P. Hoffmann. “Social Learning, Social Bonds, Self-Control and Adolescent Nicotine Vaping”. In: *Substance Use and Misuse* 56.6 (2021), pp. 819–830.

[148] M. A. Pentz et al. “Parent, peer, and executive function relationships to early adolescent e-cigarette use: A substance use pathway?” In: *Addictive Behaviors* 42 (2015), pp. 73–78.

[149] J. Cho et al. “Developmental patterns of tobacco product and cannabis use initiation in high school”. In: *Addiction* 116.2 (Feb. 2021), pp. 382–393.

[150] B. Glaser, K. H. Shelton, and M. B. M. van den Bree. “The Moderating Role of Close Friends in the Relationship Between Conduct Problems and Adolescent Substance Use”. In: *Journal of Adolescent Health* 47.1 (2010), pp. 35–42.

[151] R. Tokle, G. S. Brunborg, and T. F. Vedøy. “Adolescents’ Use of Nicotine-Free and Nicotine E-Cigarettes: A Longitudinal Study of Vaping Transitions and Vaper Characteristics”. In: *Nicotine & Tobacco Research* 24.3 (Mar. 2022), pp. 400–407.

[152] C. P. Maynard. “Vaping in the Context of Delinquency Among Adolescents: A Cross-Sectional Comparison Among Alcohol and Cigarette Use”. In: *Journal of Drug Issues* 51.3 (Mar. 2021), pp. 531–546.

[153] K. W. Bold et al. “Early age of e-cigarette use onset mediates the association between impulsivity and e-cigarette use frequency in youth”. In: *Drug and alcohol dependence* 181 (2017), pp. 146–151.

[154] I. N. Defoe et al. “Disentangling longitudinal relations between youth cannabis use, peer cannabis use, and conduct problems: developmental cascading links to cannabis use disorder”. In: *Addiction* 114.3 (Mar. 2019), pp. 485–493.

[155] B. F. O’Donnell et al. “Decision making and impulsivity in young adult cannabis users”. In: *Frontiers in psychology* 12 (2021), p. 679904.

[156] D. B. Kandel and K. Andrews. “Processes of Adolescent Socialization by Parents and Peers”. In: *International Journal of the Addictions* 22.4 (Jan. 1987), pp. 319–342.

[157] W. A. Mason, J. E. Hitchings, and R. L. Spoth. “The interaction of conduct problems and depressed mood in relation to adolescent substance involvement and peer substance use”. In: *Drug and Alcohol Dependence* 96.3 (2008), pp. 233–248.

[158] D. M. Fergusson, L. J. Horwood, and E. M. Ridder. “Conduct and attentional problems in childhood and adolescence and later substance use, abuse and dependence: Results of a 25-year longitudinal study”. In: *Drug and Alcohol Dependence* 88.SUPPL.1 (2007), S14–S26.

[159] I. Díaz et al. “Nonparametric efficient causal mediation with intermediate confounders”. In: *Biometrika* 108.3 (2021), pp. 627–641.

[160] N. S. Hejazi et al. “Nonparametric causal mediation analysis for stochastic interventional (in)direct effects”. In: *Biostatistics* (Feb. 2022), kxac002.

[161] C. H. Miles. “On the Causal Interpretation of Randomized Interventional Indirect Effects”. In: *arXiv preprint arXiv:2203.00245* (2022).

[162] *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association, 2013. ISBN: 9780890425541.

[163] Substance Abuse and Mental Health Services. *DSM-5 Changes - Implications for Child Serious Emotional Disturbance*. Tech. rep. Rockville, MD, 2016.

[164] B. A. Primack et al. “Progression to traditional cigarette smoking after electronic cigarette use among us adolescents and young adults”. In: *JAMA Pediatrics* 169.11 (2015), pp. 1018–1023.

[165] D. M. Fergusson, J. M. Boden, and L. J. Horwood. “The developmental antecedents of illicit drug use: Evidence from a 25-year longitudinal study”. In: *Drug and Alcohol Dependence* 96.1-2 (2008), pp. 165–177.

[166] A. Rajabi et al. “Association between tobacco smoking and opioid use: A meta-analysis”. In: *Addictive Behaviors* 92 (2019), pp. 225–235.

[167] R. A. Brown et al. “Cigarette Smoking, Major Depression, and Other Psychiatric Disorders among Adolescents”. In: *Journal of the American Academy of Child & Adolescent Psychiatry* 35.12 (1996), pp. 1602–1610.

[168] M. Azevedo Da Silva and S. S. Martins. “Mutual Influences on Bullying Perpetration and Substance Use Among Adolescents in the United States”. In: *Journal of Adolescent Health* 67.3 (2020), pp. 384–391.

[169] I. Díaz and N. S. Hejazi. “Causal mediation analysis for stochastic interventions”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.3 (2020), pp. 661–683.

[170] R. C. Kessler et al. “Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication”. In: *Archives of general psychiatry* 62.6 (2005), pp. 593–602.

[171] R. M. Rapee et al. “Adolescent development and risk for the onset of social-emotional disorders: A review and conceptual model”. In: *Behaviour Research and Therapy* 123 (2019), p. 103501.

[172] B. W. Chaffee, S. L. Watkins, and S. A. Glantz. “Electronic cigarette use and progression from experimentation to established smoking”. In: *Pediatrics* 141.4 (2018), e20173594.

[173] T. Korhonen et al. “Externalizing Behavior Problems and Cigarette Smoking as Predictors of Cannabis Use: The TRAILS Study”. In: *Journal of the American Academy of Child & Adolescent Psychiatry* 49.1 (2010), pp. 61–69.

[174] N. A. of Sciences and E. Medicine. “Public health consequences of e-cigarettes”. In: (2018).

[175] A. Mcneill et al. *E-cigarettes : an evidence update*. Tech. rep. 2015, p. 111.

[176] A. McNeill et al. *Evidence review of e-cigarettes and heated tobacco products 2018. A report commissioned by Public Health England*. Tech. rep. 2018, pp. 1–243.

[177] A. McNeill et al. “Vaping in England: An Evidence Update Including Mental Health and Pregnancy, March 2020”. In: *Public Health England* March (2020), pp. 1–217.

[178] D. M. Fergusson, J. M. Boden, and L. J. Horwood. “The developmental antecedents of illicit drug use: Evidence from a 25-year longitudinal study”. In: *Drug and Alcohol Dependence* 96.1-2 (2008), pp. 165–177.

[179] A. L. Groom et al. “The influence of friends on teen vaping: A mixed-methods approach”. In: *International journal of environmental research and public health* 18.13 (2021), p. 6784.

[180] J. P. Hoffmann. “Social Learning, Social Bonds, Self-Control and Adolescent Nicotine Vaping”. In: *Substance Use and Misuse* 56.6 (2021), pp. 819–830.

[181] A. M. Glasser et al. “Patterns of E-cigarette Use and Subsequent Cigarette Smoking Cessation Over 2 Years (2013/2014–2015/2016) in the Population Assessment of Tobacco and Health Study”. In: *Nicotine & Tobacco Research* 23.4 (Apr. 2021), pp. 669–677.

[182] R. Chen et al. “Effectiveness of e-cigarettes as aids for smoking cessation: evidence from the PATH Study cohort, 2017–2019”. In: *Tobacco Control* (Jan. 2022), pp. 2021–056901.

[183] *Map of Current Cigarette Use Among Adults*. Tech. rep. Centers for Disease Control and Prevention, 2019.

[184] Gaby Galvin. “Vaping Rates Are Highest in These 10 States”. In: *US News* (2019).

[185] A. S. Friedman and S. Xu. “Associations of Flavored e-Cigarette Uptake With Subsequent Smoking Initiation and Cessation”. In: *JAMA Network Open* 3.6 (June 2020), e203826–e203826.

[186] C. Notley et al. “Youth use of e-liquid flavours—a systematic review exploring patterns of use of e-liquid flavours and associations with continued vaping, tobacco smoking uptake or cessation”. In: *Addiction* 117.5 (May 2022), pp. 1258–1272.

[187] S. S. Hawkins et al. “Flavoured tobacco product restrictions in Massachusetts associated with reductions in adolescent cigarette and e-cigarette use”. In: *Tobacco Control* 31.4 (July 2022), p. 576.

[188] L. Li et al. “How Does the Use of Flavored Nicotine Vaping Products Relate to Progression Toward Quitting Smoking? Findings From the 2016 and 2018 ITC 4CV Surveys”. In: *Nicotine & Tobacco Research* 23.9 (Sept. 2021), pp. 1490–1497.

[189] M. Rorie et al. “Structure and deviancy training in after-school programs”. In: *Journal of Adolescence* 34.1 (2011), pp. 105–117.

[190] K. A. Dodge, T. J. Dishion, and J. E. Lansford. “Deviant Peer Influences in Intervention and Public Policy for Youth”. In: *Social Policy Report* 20.1 (Mar. 2006), pp. 1–20.

[191] R. Spoth et al. “Universal intervention effects on substance use among young adults mediated by delayed adolescent substance initiation.” In: *Journal of consulting and clinical psychology* 77.4 (2009), p. 620.

[192] M. Malmberg et al. “Substance use risk profiles and associations with early substance use in adolescence”. In: *Journal of Behavioral Medicine* 33.6 (2010), pp. 474–485.

[193] K. E. Rudolph et al. “When effects cannot be estimated: redefining estimands to understand the effects of naloxone access laws”. In: *Epidemiology* 33.5 (2022), pp. 689–698.

[194] M. L. Petersen. “Applying a causal road map in settings with time-dependent confounding”. In: *Epidemiology* 25.6 (2014), pp. 898–901.

[195] R. Jessor and S. Jessor. “A social-psychological framework for studying drug use.” In: *NIDA research monograph*. Ed. by D. J. Lettieri, M. Sayers, and H. W. Pearson. Vol. 30. Rockville, Maryland: US Department of Health, Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, 1980, pp. 102–109.

[196] J. E. Donovan and R. Jessor. “Problem drinking and the dimension of involvement with drugs: A Guttman scalogram analysis of adolescent drug use”. In: *American Journal of Public Health* 73.5 (1983), pp. 543–552.

[197] J. E. Donovan and R. Jessor. “Structure of Problem Behavior in Adolescence and Young Adulthood”. In: *Journal of Consulting and Clinical Psychology* 53.6 (1985), pp. 890–904.

[198] J. E. Donovan and R. Jessor. “Structure of Problem Behavior in Adolescence and Young Adulthood”. In: *Journal of Consulting and Clinical Psychology* 53.6 (1985), pp. 890–904.

[199] M. M. Vanyukov et al. “Preliminary evidence for an association of a dinucleotide repeat polymorphism at the MAOA gene with early onset alcoholism/substance abuse”. In: *American Journal of Medical Genetics - Neuropsychiatric Genetics* 60.2 (1995), pp. 122–126.

[200] M. M. Vanyukov et al. “Common liability to addiction and "gateway hypothesis": Theoretical, empirical and evolutionary perspective”. In: *Drug and Alcohol Dependence* 123.SUPPL.1 (2012), S3–S17.

[201] G. J. Huba, J. A. Wingard, and P. M. Bentler. “Framework for an interactive theory of drug use.” In: *NIDA research monograph*. Ed. by D. J. Lettieri, M. Sayers, and H. W. Pearson. Vol. 30. Rockville, Maryland: US Department of Health, Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, 1980, pp. 95–101.

[202] R. JESSOR. “Problem-Behavior Theory, Psychosocial Development, and Adolescent Problem Drinking”. In: *British Journal of Addiction* 82.4 (1987), pp. 331–342.

[203] B. R. Garner, V. K. Belur, and M. L. Dennis. “The GAIN short screener (GSS) as a predictor of future arrest or incarceration among youth presenting to substance use disorder (SUD) treatment”. In: *Substance Abuse: Research and Treatment* 7 (2013), pp. 199–208.

[204] M. L. Dennis, Y. F. Chan, and R. R. Funk. “Development and validation of the GAIN Short Screener (GSS) for internalizing, externalizing and substance use disorders and crime/violence problems among adolescents and adults”. In: *American Journal on Addictions* 15.SUPPL. 1 (2006), s80–s91.

[205] C. M. Hamilton et al. “The PhenX toolkit: Get the most from your measures”. In: *American Journal of Epidemiology* 174.3 (2011), pp. 253–260.

[206] M. M. VANYUKOV, H. B. MOSS, and R. E. TARTER. “Assortment for the Liability to Substance Abuse and Personality Traits”. In: *Annals of the New York Academy of Sciences* 708.1 (1994), pp. 102–107.

[207] M. M. Vanyukov et al. “The MAOA promoter polymorphism, disruptive behavior disorders, and early onset substance use disorder: Gene-environment interaction”. In: *Psychiatric Genetics* 17.6 (2007), pp. 323–332.

[208] M. M. Vanyukov et al. “Measurement of the risk for substance use disorders: Phenotypic and genetic analysis of an index of common liability”. In: *Behavior Genetics* 39.3 (2009), pp. 233–244.

[209] R. E. Tarter et al. “Neurobehavior disinhibition in childhood predisposes boys to substance use disorder by young adulthood: Direct and mediated etiologic pathways”. In: *Drug and Alcohol Dependence* 73.2 (2004), pp. 121–132.

[210] R. E. Tarter et al. “Deviant socialization mediates transmissible and contextual risk on cannabis use disorder development: A prospective study”. In: *Addiction* 106.7 (2011), pp. 1301–1308.

[211] L. Kirisci et al. “Age of alcohol and cannabis use onset mediates the association of transmissible risk in childhood and development of alcohol and cannabis disorders: evidence for common liability.” In: *Experimental and clinical psychopharmacology* 21.1 (2013), pp. 38–45.

[212] L. Kirisci et al. “Longitudinal modeling of transmissible risk in boys who subsequently develop cannabis use disorder”. In: *American Journal of Drug and Alcohol Abuse* 39.3 (2013), pp. 180–185.

[213] L. Kirisci et al. “Prediction of cannabis use disorder between boyhood and young adulthood: Clarifying the phenotype and environtype”. In: *American Journal on Addictions* 18.1 (2009), pp. 36–47.

[214] T. P. Beauchaine and T. McNulty. “Comorbidities and continuities as ontogenic processes: Toward a developmental spectrum model of externalizing psychopathology”. In: *Development and Psychopathology* 25.4 PART 2 (2013), pp. 1505–1528.

[215] A. E. Talley, K. J. Sher, and A. K. Littlefield. “Sexual orientation and substance use trajectories in emerging adulthood”. In: *Addiction* 105.7 (July 2010), pp. 1235–1245.

[216] K. Marynak et al. “Exposure to Electronic Cigarette Advertising Among Middle and High School Students — United States, 2014–2016”. In: *MMWR. Morbidity and Mortality Weekly Report* 67.10 (2018), pp. 294–299.

[217] E. J. D’Amico et al. “Planting the seed for marijuana use: Changes in exposure to medical marijuana advertising and subsequent adolescent marijuana use, cognitions, and consequences over seven years”. In: *Drug and Alcohol Dependence* 188 (2018), pp. 385–391.

[218] N. Chadi, C. Minato, and R. Stanwick. “Cannabis vaping: Understanding the health risks of a rapidly emerging trend”. In: *Paediatrics and Child Health (Canada)* 25.Supplement_1 (2020), S16–S20.

Appendix A: Aim 1

A.1 Search equation

A.1.1 Google Scholar

(e-cigarettes OR vaping) AND (cannabis OR cocaine OR heroin OR opiate OR stimulant OR nonmedical OR amphetamine OR polysubstance) AND longitudinal AND participants

A.1.2 Pubmed (MEDLINE) and PsychINFO

("Electronic Nicotine Delivery Systems"[Mesh]) OR "Vaping"[Mesh]) AND ("Cannabis"[Mesh] OR "Cocaine"[Mesh] OR "Heroin"[Mesh] OR "Opiate Alkaloids"[Mesh] OR "Central Nervous System Stimulants"[Mesh] OR "Amphetamines"[Mesh] OR nonmedical OR polysubstance)

A.2 Covidence extraction form

General information

- Title: [Free text]
- Lead author contact details: [Free text]
- Country in which the study conducted
  - United States
  - UK
  - Canada
  - Australia
  - Other: [Free text]

Characteristics of included studies

- Research question: [Free text]
- Start date: [Free text]
- End date: [Free text]

Participants

- Inclusion criteria: [Free text]
- Exclusion criteria: [Free text]
- Method of recruitment of participants: [Free text]
- Population description (youth/adult, gender, etc.)
- Total number of participants before loss to follow-up

	Operationalization (lifetime, past-year, past-month, etc.)
Exposure 1	
Exposure 2	
Exposure 3	

Table A.1: Table used for extracting exposure information

Exposure

Outcome

	Operationalization
Outcome 1	
Outcome 2	
Outcome 3	
Outcome 4	
Outcome 5	

Table A.2: Table used for extracting outcome information

Prevalent or incident outcomes?: [Prevalent / Incident / Both]

Follow-up duration in years?: [<1 / 1 to < 2 / 2 to < 3 / 3 to < 4 / 4 or more]

Covariates considered

	Covariate	Operationalization	Role (e.g., confounder)
Covariate 1			
Covariate 2			
Covariate 3			
Covariate 4			
Covariate 5			
Covariate 6			
Covariate 7			
Covariate 8			
Covariate 9			
Covariate 10			
Covariate 11			
Covariate 12			
Covariate 13			
Covariate 14			
Covariate 15			

Table A.3: Table used for extracting covariate information

Biases addressed

	How the bias was addressed
Confounding	
Information bias	
Selection bias (internal validity)	

Table A.4: Table used for extracting bias information

Estimation method: [Free text]

Effect modification analysis?: [Yes / No]

If there were any effect modifiers, what were they?: [Free text, NA if none]

What were the results of any effect modification analysis?: [Free text, NA if none]

Mediation analysis?: [Yes / No]

If there were any mediators, what were they?: [Free text, NA if none]

What were the results of any mediation analysis?: [Free text, NA if none]

What were the direct effect results?: [Free text, NA if none]

Brief theoretical explanation of results from discussion: [Free text]

Theoretical explanation of results verbatim from discussion: [Free text]

Study funding sources: [Free text]

Possible conflicts of interest for study authors: [Free text]

Table A.5: First author, title, and year of publication of 17 included studies

No.	First author	Title	Year
1	Cassidy, Rachel N.	Initiation of vaporizing cannabis: Individual and social network predictors in a longitudinal study of young adults	2018
2	Lozano, Paula	A longitudinal study of electronic cigarette use and onset of conventional cigarette smoking and marijuana use among Mexican adolescents	2017
3	Dai, Hongying	Electronic Cigarettes and Future Marijuana Use: A Longitudinal Study	2018
4	Audrain-McGovern, Janet	Adolescent E-Cigarette, Hookah, and Conventional Cigarette Use and Subsequent Marijuana Use	2018
5	Unger, Jennifer B.	E-cigarette use and subsequent cigarette and marijuana use among Hispanic young adults	2016

Continued on next page

Table A.5: First author, title, and year of publication of 17 included studies (cont.)

No.	First author	Title	Year
6	Park, Eunhee	Adolescent E-cigarette use trajectories and subsequent alcohol and marijuana use	2020
7	Wong, Su-Wi	Is E-cigarette use a gateway to marijuana use? Longitudinal examinations of initiation, reinitiation, and persistence of e-cigarette and marijuana use	2020
8	Rogers, Shannon M.	Multidirectional Pathways of Tobacco and Marijuana Use, Including Comorbid Use, Among Young Adults (Aged 18-25 Years) in Texas: A Six-Wave Cross-Lagged Model	2021
9	Pokhrel, Pallav	Predictors of marijuana vaping onset and escalation among young adults	2020
10	Evans-Polce, Rebecca J.	Prospective associations of e-cigarette use with cigarette, alcohol, marijuana, and nonmedical prescription drug use among US adolescents	2020
11	Lee, Juhan	Predictors of vaping marijuana initiation among US adolescents: Results from the Population Assessment of Tobacco and Health (PATH) study Wave 3 (2015-2016)	2021
12	Butler, Alexandra E.	Bidirectional associations between cannabis, e-cigarette, and cigarette use among Canadian youth: findings from the COMPASS Study	2021
13	Bentivegna, Kathryn	Electronic cigarettes associated with incident and polysubstance use among youth	2021
14	Ksinan, Albert J.	E-cigarette use is prospectively associated with initiation of cannabis among college students	2020
15	Ortega, Adrian	Longitudinal investigation of the bidirectional associations between initiation of e-cigarettes and other substances in adolescents	2021

Continued on next page

Table A.5: First author, title, and year of publication of 17 included studies (cont.)

No.	First author	Title	Year
16	Seidel, Ann-Kathrin	Use of electronic cigarettes as a predictor of cannabis experimentation: A longitudinal study among German youth	2021
17	Silveira, Marushka	Longitudinal associations between youth tobacco and substance use in waves 1 and 2 of the Population Assessment of Tobacco and Health (PATH) Study	2018

Table A.6: Assessment of associations in 17 studies included

Study	Exposure[a,b]	Outcome	Measure(s) of association[c,d]	Estimate(s) and (95% CI[e], SE[f], or p-values)	Outcome prevalence
Cassidy	Lifetime ENDS use	Marijuana vaping with ENDS or otherwise since Time 1 (initiation)	OR	2.13 (1.20 -3.79)	34% overall
Lozano	Ever tried e-cigarettes	Past-12-month marijuana use	RR	1.42 (0.84-2.37)	Unknown
Dai	Ever:	Past 12-month:			
	E-cigarette use	Marijuana use	OR	1.9 (1.4-2.5)	7.7% among
	E-cigarette use	Heavy marijuana use	OR	2.5 (1.2-5.3)	unexp.
	No. cartridges used	Marijuana use	OR	1.3 (1.1-1.5)	2.5% among
	No. cartridges used	Heavy marijuana use	OR	1.2 (1.0-1.5)	unexp.
					NA (continuous)
					NA (continuous)
Audrain-McGovern	Ever e-cigarette use	Past 30-day combustible, vaping, edible marijuana use	OR	3.67 (2.51-5.36)	8.7% overall
	Ever e-cigarette use	Ever combustible, vaping, edible marijuana use	OR	3.63 (2.69-4.90)	18.4% overall
Unger	Past-month e-cigarette use	Past-month marijuana use	OR	1.97 (1.01, 3.86)	12.20% among unexposed

Continued on next page

Table A.6: Assessment of associations in 17 studies included (cont.)

Study	Exposure[a,b]	Outcome	Measure(s) of association[c,d]	Estimate(s) and (95% CI[e], SE[f], or p-values)	Outcome prevalence
Park	Past-6-month e-cigarette use at:	Past-6-month marijuana use at:			
	13	13.5	log odds	-0.437 (SE=0.01,	1.4% overall
	13.5	14	log odds	p<0.05)	1.2% overall
	14	14.5	log odds	-0.059 (0.462)	5.3% overall
	14.5	15	log odds	0.141 (0.084)	8.8% overall
	15	15.5	log odds	0.137 (0.093)	12.1% overall
	15.5	16	log odds	0.210 (0.050,	15.8% overall
	16	16.5	log odds	p<0.05)	17.7% overall
	16.5	17	log odds	0.053 (0.044)	17.2% overall
	17	17.5	log odds	0.357 (0.081, p<0.05)	25.3% overall
				0.316 (0.083, p<0.05)	
				-0.030 (0.121)	
Wong	Past-12-month e-cigarette use at:	Wave 3 past-month marijuana:			
	Wave 2	Initiation	OR	2.08 (1.11-3.90)	Not reported
	Wave 2	Reinitiation	OR	1.37 (1.01-1.87)	Not reported
	Wave 2	Sustained	OR	1.19 (0.95-1.42)	Not reported

Continued on next page

Table A.6: Assessment of associations in 17 studies included (cont.)

Study	Exposure[a,b]	Outcome	Measure(s) of association[c,d]	Estimate(s) and (95% CI[e], SE[f], or p-values)	Outcome prevalence
Rogers	Past-30-day ENDS use at:	Past-30-day marijuana use at:			
	Wave 1	Wave 2	OR	Not reported	24.49% overall
	Wave 2	Wave 3	OR	0.87 (0.79-0.95)	Not reported
	Wave 3	Wave 4	OR	1.24 (1.16-1.32)	Not reported
	Wave 4	Wave 5	OR	1.09 (1.05-1.14)	Not reported
	Wave 5	Wave 6	OR	1.20 (1.16-1.25)	Not reported
Pokhrel	Lifetime e-cigarette use (even just a puff)	Lifetime marijuana vaping at follow-up	OR	4.63 (1.98-10.8)	4.2% overall
	Lifetime dual cigarette/e-cigarette use	Lifetime marijuana vaping at follow-up	OR	13.4 (5.37-33.3)	4.2% overall
Evans-Polce	Past-30-day e-cigarette use	Past-30-day marijuana use	OR	2.03 (1.07-3.86)	19.31% overall
	Past-30-day e-cigarette use	Past 30-day nonmedical prescription tranquilizer, sedative, stimulant, or opioid use use	OR	2.57 (1.16-5.71)	4.98% overall
Lee	Current ENDS use	Marijuana vaping initiation	OR	2.16 (1.20-3.89)	6.40% among unexposed
Butler	E-cigarette use (at least monthly in the last 12 months) at Time 1	Cannabis use (at least monthly in the last 12 months) at Time 2	OR	3.09 (2.45-3.91)	8.9% overall

Continued on next page

Table A.6: Assessment of associations in 17 studies included (cont.)

Study	Exposure[a,b]	Outcome	Measure(s) of association[c,d]	Estimate(s) and (95% CI[e], SE[f], or p-values)	Outcome prevalence
Bentivegna	Ever e-cigarette use in Wave 1	Past 12-month use in Wave 2 or Wave 3:			
		Marijuana	OR	2.59 (1.90-3.52)	15.7% overall
		Nonprescribed Ritalin/Adderall	OR	1.89 (1.09-3.28)	2.3% overall
		Polysubstance	OR	2.09 (1.43-3.05)	3.7% overall
		Marijuana vaping	OR	2.26 (1.56-3.27)	7.3% overall
		Nonprescribed painkiller, sedative, or tranquilizers	OR	1.21 (0.79-1.87)	10.9% overall
Ksinan	E-cigarette use at:	Cannabis use at:			
	Year 1	Year 2	log odds	b=0.09, p =0.02	Not reported
	Year 2	Year 3	log odds	b=0.01, p=0.698	Not reported
	Year 3	Year 4	log odds	b=0.09, p =0.031	Not reported
Ortega	Ever use of e-cigarettes at baseline	Ever use of marijuana at year 1	Probit beta	-3.687 (SE=12.54, ≥ 0.05)	6.5% overall
	Ever use of e-cigarettes at baseline	Ever use of marijuana at year 2	Probit beta	-3.577 (62.03, ≥ 0.05)	9.4% overall
Seidel	Lifetime e-cigarette use	Cannabis initiation during the next 18 months	RR	1.83 (1.48-2.25)	10.40% among unexposed
Silveira	Ever e-cigarette use at Wave 1	Past-12-month marijuana use at Wave 2	OR	1.9 (1.6-2.2)	16.6% overall

Continued on next page

Table A.6: Assessment of associations in 17 studies included (cont.)

Study	Exposure[a,b]	Outcome	Measure(s) of association[c,d]	Estimate(s) and (95% CI[e], SE[f], or p-values)	Outcome prevalence
	Ever e-cigarette use at Wave 1	Past-12-month non-prescribed use of Ritalin or Adderall at Wave 2	OR	3.2 (2.3-4.4)	10.5% overall
	Ever e-cigarette use at Wave 1	Past-12-month non-prescribed use painkillers, sedatives, or tranquilizers at Wave 2	OR	1.3 (1.0-1.8)	11.4% overall
	Ever e-cigarette use at Wave 1	Past-12-month other drug use at Wave 2	OR	2.5 (1.7-3.5)	8.6% overall

[a] Past-year and past-12-month time frames are equal. Current, past-month and past-30-day time frames are equal. Ever and lifetime time frames are equal.
[b] ENDS: Electronic nicotine delivery system(s)
[c] OR: Odds ratio
[d] RR: Risk ratio
[e] CI: Confidence interval
[f] SE: Standard error

Table A.7: Assessment of covariate or confounding control in 17 included studies

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Cassidy	None.	Many personal, interpersonal, and contextual factors, such as participants' other substance use, earlier marijuana use, poor school performance, psychopathology, sensation seeking predisposition, impulsivity, friends or family members who use substances, and contextual effects (exposure to advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues).	All likely increase risks of e-cigarette use, initiating marijuana vaping, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimate (OR=2.13, 95%CI: 1.20-3.79) could be more positive than if these factors were controlled. Unmeasured positive confounding possibly explains the association. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana vaping risk would have the opposite impact and are unlikely.,[œ]	2.28 (1.42)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Lozano	Sociodemographic characteristics assessed included age, sex and parental education, which was defined as the highest level reported for either parent (i.e., primary, secondary, high school, university, unknown). Social network smoking behavior included: parent smoker (either vs. none), sibling smoker (any vs. none), smoking among close friends (any vs. none). Personal risk factors included a four-item scale of sensation seeking (i.e., "I like to do frightening things,"); trial of alcohol; binge drinking (more than 3 alcoholic beverages in the last 30 days); trial of drugs (ever use of marijuana, cocaine). Internet tobacco product advertising was queried with a general question that could capture either e-cigarette or	Missing psychopathology and impulsivity, and physical contextual factors (exposure to physical advertisements, ease of access, neighborhood cues).	All likely increase risks of e-cigarette use, past-12-month marijuana use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimate (RR=1.42, 95%CI: 0.87-2.37) could be more positive than if these factors were controlled. Unmeasured positive confounding possibly explains the association, particularly since the lower CI is already below 1. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[d,e]	2.19 (1.00)

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Dai	Age group, gender, race/ethnicity, grade performance, parental education, region, cigarette ever use, alcohol ever use, NMPD[a] ever use, other illicit drug ever use, sensation seeking score.	Missing more proximal influences on participants, such as rules on using tobacco in the home, whether participants live with someone who uses tobacco (as a parent who uses tobacco may live elsewhere). Contextual factors (exposure to advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues) are missing, as are other personal traits (psychopathology, impulsivity).	All likely increase risks of e-cigarette use, past-12-month marijuana use and heavy use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimates (OR=1.9, 95%CI: 1.4-2.5) and (OR=2.5, 95%CI: 1.2-5.3) could be more positive than if these factors were controlled. Unmeasured positive confounding possibly explains the associations, especially the association with non-heavy use. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	3.21 (1.00) 3.26 (1.69)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Audrain-McGovern	Age, sex, race/ethnicity, parental education, family history of combustible cigarette and drug use, peer combustible cigarette use and marijuana use, Temperment and Character Inventory Impulsivity subscale, and Center for Epidemiologic Studies on Depression Scale.	Missing more proximal influences on participants, such as rules on using tobacco in the home, whether participants live with someone who uses tobacco (as a parent who uses tobacco may live elsewhere), participants' own other substance use, and contextual effects (exposure to advertisements, ease of access, neighborhood cues).	All likely increase risks of e-cigarette use, past-30-day and marijuana use and initiation, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimates (OR=3.67, 95%CI: 2.51=5.36), (OR=3.63, 95%CI: 2.69-4.90) could be more positive than if these factors were controlled. Potential unmeasured positive confounders—perhaps the strongest of which are one's other drug and alcohol use—possibly explains the associations. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	6.8 (4.46) 3.22 (2.66)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Unger	Age, sex, other substance use (marijuana, alcohol and other tobacco products).	Missing many personal factors like psychopathology, school grade performance, and psychopathology. Also missing contextual effects (exposure to advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues), as well as family and home life including tobacco rules in the home and living with a tobacco user), and friends and family tobacco and cannabis use.	All likely increase risks of e-cigarette use, past-month marijuana use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimate (OR=1.97, 95%CI: 1.01-3.86) could be more positive than if these factors were controlled. Potential unmeasured positive confounding possibly explains the associations, particularly since the OR's lower CI is nearly 1. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	3.35 (1.11)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Park	Sex, mother's education level, mother-reported household income.	Authors write in terms that they are interested in effects on marijuana use, but earlier marijuana use not controlled. Also many important personal, interpersonal, and contextual factors not controlled for, including age, attitudes towards e-cigarettes and other drugs, other drug use, psychopathology or impulsivity, people close to individual using tobacco or other drugs, school grade performance, exposure to advertisements, ease of access, neighborhood cues.	All likely increase risks of e-cigarette use, past-6-month marijuana use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimates (see Table 2.2) could be more positive than if these factors were controlled. That is, positive and negative associations would be more negative if unmeasured confounders were controlled. In this study, most associations were positive, so attenuation would be expected if assumptions that unmeasured confounders positively relate to each other and all measured variables hold. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.	2.45 1.32 1.15 1.14 1.46 1.18 1.67 1.62 1.14

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Wong	Baseline early onset cigarette use <=18 years old, cigarette smoking frequency (days in past month), early onset of e-cigarette use <=18 years old, locus of control, social network account, perceived life quality, perceived social relationship satisfaction, nicotine product house rule, general opinion to tobacco use, sex, age, race/ethnicity, eduaction, married, employed.	Missing only contextual factors (exposure to advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues), social influences of vaping and using drugs (friends and family), other substance use (heroin, cocaine, etc.), psychopathology, impulsivity, and sensation seeking.	All likely increase risks of e-cigarette use, past-month marijuana use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimates (OR=2.08, 95%CI: 1.11-3.90; OR=1.37, 95%CI: 1.01-1.87; OR=1.19, 95%CI: 0.95-1.42), could be more positive than if these factors were controlled. Potential unmeasured positive confounding possibly explains the associations, particularly since the ORs' lower CIs are close to or cross under 1. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	3.58 (1.46) 2.08 (1.11) 1.67 (1)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Rogers	Age, race/ethniccity, and sex.	Missing critical confounders such as familial/peer tobacco and other drug influences, contextual factors (exposure to advertisements, ease of access, neighborhood cues), and other personal factors (participants' other substance use, earlier e-cigarette and marijuana use, psychopathology, sensation seeking, impulsivity, sexual identity, living situation (roommates vs. home, etc.).	All likely increase risks of e-cigarette use, past-30-day marijuana use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimates (OR=0.87, 95%CI: 0.79-0.95; OR=1.24, 95%CI: 1.16-1.32; OR=1.09, 95%CI: 1.05-1.14; OR=1.20, 95%CI: 1.16-1.25), could be more positive than if these factors were controlled. Potential unmeasured positive confounding possibly explains the associations, particularly since the ORs' lower CIs are close to or cross under 1. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	1.35 (1.19) 1.47 (1.37) 1.26 (1.18) 1.42 (1.37)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Pokhrel	Age, ethnicity, family income, sensation seeking, cigarette smoking history, presence of cigarette, e-cigarette, marijuana users, those who visit vape shops often, and those who know someone who works at a vape shop in social networks, ever exposed to direct marketing through email, text, mail, or social media.	Missing some personal factors, such as marijuana use history, psychopathology, sexual identity, parents' marital status, school performance, and contextual factors (legal environment around e-cigarettes and marijuana).	All likely increase risks of e-cigarette use, marijuana initiation, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimate (OR=4.63, 95%CI: 1.98-10.80), could be more positive than if these factors were controlled. The very large association and appreciable confidence intervals suggests that unmeasured confounding may not attenuate them to the null. More, the outcome prevalence among the unexposed (4.20%) is probably too low for the appreciable OR to overestimate a null RR. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,d,e]	8.73 (3.37) 26.29 (10.21)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Evans-Polce	Past-30-day cigarette use, Past-30-day alcohol use, Past-30-day NMPD[a] use, Gender, Race/ethnicity, Parent education, Region, College plans.	Missing proximal influences on participants, such as rules on using tobacco in the home, whether participants live with someone who uses tobacco (parent, roommate), personal factors, such as sensation seeking, impulsivity, and psychopathology, and contextual effects (exposure to advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues).	All likely increase risks of e-cigarette use, past-30-day marijuana use, nonmedical prescription drug use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimates (OR=2.03, 95%CI: 1.07-3.86; OR=2.57, 95%CI: 1.16-5.71) could be more positive than if these factors were controlled. Potential unmeasured positive confounding possibly explains the associations, particularly since the ORs' lower CI is nearly 1. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	2.2 (1.22) 4.58 (1.59)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Lee	Cigarette smoking, Other tobacco use, Other marijuana product use, alcohol use, other drug use (included Ritalin or Adderall, painkillers, sedatives or tranquilizers, cocaine or crack, stimulants like methamphetamine or speed, any other drugs like heroin, inhalants, solvents or hallucinogens), age, sex, race, Hispanic, parental education level, parental ENDS[h]use, peer ENDS[h]use, internalizing and externalizing tendencies.	Missing proximal influences on participants, such as rules on using tobacco in the home, whether participants live with someone who uses tobacco (i.e., living with a tobacco user), and rules about tobacco use in the home. Also missing are contextual factors (exposure to advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues), peer marijuana vaping, type of marijuana product a respondent used if at all, frequency of marijuana use if any, age of onset if at all, motivation for marijuana vaping, perceptions and attitudes related to marijuana vaping, and personal factors such as school performance and sexual identity.	All likely increase risks of e-cigarette use, marijuana vaping initiation, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimate (OR=2.16, 95%CI: 1.20-3.89) could be more positive than if these factors were controlled. Potential unmeasured positive confounding possibly explains the associations, particularly since the OR's lower CI is close to 1. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	3.74 (1.69)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Butler	School grade, Gender, Self-identified ethnicity, weekly spending money of students	Missing proximal influences on participants, such as rules on using tobacco in the home, whether participants live with someone who uses tobacco, family life (parents' relationship status and living situation), participants' other substance use, psychopathology, impulsivity, sensation seeking, and contextual factors (exposure to advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues).	All likely increase risks of e-cigarette use, monthly past-year marijuana use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimate (OR=3.09, 95%CI: 2.45-3.91) could be more positive than if these factors were controlled. The extensiveness of potential unmeasured positive confounding possibly explains the associations even though the lower CI is very positive. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	5.63 (2.45)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Bentivegna	Age group, Gender, Race, Ethnicity, Grade performance, Parental highest education level, Ever cigarette use, Ever alcohol use, Sensation seeking mean score, feeling trapped, lonely, sad, blue, depressed, or hopeless about the future.	Missing other measures of psychopathology, impulsivity, and contextual/environmental influences, such as living with a tobacco user or rules about tobacco use in the home and friends' substance use. Additionally exposure to other contextual factors (advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues).	All likely increase risks of e-cigarette use, past-12-month marijuana use, nonmedical prescription drug use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimates (OR=2.59, 95%CI: 1.90-3.52; OR=1.89, 95%CI: 1.09-3.28; OR=2.09, 95%CI: 1.43-3.05; OR=2.26, 95%CI: 1.56-3.27) could be more positive than if these factors were controlled. Potential unmeasured positive confounding possibly explains the associations, although the lower limits of most ORs' CIs are appreciably greater than 1, suggesting robustness. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.[c,e]	2.6 (2.1) 3.19 (1.40) 3.60 (2.21) 3.95 (2.49) 1.71 (1.00)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Ksinan	Depressive symptoms at every wave, polysubstance use (of opioid, stimulants, or cocaine: average of past 12-month 0=never, 1=1-5 times in the past 12 months, 2=6+ times in past 12 months), alcohol use (grams of ethanol in past 30 days), sensation seeking at baseline only (3 items), ethnicity, race, age, sex.	Missing proximal influences on participants, such as rules on using tobacco in the home, whether participants live with someone who uses tobacco. Contextual factors (exposure to advertisements, ease of access, neighborhood cues) and personal factors, such as participants' psychopathology and impulsivity, are missing.	All likely increase risks of e-cigarette use, past-6-month marijuana use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported estimates (see Table 2.2) could be more positive than if these factors were controlled. That is, positive and negative associations would be more negative if unmeasured confounders were controlled. In this study, two relevant associations were positive and one was null. Attenuation would be expected if assumptions that unmeasured confounders positively relate to each other and all measured variables hold. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.	1.40 1.11 1.40

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Ortega	None.	Many personal, interpersonal, and contextual factors, such as participants' other substance use, earlier marijuana use, school grade performance, psychopathology, sensation seeking predisposition, impulsivity, relationships with friends or family who use substances, and contextual factors (exposure to advertisements, ease of access, neighborhood cues).	All likely increase risks of e-cigarette use and marijuana use initiation at two future time points, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the negative Probit estimates (-3.687 and -3.577) could be more positive than if these factors were controlled. In terms of the probit results, the probability of observing marijuana initiation would be even lower if the unmeasured confounders were controlled. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) marijuana use risk would have the opposite impact and are unlikely.	N/A

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Seidel	Gender, age, school type, migrant status, sensation seeking, peer cannabis use, conventional cigarette use, and alcohol use.	Authors state there is likely unmeasured confounding by socioeconomic position of parents, other substance use among parents or friends or other psychological variables like depression, anxiety or impulsivity. There is also likely unmeasured confounding by other contextual factors (exposure to advertisements, ease of access, neighborhood cues), other personal factors (like school performance, sexual identity, and externalizing psychopathology).	All likely increase risks of e-cigarette use, past-12-month marijuana use, nonmedical prescription drug use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported main association (RR=1.83, 95%CI: 1.48-2.25) could be more positive than if these factors were controlled. The other reported associations were negative because they were estimates of interactions with less risky factors (low sensation seeking, RR=0.77, 95%CI: 0.61-0.97; no cigarette use, RR=0.48, 95%CI: 0.37-0.64). If these were expressed as interactions with more risky factors, the associations would be positive, and the same bias as the main association would apply. Potential unmeasured positive	

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
			confounding possibly explains the interaction between cannabis initiation and the interaction between e-cigarette use and low sensation seeking, but the other associations appear more robust. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) cannabis initiation risk would have the opposite impact and are unlikely.[d,e]	3.06 (2.32)

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
Silveira	Age, gender, race/ethnicity, sensation seeking, past-year mental health problems, and Wave 1 ever tobacco or substance use in the models predicting Wave 2 tobacco use and substance use.	Missing measures of psychopathology and contextual/environmental influences, such as living with a tobacco user or rules about tobacco use in the home. Additionally exposure to other contextual factors (advertisements, legal environment around e-cigarettes and marijuana, ease of access, neighborhood cues).	All likely increase risks of e-cigarette use, past-12-month marijuana use, nonmedical prescription drug use, and other prohibited substance use, and likely are positively associated with each other. This potential unmeasured (positive) confounding means that the reported main association (OR=1.9, 95%CI: 1.6-2.2; OR=3.2, 95%CI: 2.3-4.4; OR=1.3, 95%CI: 1.0-1.8; OR=2.5, 95%CI: 1.7-3.5) could be more positive than if these factors were controlled. Potential unmeasured positive confounding possibly explains the associations between e-cigarette use and cannabis use, nonprescribed use of painkillers,	

Continued on next page

Table A.7: Assessment of confounding control in 17 included studies (cont.)

Study	Covariates controlled	Potential unmeasured confounders	Could potential unmeasured confounding explain association?	E-value (lower CI)[e]
			sedative, or tranquilizers, use of cocaine or crack, or other stimulants (methamphetamine or speed), heroin, inhalants, solvents, or hallucinogens. Unmeasured confounders that decrease (increase) e-cigarette use risk and increase (decrease) cannabis initiation risk would have the opposite impact and are unlikely.[c,e]	2.10 (1.84) 5.85 (4.03) 1.92 (1.00) 4.44 (2.79)

[a] NMPD: Nonmedical prescription drug
[b] ENDS: Electronic nicotine delivery system(s)
[c] OR: Odds ratio
[d] RR: Risk ratio
[e] CI: Confidence interval

Appendix B: Aims 2 and 3

Table B.1: Dates and numbers of interviews from data collection waves from PATH restricted data files user guide

	Data collection		Interviews conducted		
Wave	**Start date**	**End date**	**Adults**	**Youth**	**Parents of youth**
1	September 12, 2013	December 14, 2014	32,320	13,651	13,588
2	October 23, 2014	October 30, 2015	28,362	12,172	12,129
3	October 19, 2015	October 23, 2016	28,148	11,814	11,807
4	December 1, 2016	January 3, 2018	33,822	14,798	14,709
5	December 1, 2018	November 30, 2019	34,309	12,098	12,027

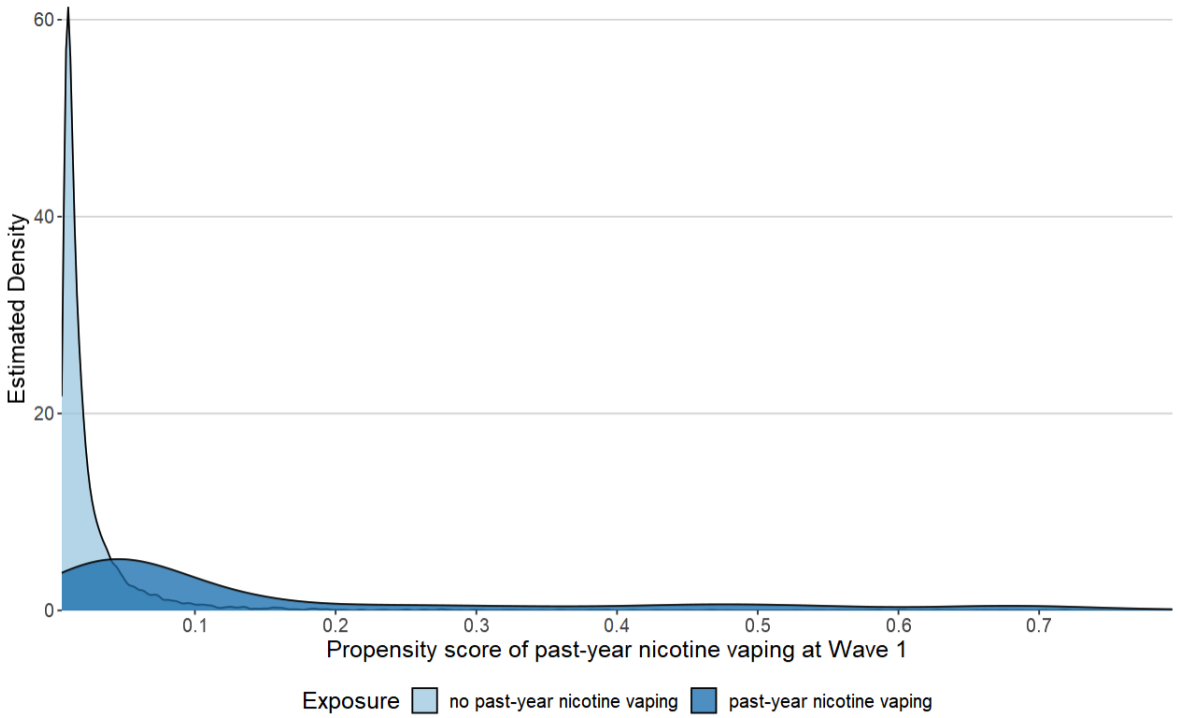

Figure B.1: Plot showing little overlap of conditional probabilities of nicotine vaping (i.e., propensity scores) for observed nicotine vaping and non-vaping groups (i.e., positivity violation) at Wave 1 of PATH, 9,572 baseline participants aged 12 to 16 years old

B.1 Approach to confounding

Consistent with earlier literature that considered vaping and other substance use to be risky behaviors,[12,70] we used problem behavior theory,[46,195–198] common liability theory,[199,200] and interactive theory of substance use as conceptual frameworks to motivate covariate selection.[201] We used Zingberg's drug, set, and setting framework to consider how each construct plays a role within the context of other constructs, the study population, and the exposure and outcomes.[115]

Briefly, problem behavior theory holds that behavioral outcomes, such as substance use, result from three levels of psychosocial influence: demography-social structure and socialization, personality system, perceived environment system.[46,195–197] Demography-social structure and socialization is the most causally distal system and refers to parents' education and the family structure, as well as media influence, friends' interests, maternal controls, and parents' religiosity and beliefs. These factors may influence problem or positive behaviors through more proximal influences, the personality and perceived environment systems.[202] The personality system concerns sets of personal controls and beliefs that may exert influences on problem behaviors.

In the studies in Chapters 3 and 4, the following personal control influences were considered as measured confounders: value on academic achievement, affect, and locus of control. The perceived environment system refers to supports, influence, controls, models, and expectations of others that individuals can perceive and to which they may ascribe meaning.[202] Parental

support and controls, friends' support and controls, parental and friends' influences, and parents' and friends' approval of problem behavior were considered as potential measured confounders. A prior study using PATH adult participant data invoked only problem behavior theory.[70] That study used the variables age of cigarette use onset, cigarette use frequency, age of alcohol use onset, and alcohol use frequency to measure the behavioral system in problem behavior theory. Personality system variables included locus of control, perceived life quality, and perceived social relationship satisfaction. Perceived environmental system variables included rules about nicotine in the house and general opinion toward tobacco use. The variables that the authors controlled in that study were controlled in the present study.

Common liability theory holds that individuals may be genetically predisposed to all forms of substance use and addiction (i.e., common causes of all variations).[199,200] Genetic variables were not available, but baseline behavioral disinhibition was. This construct was captured using self-reports from some questions in the GAIN-SS externalizing subscale.[203–205] Parent-reported attention-deficit/hyperactivity disorder diagnosis by a physician, which common liability theorists contend is the manifestation of transmissible risk for substance use and addiction,[114,199,200,206–213] also captured the behavioral disinhibition construct. Neurodevelopmental science supports the hypothesis that a trait characteristic of behavioral disinhibition is transmitted from parents to their children,[214] and may increase the risk for general substance use behaviors.

A third framework, interactive theory, contends that large constellations of personal and exogenous forces influence lifestyle behaviors, including substance use, and that earlier theories had ignored some of these forces.[50] The four major constellations of forces are: biological, intrapersonal, interpersonal, and sociocultural,[50] which largely overlap with problem behavior theory and common liability theory. For instance, at intrapersonal level, we may be concerned with the personality system in problem behavior theory, and at the biological level, we may be concerned with genetic effects in common liability theory. Overall health (biological), socioeconomic resources (intrapersonal), and substance availability (sociocultural) are contained in interactive theory but not other frameworks;[50] therefore, these constructs were included as confounders in the present study.

These three major conceptual frameworks motivated which covariates were selected from the PATH data set to serve as confounders in each analysis; however, these frameworks are agnostic to context in which each covariate construct may play a role. Zinberg's drug, set, and setting is a conceptual model that allows for flexibility for how each of the substantive constructs may work, and therefore, how each may be considered in separate analyses.[115] "Drug" describes attributes of the substance (e.g., route of administration, pharmacological action, potency, feelings/highness); "set" describes the attitude of the person at the time of use; "setting" describes the influence of the physical and social setting within which use occurs.[115] In Chapter 3, nicotine vaping was posited to affect cannabis initiation. A main reason for substance use is coping with stress, and we controlled indicators of internalizing psychopathology to measure these feelings. Some of the most vulnerable to increased stress during the sensitive transitional period are members of the LGBTQ+ community. We therefore controlled for sexual orientation hoping that responses would tap into identity-related stress.[215] Exposure to vaping advertising was considered a "setting" because the proliferation and normalization of vaping in media,

storefronts, and elsewhere may influence whether individuals vape nicotine and other products, such as cannabis;[216–218] therefore, measured variables, such as "In past 30 days, has noticed e-cigarettes or other electronic nicotine products being advertised" were controlled in the analysis. This variable may capture how normatively individuals perceive vaping proliferation in their physical setting.

Thus, the general process for covariate control was to apply three theories mentioned earlier as a starting point for which variables may confound the analyses, and then Zinberg's drug, set, and setting framework to think through other potentially contextually relevant confounders. Unmeasured confounding was considered in a sensitivity analysis.

Table B.2: List of confounders selected for incremental propensity score intervention and stochastic interventional in(direct) effects analyses

Variable	Operationalization
Age[a]	12 = 1 13 = 2 14 = 3 15 = 4 16 = 5
Parent or guardian: marital status (Wave 1-Wave 4) / Own marital status (Wave 3-Wave 4)	0. Not currently married 1. Currently married
Race[b]	0.White race alone 1. Black race alone 2. Asian race alone 3. Other race, including multiracial
In the past 12 months, total household income	0. $49,000 or less 1. $50,000 or more
Highest grade of year of school completed by parent	0. Up to some college but no degree 1. At least an associate's degree
Like new and exciting experiences, even if I have to break the rules	1. Agree or strongly agree 0. Neither agree nor disagree, disagree, or strongly disagree

Continued on next page

Table B.2: List of confounders selected for incremental propensity score intervention analyses

Variable	Operationalization
Like to do frightening things	1. Agree or strongly agree 0. Neither agree nor disagree, disagree, or strongly disagree
Prefer friends who are exciting and unpredictable	1. Agree or strongly agree 0. Neither agree nor disagree, disagree, or strongly disagree
Last time you had significant problems with: Feeling very trapped, lonely, sad, blue, depressed or hopeless about the future	0. Not in the past month 1. In the past month
Last time you had significant problems with: 1. Sleep trouble - such as bad dreams, sleeping restlessly or falling asleep during the day 2. Feeling anxious, nervous, tense, scared, panicked or like something bad was going to happen 3. Becoming very distressed and upset when something reminded you of the past	0. Not in the past 12 months 1. In the past 12 months

Continued on next page

Table B.2: List of confounders selected for incremental propensity score intervention analyses

Variable	Operationalization
Last time you did the following two or more times 1. Lied or conned to get things you wanted or avoid having to do something 2. Had a hard time paying attention at school, work, or home 3. Had a hard time listening to instructions at school, work, or home 4. Were a bully or threatened other people 5. Gave answers before the other person finished asking the question 6. Started physical fights with other people 7. Felt restless or the need to run around or climb on things	0. Not in the past 12 months 1. In the past 12 months
Youth's school grade performance	0. Mostly B's or worse 1. Mostly A's or mostly A's and B's
Tobacco use in the past 30 days (cigarettes, cigars, cigarillos, filtered cigars, pipes, hookah, smokeless tobacco)	0. No 1. Yes
Sexual orientation	0. Straight 1. Not straight
Youth's overall health status according to parent	0. Very good or excellent 1. Good, fair, or poor
Someone other than youth uses tobacco in the home	0. No 1. Yes

Continued on next page

Table B.2: List of confounders selected for incremental propensity score intervention analyses

Variable	Operationalization
Rules about using tobacco products that are burned or not burned inside the home	0. Not allowed at any time inside my home 1. Allowed at some times or any time inside my home
Anyone who lives with you now uses any form of tobacco	0. No 1. Yes
Would use an e-cigarette if one of your best friends offered you one	0. Probably or definitely not 1. Probably or definitely yes
Ever used alcohol at all (Wave 1) / Used alcohol in the past 12 months (Wave 2-Wave 4)	0. No 1. Yes
Youth has another parent who lives somewhere else / how often youth stays with other parent who lives somewhere else	0. Yes, never 1. Yes, less than or about half the time 2. No
Youth has ever been told by a doctor or other health professional that he/she has ADHD or ADD	0. No 1. Yes
Youth has a curfew or set time to be home on school nights or weekends	0. No 1. Yes
Reaction if parent/guardian found you using tobacco	0. Be very upset 1. Not be too upset or have no reaction
Parent or guardian talk with you, even once, about not using any type of tobacco in the past 12 months	0. No 1. Yes
Used in the past 12 months, when prohibited substance initiation was the outcome: Marijuana, hash, THC, or grass	0. No 1. Yes

Continued on next page

Table B.2: List of confounders selected for incremental propensity score intervention analyses

Variable	Operationalization
Used in the past 12 months, when cannabis initiation was the outcome: 1. Prescription drug not prescribed to you: Ritalin or Adderall 2. Prescription drug not prescribed to you: Painkillers, sedatives or tranquilizers 3. Cocaine or crack 4. Stimulants like methamphetamine or speed 5. Any other drugs like heroin, inhalants, solvents or hallucinogens	0. No 1. Yes
In the past 30 days, how often noticed tobacco ads when visiting convenience stores, small markets, or liquor stores (measured from Wave 2-Wave 4) or In the past 30 days, has noticed e-cigarettes being advertised (measured from Wave 2-Wave 4): 1. On posters or billboards 2. In newspapers or magazines 3. On websites or social media sites 4. On radio, on television	0. Never or No 1. Rarely, sometimes or often or Yes

[a] Due to small cell counts when stratifying, I grouped 12-, 13-, and 14 year-olds together and 15- and 16-year-olds together in Aim 3.

[b] Due to small cell counts when stratifying, I grouped Black, Asian, and Other race categories together in Aim 3.

B.2 Revised incremental propensity score interventions code (based on Kennedy's npcausal::ipsi)

```
ipsi <- function(y, a, x.trt, x.out, time, age, svywt, id, delta.seq,
                 nsplits = 5, ci_level = 0.95,
                 progress_bar = TRUE, return_ifvals = FALSE,
                 fit = "rf", estimand,
                 sl.lib = c("SL.mean", "SL.earth", "SL.glm")) {

  require("SuperLearner")
  require("earth")
  require("glmnet")
  require("xgboost")

  # setup storage
  ntimes <- length(table(time))
  end <- max(time)
  start <- min(time)
  n <- length(unique(id))
  dat <- data.frame(time = time, age = age, id = id, y = y, a = a, svywt =
  svywt)
  k <- length(delta.seq)
  ifvals <- matrix(nrow = n, ncol = k)
  risk <- rep(NA, k)
  wt <- matrix(nrow = n * ntimes, ncol = k)
  cumwt <- matrix(nrow = n * ntimes, ncol = k)
  rt <- matrix(nrow = n * ntimes, ncol = k)
  vt <- matrix(nrow = n * ntimes, ncol = k)
  x.trt <- data.frame(x.trt)
  x.out <- data.frame(x.out); x.out$a <- a

  # bounding and trimming functions
  bound <- function(x, min = 0.00001, max = 0.999999){
    ifelse(x < lower, lower, ifelse(x > upper, upper))
  }
  trim <- function(x, lower = quartile(0.25), upper = quartile(0.75)){
    ifelse(x < lower, lower, ifelse(x > upper, upper))
  }

  if (progress_bar) {
    pb <- txtProgressBar(
      min = 0, max = 2 * nsplits * length(delta.seq) + 3,
      style = 3
    )
  }
```

```
s <- sample(rep(seq_len(nsplits), ceiling(n / nsplits))[seq_len(n)])
slong <- rep(s, rep(ntimes, n))

if (progress_bar) {
  pbcount <- 0
}

for (split in seq_len(nsplits)) {
  if (progress_bar) {
    Sys.sleep(0.1)
    setTxtProgressBar(pb, pbcount)
    pbcount <- pbcount + 1
  }

  # fit treatment model
  if (fit=="sl"){
    trtmod <- SuperLearner(dat$a[slong!=split], x.trt[slong!=split,],
                           newX = x.trt, SL.library = sl.lib, family =
                           binomial)
    dat$ps <- bound(trtmod$SL.predict)
    pi.t <- dat$ps
  }

  for (j in seq_len(k)) {
    if (progress_bar) {
      Sys.sleep(0.1)
      setTxtProgressBar(pb, pbcount)
      pbcount <- pbcount + 1
    }
    delta <- delta.seq[j]

    # compute weights
    wt[, j] <- (delta * dat$a + 1 - dat$a) / (delta * dat$ps + 1 - dat$ps)
    cumwt[, j] <- as.numeric(t(aggregate(wt[, j],
                                         by = list(dat$id),
                                         cumprod
    )[, -1]))
    vt[, j] <- (1 - delta) * (dat$a * (1 - dat$ps) -
                                (1 - dat$a) * delta * dat$ps) / delta

    # fit outcome models
    outmod <- vector("list", ntimes)
    m1 <- vector("list", ntimes)
    m0 <- vector("list", ntimes)
```

```
mA <- vector("list", ntimes)
rtp1 <- dat$y[dat$time == end]
if (progress_bar) {
  Sys.sleep(0.1)
  setTxtProgressBar(pb, pbcount)
  pbcount <- pbcount + 1
}
for (i in seq_len(ntimes)) {
  t <- rev(unique(dat$time))[i]

  if(estimand=="ATE") {
  # counterfactual case for treatment: A = 1
  newx1 <- x.out[dat$time == t, ]
  newx1$a <- 1

  # counterfactual case for no treatment: A = 0
  newx0 <- x.out[dat$time == t, ]
  newx0$a <- 0

  if (fit=="sl"){
    print(c(i,j)); flush.console()
    outmod[[i]] <- SuperLearner(rtp1[s!=split],
                                x.out[dat$time==t &
                                slong!=split,],SL.library = sl.lib,
                                newX=rbind(newx1,newx0))

    # Predicted outcomes for ATE
    m1 <- ifelse(dat$y[dat$time == t] == 1 & dat$a[dat$time == t] ==
    1, 1, outmod[[i]]$SL.predict[1:dim(newx1)[1]])
    m0 <- ifelse(dat$y[dat$time == t] == 1 & dat$a[dat$time == t] ==
    0, 1,
    outmod[[i]]$SL.predict[(dim(newx1)[1]+1):(dim(newx1)[1]+dim(newx0)[1])])

    dat$m1[dat$time == t] <- bound(m1[[t]])
    dat$m0[dat$time == t] <- bound(m0[[t]])

    # ATE pseudo-outcome
    pi.t[dat$time == t] <- dat$ps[dat$time == t]
    rtp1 <- (delta * pi.t[dat$time == t] * dat$m1[dat$time == t] + (1
   - pi.t[dat$time == t]) * dat$m0[dat$time == t]) /
      (delta * pi.t[dat$time == t] + 1 - pi.t[dat$time == t])
    rt[dat$time == t, j] <- rtp1
  }
  }
```

```
    # compute influence function values
    ifvals[s == split, j] <- ((cumwt[, j] * dat$y)[dat$time == end] +
                                aggregate(cumwt[, j] * vt[, j] * rt[, j] *
                                svywt,
                                by = list(dat$id), sum)[, -1])[s == split]
  }
}

# compute estimator
for (j in seq_len(k)) {
  risk[j] <- mean(ifvals[, j])
}

ifvals.dif <- ifvals[,10] - ifvals[-10]
risk.dif <- apply(ifvals.dif, 2, mean)

# compute asymptotic variance
se <- sqrt(apply(ifvals, 2, var)) / sqrt(n)
ci_norm_bounds <- abs(stats::qnorm(p = (1 - ci_level) / 2))
eff.ll <- risk - ci_norm_bounds * se
eff.ul <- risk + ci_norm_bounds * se

se.dif <- sqrt(apply(ifvals.dif, 2, var)) / sqrt(n)
eff.ll.dif <- risk.dif - ci_norm_bounds * se.dif
eff.ul.dif <- risk.dif + ci_norm_bounds * se.dif

# multiplier bootstrap
if (progress_bar) {
  Sys.sleep(0.1)
  setTxtProgressBar(pb, pbcount)
  pbcount <- pbcount + 1
}

res <- data.frame(
  increment = delta.seq, risk = risk, se = se,
  est.ll = est.ll, est.ul = est.ul
)
res.dif <- data.frame(
  increment = delta.seq, risk.dif = risk.dif, se.dif = se.dif,
  est.ll.dif = est.ll.dif, est.ul.dif = est.ul.dif
)

# output
return(print(list(
  res = res, res.dif = res.dif, ifvals.dif = ifvals.dif, dat = dat)))
```

```
}
```

Table B.3: Aim 2 missing exposure data

Wave	Positive	Negative	Missing	Proportion missing
1	302	9,218	51	0.5%
2	587	7,252	1,732	18.1%
3	914	6,640	2,017	21.1%
4	853	6,063	2,655	27.7%

Table B.4: Aim 2 missing outcome data (cannabis initiation)

Wave	Positive	Negative	Missing	Proportion missing
2	620	7,666	1,285	13.4%
3	954	7,103	1,514	15.8%
4	1,490	6,133	1,948	20.4%
5	2,084	4,873	2,614	27.3%

Table B.5: Aim 2 Missing outcome data (prohibited substance initiation)

Wave	Positive	Negative	Missing	Proportion missing
2	578	7,688	1,305	13.6%
3	786	7,192	1,593	16.6%
4	722	6,887	1,962	20.5%
5	574	6,355	2,642	27.6%

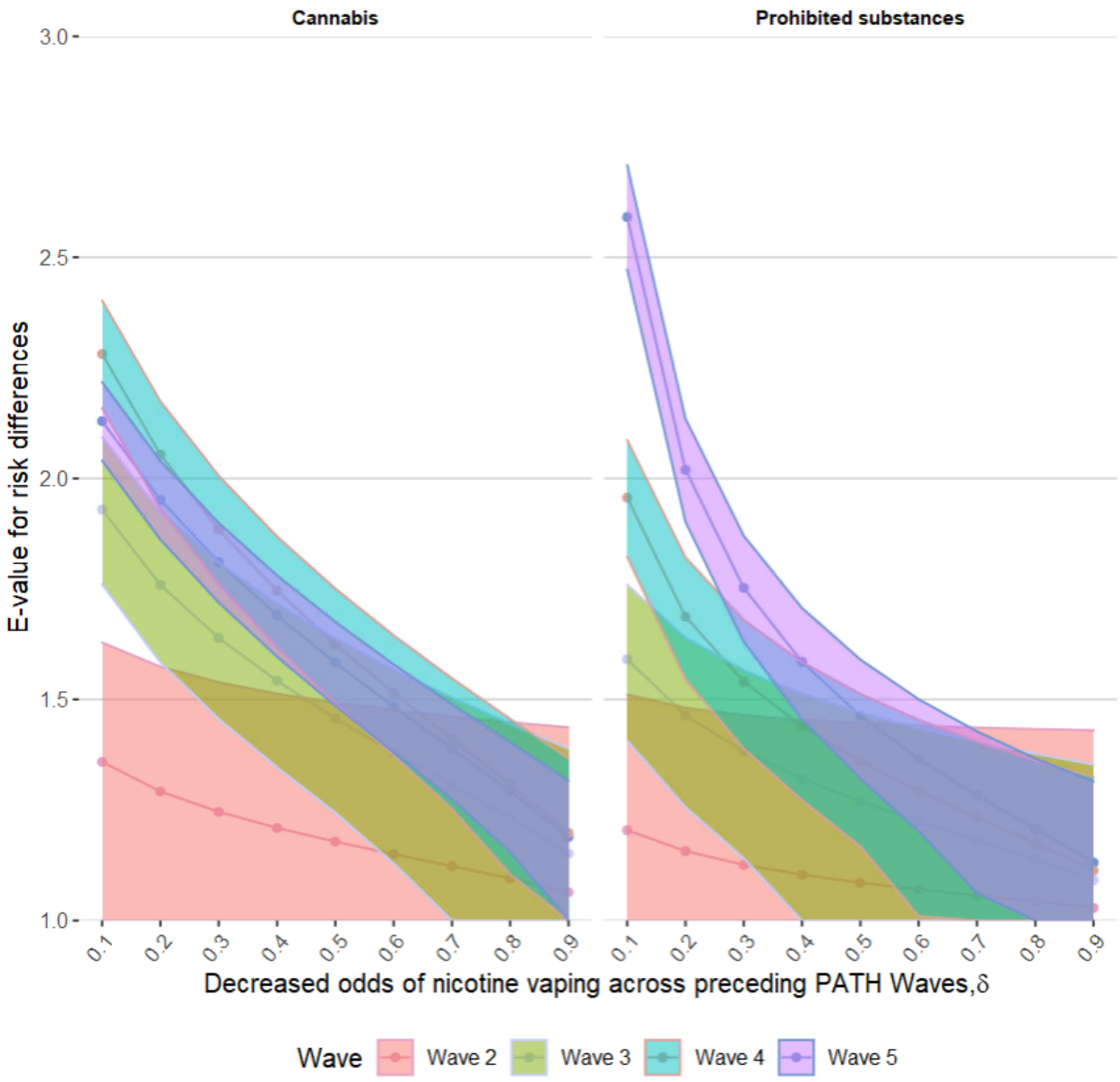

Figure B.2: E-values for risk differences from Figure 3.4 by each follow-up Wave

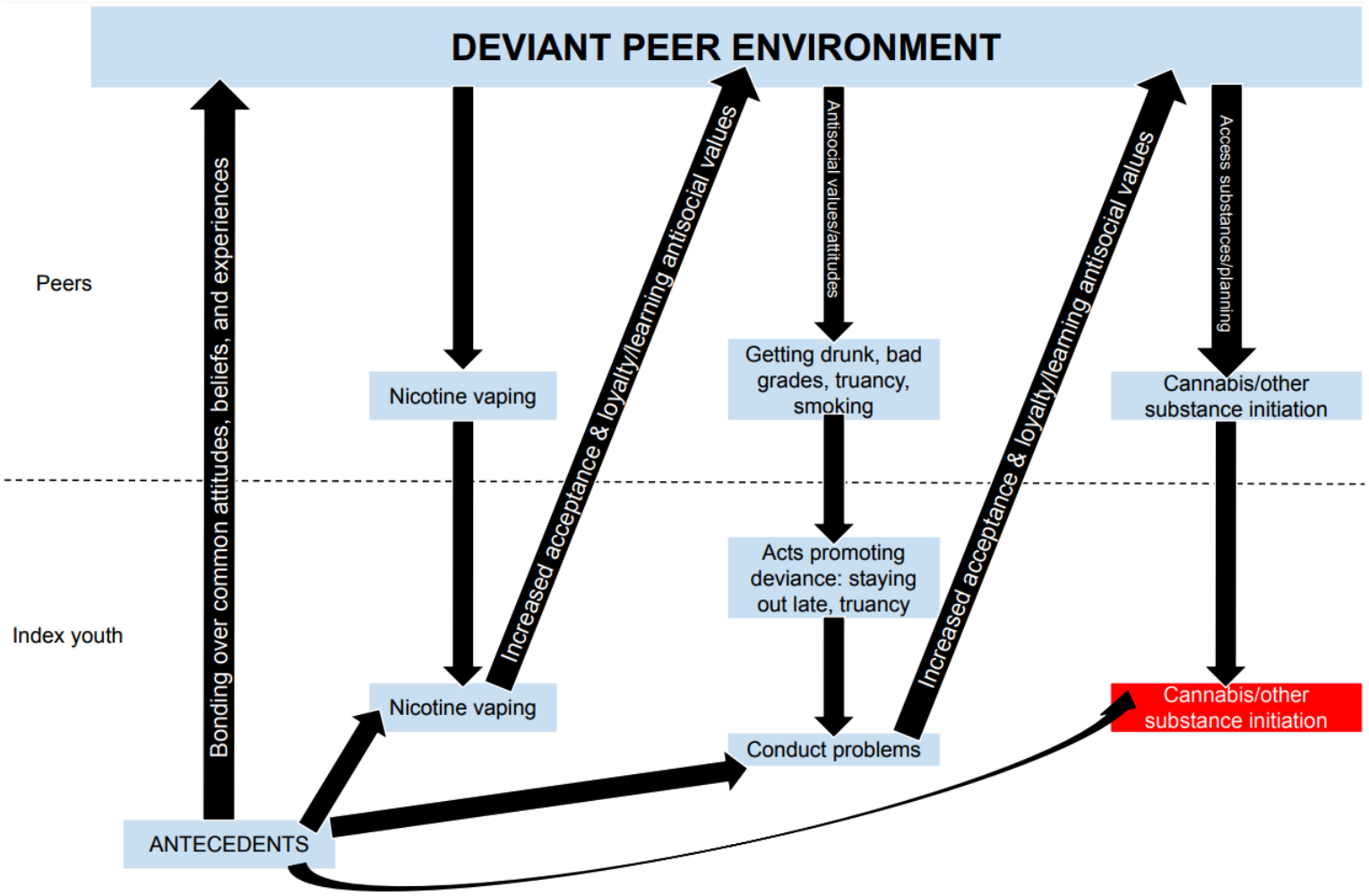

Figure B.3: Conceptual model showing nicotine vaping's posited indirect effects on cannabis initiation and other substance initiation through its effects on conduct problems within the context of a deviant peer environment

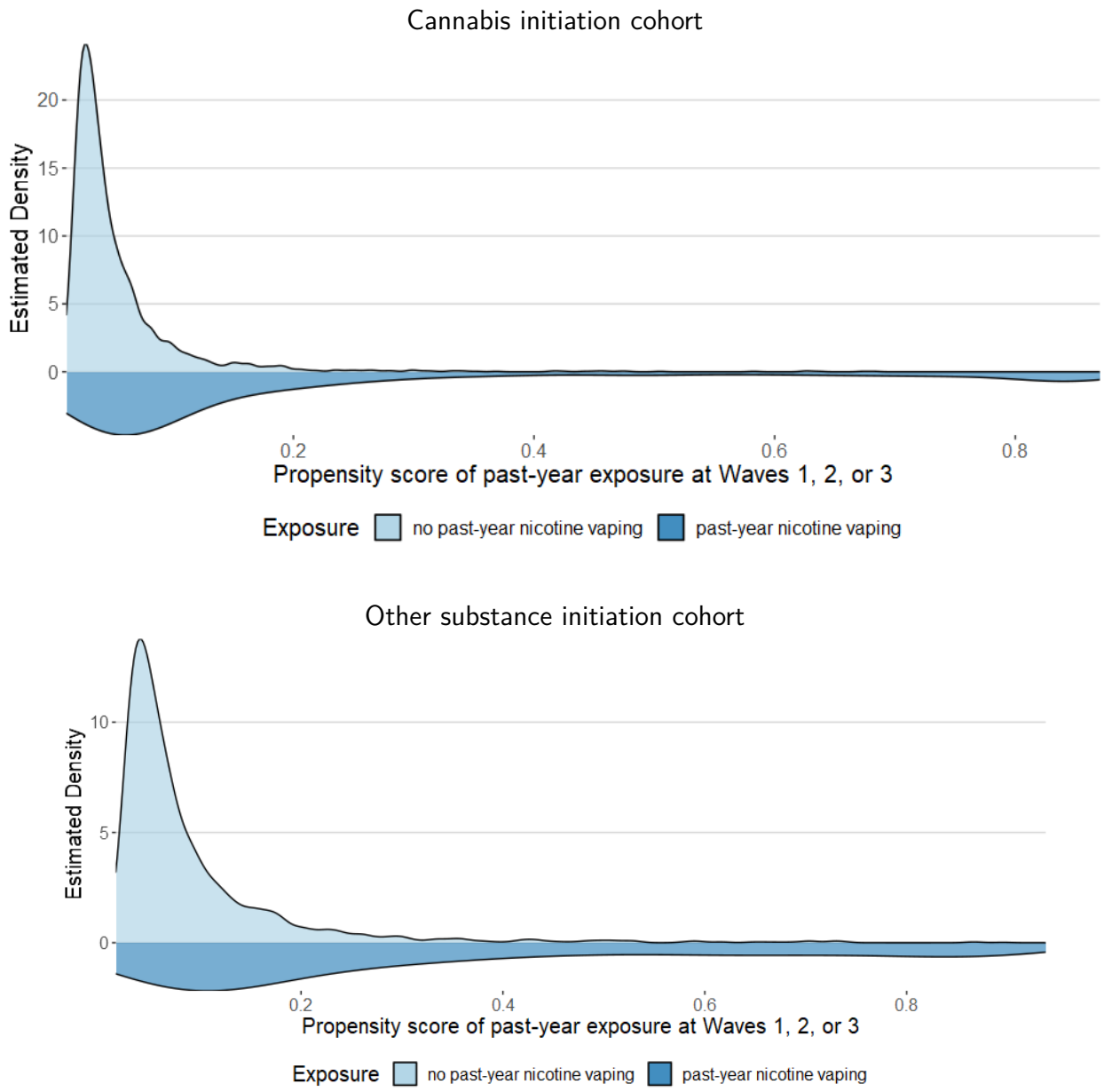

Figure B.4: Plot showing little overlap of conditional probabilities of nicotine vaping (i.e., propensity scores) for observed nicotine vaping and non-vaping groups (i.e., positivity violation) at Waves 1, 2, or 3 of PATH, 2,949 and 2,995 baseline participants aged 12 to 16 years old in cannabis initiation cohort (top) and other substance initiation cohort (bottom), respectively

Table B.6: Aim 3 missing data among 2,949 participants who never reported conduct problems before Wave 4 and who never reported cannabis use before Wave 5

Variable (all past-year)	Positive	Negative	Missing	Proportion missing
Wave 1 or 2 or 3 nicotine vaping	112	1,922	915	31.0%
Wave 4 conduct problems				
Lying or conning	350	1,717	882	29.9%
Bullying perpetration	52	2,017	880	29.8%
Starting fights	39	2,031	879	29.8%
Wave 5 cannabis initiation	283	1,589	1,077	36.5%

Table B.7: Aim 3 missing data among 2,995 baseline participants aged 12 to 16 years old who never reported conduct problems before Wave 4 and who never reported prohibited substance use before Wave 5

Variable (all past-year)	Positive	Negative	Missing	Proportion missing
Wave 1 or 2 or 3 nicotine vaping	226	1,817	952	31.8%
Wave 4 conduct problems				
Lying or conning	373	1,728	894	29.8%
Bullying perpetration	51	2,051	893	29.8%
Starting fights	48	2,055	892	29.8%
Wave 5 prohibited substance initiation	84	1,805	1,106	36.9%

B.3 Stochastic direct and indirect effects code (based on Hejazi's pub_medshift_interv_biostats::example)

```
# estimate nuisance parameters
fitnuisance <- function(data, nfolds,
                        sl_lib = c("SL.glm", "SL.mean")) {
  ## set SL candidates
  candidatesg <- candidatese <- candidatesm <- candidatesb <-
    candidatesd <- sl_lib

  ## extract data
  #Y <- as.numeric(data[["Y5"]])
  Y <- as.numeric(data[["Y4"]])
  #Z <- as.data.frame(data[,c(99,102,103)])
  Z <- as.data.frame(data[,c(69,72,73)])
  #L <- as.numeric(data[["L3"]])
  L <- as.numeric(data[["L2"]])
```

```r
#A <- as.numeric(data[["APY123"]])
A <- as.numeric(data[["APY12"]])
W <- as.data.frame(data[,c(1:31)])

## observations and cross-validation
n <- length(A)
valSets <- split(sample(seq_len(n)), rep(seq_len(nfolds), length = n))

## fit nuisance functions
fitg <- mySL(
  Y = A,
  X = W,
  family = stats::binomial(),
  SL.library = candidatesg,
  validRows = valSets
)
fite <- mySL(
  Y = A,
  X = data.frame(Z, W),
  family = stats::binomial(),
  SL.library = candidatese,
  validRows = valSets
)
fitm <- mySL(
  Y = Y,
  X = data.frame(Z, L, A, W),
  family = binomial(),
  SL.library = candidatesm,
  validRows = valSets
)
fitb <- mySL(
  Y = L,
  X = data.frame(A, W),
  family = stats::binomial(),
  SL.library = candidatesb,
  validRows = valSets
)
fitd <- mySL(
  Y = L,
  X = data.frame(Z, A, W),
  family = stats::binomial(),
  SL.library = candidatesd,
  validRows = valSets
)
```

```
  return(list(fitg = fitg, fite = fite, fitm = fitm, fitb = fitb,
              fitd = fitd, valSets = valSets))
}


# construct estimators of direct and indirect effects
estimators <- function(data, delta, nuisance,
                       sl_lib = c("SL.myglm", "SL.mean")) {
  ## set SL candidates
  candidatesv  <- candidatess <- candidatesq <- candidatesubar <- sl_lib

  ## extract data
  #Y <- as.numeric(data[["Y5"]])
  Y <- as.numeric(data[["Y4"]])
  #Z <- as.data.frame(data[,c(99,102,103)])
  Z <- as.data.frame(data[,c(69,72,73)])
  #L <- as.numeric(data[["L3"]])
  L <- as.numeric(data[["L2"]])
  #A <- as.numeric(data[["APY123"]])
  A <- as.numeric(data[["APY12"]])
  W <- as.data.frame(data[,c(1:31)])

  g1  <- truncate(nuisance$fitg$SL.predict[, 1])
  e1  <- truncate(nuisance$fite$SL.predict[, 1])
  b1A <- nuisance$fitb$SL.predict[, 1]
  d1A <- truncate(nuisance$fitd$SL.predict[, 1])

  g <- A * g1 + (1 - A) * (1 - g1)
  e <- A * e1 + (1 - A) * (1 - e1)
  b <- L * b1A + (1 - L) * (1 - b1A)
  d <- L * d1A + (1 - L) * (1 - d1A)

  b11  <- predict(nuisance$fitb, newdata = data.frame(A = 1, W))$pred[, 1]
  b10  <- predict(nuisance$fitb, newdata = data.frame(A = 0, W))$pred[, 1]
  d11  <- truncate(predict(nuisance$fitd,
                           newdata = data.frame(Z, A = 1, W))$pred[, 1])
  d10  <- truncate(predict(nuisance$fitd,
                           newdata = data.frame(Z, A = 0, W))$pred[, 1])
  m11  <- predict(nuisance$fitm,
                  newdata = data.frame(Z, L = 1, A = 1, W))$pred[, 1]
  m10  <- predict(nuisance$fitm,
                  newdata = data.frame(Z, L = 1, A = 0, W))$pred[, 1]
  m01  <- predict(nuisance$fitm,
                  newdata = data.frame(Z, L = 0, A = 1, W))$pred[, 1]
  m00  <- predict(nuisance$fitm,
```

```
                newdata = data.frame(Z, L = 0, A = 0, W))$pred[, 1]

m <- nuisance$fitm$SL.predict[, 1]
m1A <- A * m11 + (1 - A) * m10
m0A <- A * m01 + (1 - A) * m00

uA <- m1A * b1A + m0A * (1 - b1A)
u1 <- m11 * b11 + m01 * (1 - b11)
u0 <- m10 * b10 + m00 * (1 - b10)

vout <- m * b / d
sout <- m * b / d * g / e

if (sd(vout) < .Machine$double.eps * 10) {
  v1 <- v0 <- rep(mean(vout), length(vout))
} else {
  # the call below breaks
  fitv <- mySL(
    Y = vout,
    X = data.frame(L, A, W),
    family = stats::gaussian(),
    SL.library = candidatesv,
    validRows = nuisance$valSets
  )
  v1 <- predict(fitv, newdata = data.frame(L = 1, A, W))$pred[, 1]
  v0 <- predict(fitv, newdata = data.frame(L = 0, A, W))$pred[, 1]
}

if (sd(sout) < .Machine$double.eps * 10) {
  s1 <- s0 <- rep(mean(sout), length(sout))
} else {
  fits <- mySL(
    Y = sout,
    X = data.frame(L, A, W),
    family = stats::gaussian(),
    SL.library = candidatess,
    validRows = nuisance$valSets
  )
  s1 <- predict(fits, newdata = data.frame(L = 1, A, W))$pred[, 1]
  s0 <- predict(fits, newdata = data.frame(L = 0, A, W))$pred[, 1]
}

if (sd(uA) < .Machine$double.eps * 10) {
  ubar1 <- ubar0 <- rep(mean(uA), length(uA))
} else {
```

```
  fitubar <- mySL(
    Y = uA,
    X = data.frame(A, W),
    family = stats::gaussian(),
    SL.library = candidatesubar,
    validRows = nuisance$valSets
  )
  ubar1 <- predict(fitubar, newdata = data.frame(A = 1, W))$pred[, 1]
  ubar0 <- predict(fitubar, newdata = data.frame(A = 0, W))$pred[, 1]
}

ubarA  <- A * ubar1 + (1 - A) * ubar0
q1     <- ubar1 - ubar0

if (sd(u1 - u0) < .Machine$double.eps * 10) {
  q2  <- rep(mean(u1 - u0), length(u1))
} else {
  fitq2 <- mySL(
    Y = u1 - u0,
    X = data.frame(W),
    family = stats::gaussian(),
    SL.library = candidatesq,
    validRows = nuisance$valSets
  )
  q2  <- fitq2$SL.predict[, 1]
}

g1delta <- gdelta1(g1, delta)
gdelta <- A * g1delta + (1 - A) * (1 - g1delta)

HD <- b / d * (1 - gdelta / e)
HI <- b / d * gdelta * (1 / e -  1 / g)
KD <- v1 - v0 - gdelta / g * (s1 - s0)
KI <- gdelta / g * (s1 - s0 - v1 + v0)
MD <- - g1delta * (1 - g1delta) / (g1 * (1 - g1)) * q2
MI <- g1delta * (1 - g1delta) / (g1 * (1 - g1)) * (q2 - q1)
J  <- - gdelta / g

de <- ubar1 * g1 + ubar0 * (1 - g1) - (u1 * g1delta + u0 * (1 - g1delta))
* svywt
ie <- (u1 - ubar1) * g1delta + (u0 - ubar0) * (1 - g1delta) * svywt
#de <- (u1 * g1delta + u0 * (1 - g1delta)) - ubar1 * g1 + ubar0 * (1 -
g1)
#ie <- (ubar1 - u1) * g1delta + (ubar0 - u0) * (1 - g1delta)
```

```
eifD <- (HD * (Y - m) + KD * (L - b1A) + MD * (A - g1) +
     (uA - ubarA) + de)
eifI <- (HI * (Y - m) + KI * (L - b1A) + MI * (A - g1) + J *
     (uA - ubarA) + ie)
osde <- mean(eifD)# - mean(eifD))
osie <- mean(eifI)# - mean(eifI))
osde2 <- mean(eifD - mean(eifD))
osie2 <- mean(eifI - mean(eifI))
seosde <- sd(eifD) / sqrt(length(A))
seosie <- sd(eifI) / sqrt(length(A))

## now, compute the TMLE
stopcrit <- FALSE
iter <- 1

## iterative TMLE
while (!stopcrit) {
  # fit logistic tilting models
  tiltm <- glm(Y ~ 0 + offset(qlogis(bound(m))) + HD + HI,
               family = binomial())
  tiltb <- glm(L ~ 0 + offset(qlogis(bound(b1A))) + KD + KI,
               family = binomial())
  tiltg <- glm(A ~ 0 + offset(qlogis(bound(g1))) + MD + MI,
               family = binomial())

  # check tilting model coefs and set to zero if NA
  if (any(is.na(coef(tiltm)))) {
    tiltm$coefficients[is.na(coef(tiltm))] <- 0
  }
  if (any(is.na(coef(tiltb)))) {
    tiltb$coefficients[is.na(coef(tiltb))] <- 0
  }
  if (any(is.na(coef(tiltg)))) {
    tiltg$coefficients[is.na(coef(tiltg))] <- 0
  }


  # predict from tilting models
  m   <- predict(tiltm, type = "response")
  b1A <- predict(tiltb, type = "response")
  g1  <- predict(tiltg, type = "response")

  HD11 <- b11 / d11 * (1 - g1delta / e1)
  HI11 <- b11 / d11 * g1delta * (1 / e1 -  1 / g1)
  HD10 <- b10 / d10 * (1 - (1 - g1delta) / (1 - e1))
```

```
  HI10 <- b10 / d10 * (1 - g1delta) * (1 / (1 - e1) -  1 / (1 - g1))

  HD01 <- (1 - b11) / (1 - d11) * (1 - g1delta / e1)
  HI01 <- (1 - b11) / (1 - d11) * g1delta * (1 / e1 -  1 / g1)
  HD00 <- (1 - b10) / (1 - d10) * (1 - (1 - g1delta) / (1 - e1))
  HI00 <- (1 - b10) / (1 - d10) *
    (1 - g1delta) * (1 / (1 - e1) - (1 / (1 - g1)))

  m11 <- plogis(qlogis(bound(m11)) + coef(tiltm)[1] * HD11 +
                  coef(tiltm)[2] * HI11)
  m10 <- plogis(qlogis(bound(m10)) + coef(tiltm)[1] * HD10 +
                  coef(tiltm)[2] * HI10)
  m01 <- plogis(qlogis(bound(m01)) + coef(tiltm)[1] * HD01 +
                  coef(tiltm)[2] * HI01)
  m00 <- plogis(qlogis(bound(m00)) + coef(tiltm)[1] * HD00 +
                  coef(tiltm)[2] * HI00)

  g <- A * g1 + (1 - A) * (1 - g1)
  b <- L * b1A + (1 - L) * (1 - b1A)
  g1delta <- gdelta1(g1, delta)
  gdelta <- A * g1delta + (1 - A) * (1 - g1delta)

  HD <- b / d * (1 - gdelta / e)
  HI <- b / d * gdelta * (1 / e -  1 / g)
  KD <- v1 - v0 - gdelta / g * (s1 - s0)
  KI <- gdelta / g * (s1 - s0 - v1 + v0)
  MD <- - g1delta * (1 - g1delta) / (g1 * (1 - g1)) * q2
  MI <- g1delta * (1 - g1delta) / (g1 * (1 - g1)) * (q2 - q1)

  de <- ubar1 * g1 + ubar0 * (1 - g1) - (u1 * g1delta + u0 * (1 -
  g1delta))
  ie <- (u1 - ubar1) * g1delta + (u0 - ubar0) * (1 - g1delta)

  eifDp <- svywt *
    (HD * (Y - m) + KD * (L - b1A) + MD * (A - g1))
  eifIp <- svywt *
    (HI * (Y - m) + KI * (L - b1A) + MI * (A - g1))

  ## iterate iterator
  iter <- iter + 1
  ## note: interesting stopping criterion
  stopcrit <- max(abs(c(mean(eifDp), mean(eifIp)))) <
    max(c(sd(eifDp), sd(eifIp))) *
    log(length(A)) / length(A) | iter > 6
}
```

```
J <- -(gdelta / g)
J1 <- -(g1delta / g1)
J0 <- -((1 - g1delta) / (1 - g1))

m1A <- A * m11 + (1 - A) * m10
m0A <- A * m01 + (1 - A) * m00

uA <- m1A * b1A + m0A * (1 - b1A)
u1 <- m11 * b11 + m01 * (1 - b11)
u0 <- m10 * b10 + m00 * (1 - b10)

tiltu <- glm(uA ~ offset(qlogis(bound(ubarA))) + J, family = binomial())
ubar1 <- plogis(coef(tiltu)[1] + qlogis(bound(ubar1)) +
                  coef(tiltu)[2] * J1)
ubar0 <- plogis(coef(tiltu)[1] + qlogis(bound(ubar0)) +
                  coef(tiltu)[2] * J0)
ubarA  <- A * ubar1 + (1 - A) * ubar0

# additive effects and standard errors
## estimates of the direct and indirect effects
de_est <- (ubar1 * g1 + ubar0 * (1 - g1) -
             (u1 * g1delta + u0 * (1 - g1delta))) * svywt
ie_est <- ((u1 - ubar1) * g1delta + (u0 - ubar0) * (1 - g1delta)) * svywt

## influence functions for direct and indirect effect
eifD <- HD * (Y - m) + KD * (L - b1A) + MD * (A - g1) + (uA - ubarA) +
  de_est
eifI <- HI * (Y - m) + KI * (L - b1A) + MI * (A - g1) + J * (uA - ubarA) +
  ie_est

## tmle risk difference effects and standard errors
tmlede <- mean(de_est)
tmleie <- mean(ie_est)
setmlede <- sd(eifD) / sqrt(length(A))
setmleie <- sd(eifI) / sqrt(length(A))

# RR effects and standard errors
## estimates of decomposed direct and indirect risks
rd00 <- (ubar1 * g1 + ubar0 * (1 - g1)) * svywt
rd10 <- (u1 * g1delta + u0 * (1 - g1delta)) * svywt
ri10 <- (u1 * g1delta + u0 * (1 - g1delta)) * svywt
ri11 <- (ubar1 * g1delta + ubar0 * (1 - g1delta)) * svywt

## eif for risks
```

```
  eif00 <- rd00 - mean(rd00)
  eif10 <- rd10 - mean(rd10)
  eif11 <- ri11 - mean(ri11)

  ## estimates of direct and indirect effects
  de <- (ubar1 * g1 + ubar0 * (1 - g1)) / (u1 * g1delta + u0 * (1 -
  g1delta))
  ie <- (u1 * g1delta + u0 * (1 - g1delta)) / (ubar1 * g1delta + ubar0 * (1
 - g1delta))

  ## influence functions for direct and indirect effect risks ratios
  eifDrr <- HD * (Y - m) + KD * (L - b1A) + MD * (A - g1) + (uA - ubarA) +
  de
  eifIrr <- HI * (Y - m) + KI * (L - b1A) + MI * (A - g1) + J * (uA - ubarA)
 +  ie

  ## tmle risk ratio effects and standard errors
  derr <- mean(de)
  ierr <- mean(ie)
  seDrr <- sd(eifDrr) / sqrt(length(A))
  seIrr <- sd(eifIrr) / sqrt(length(A))

  # output
  out <- data.frame(
    parameter    = c("DE", "IE"),
    parameter_rd = c(tmlede, tmleie),
    ses_rd       = c(setmlede, setmleie),
    parameter_rr = c(derr, ierr),
    ses_rr       = c(seDrr, seIrr),
    n            = length(A))
  return(out)
}
# load packages and helper files
library(here)
library(hal9001)
library(tidyverse)
library(SuperLearner)
library(future.apply)
library(parallel)
library(Amelia)
library(ggplot2)
library(ggthemes)
source(here("H:/pub_medshift_interv_biostats-master/application/R",
"utils.R"))
set.seed(7518, "L'Ecuyer-CMRG")
```

```r
# simulation parameters
ci_norm_bounds <- abs(stats::qnorm(p = (1 - 0.95) / 2))
n_obs <- nrow(df.list.can.PY[[1]])
delta_ipsi <- seq(.1, 1, by = 0.1)

# make mock data
#data <- simdata(n_obs)
#head(data)

# load permuted data from Kara
#data <- df.list.can.PY[[i]]

# permutation bug: A is all zero for some reason?!
#table(data$A)
#data$A <- rbinom(nrow(data), 1, 0.6)

# also, too many W for me to run on my personal machine, so we reduce here
#W_random <- as.data.frame(data[, substr(names(data), 1, 1) == "W"][,
10:19])
#data_reduced <- cbind(data[, c(str_subset(names(data), "Z"), "Y", "L",
"A")],
#                      W_random)

# remove extra stuff from utils.R
#rm(list = setdiff(ls(), c("bound", "truncate", "gfun", "gdeltafun",
"gdelta1",
#                          "data", "data_reduced")))
ls()
source(here("H:/pub_medshift_interv_biostats-master/application/R",
"slfuns.R"))
source(here("H:/pub_medshift_interv_biostats-master/application/R",
"estimate_new_2.R"))

# set up SL library + smaller library for faster debugging
sl_lib <- c("SL.mean", "SL.bayesglm", "SL.earth", "SL.gam", "SL.myglm",
            "SL.myglmnet", "SL.hal9001", "SL.caretXGB", "SL.caretRF")
#sl_lib_debug <- c("SL.mean", "SL.earth", "SL.glm")
sl_lib_debug <- c("SL.mean", "SL.glm")
#sl_lib_debug <- "SL.speedglm"

# create list of datasets and other objects for use in functions below
newout.w4 <- nuisance.list.w4 <- delta.list <- vector("list", imps)
for(i in 1:imps){
  delta.list[[i]] <- delta_ipsi
```

```
  newout.w4[[i]] <- vector("list", length(delta_ipsi))
}

# estimate nuisance parameters
plan(multisession, gc = TRUE, workers = availableCores())
nuisance.w4 <- future_mapply(df.list.can, nfolds = 5, sl_lib = sl_lib_debug,
future.seed = TRUE,
                             FUN = function(df.list.can_i, nfolds_i,
                             sl_lib_i){
                               fitnuisance(data = df.list.can_i, nfolds = 5,
                               sl_lib = sl_lib_debug)
                             })

for(i in 1:imps){
  nuisance.list.w4[[i]] <- nuisance.w4[,i]
}
remove(nuisance.w4)

# construct estimates of (in)direct effects
plan(multisession, gc = TRUE, workers = availableCores())
out.can.new <- lapply(delta_ipsi, function(delta.list) {
  future_mapply(df.list.can, delta.list, nuisance.list.w4,
                FUN = function(df.list.can_i, delta.list_i,
                nuisance.list_i){
                  estimators(data = df.list.can_i, delta = delta.list_i,
                  nuisance.list_i, sl_lib = sl_lib_debug)
                })})

remove(nuisance.list.w4)

# reorganizing output into legible format
for(j in 1:length(delta_ipsi)){
  for(i in 1:imps){
    newout.w4[[i]][[j]] <- as.data.frame(out.can.new[[j]][,i])
  }
}


# meld results from different imputed data sets to plot direct effect risk
ratios
coef.matrix.can.new.derr <- matrix(NA, imps, length(delta_ipsi))
se.matrix.can.new.derr <- matrix(NA, imps, length(delta_ipsi))
for(i in 1:imps){for(j in 1:max(length(delta_ipsi))){
  coef.matrix.can.new.derr[i,j] <- newout.w4[[i]][[j]][1,4]}}
for(i in 1:imps){for(j in 1:max(length(delta_ipsi))){
```

```
  se.matrix.can.new.derr[i,j] <- newout.w4[[i]][[j]][1,5]}}
melded.can.new.derr <- mi.meld(coef.matrix.can.new.derr,
se.matrix.can.new.derr)

# meld results from different imputed data sets to plot indirect risks
delta = 1
coef.matrix.can.new.ierr <- matrix(NA, imps, length(delta_ipsi))
se.matrix.can.new.ierr <- matrix(NA, imps, length(delta_ipsi))
for(i in 1:imps){for(j in 1:max(length(delta_ipsi))){
  coef.matrix.can.new.ierr[i,j] <- newout.w4[[i]][[j]][2,4]}}
for(i in 1:imps){for(j in 1:max(length(delta_ipsi))){
  se.matrix.can.new.ierr[i,j] <- newout.w4[[i]][[j]][2,5]}}
melded.can.new.ierr <- mi.meld(coef.matrix.can.new.ierr,
se.matrix.can.new.ierr)

# create dataframe of results to plot direct effect risk ratios
plot.table.can.new.derr <- matrix(NA, length(delta_ipsi)-1, 5)
plot.table.can.new.derr[,2] <-
melded.can.new.derr$q.mi[1:length(delta_ipsi)-1]
plot.table.can.new.derr[,3] <-
melded.can.new.derr$se.mi[1:length(delta_ipsi)-1]
plot.table.can.new.derr <- as.data.frame(plot.table.can.new.derr)
colnames(plot.table.can.new.derr) <- c("delta", "risk ratio", "se", "lcl",
"ucl")
plot.table.can.new.derr$delta <- delta_ipsi[-10]
plot.table.can.new.derr$lcl <- log(plot.table.can.new.derr$`risk ratio`) -
ci_norm_bounds * plot.table.can.new.derr$se
plot.table.can.new.derr$lcl <- exp(plot.table.can.new.derr$lcl)
plot.table.can.new.derr$ucl <- log(plot.table.can.new.derr$`risk ratio`) +
ci_norm_bounds * plot.table.can.new.derr$se
plot.table.can.new.derr$ucl <- exp(plot.table.can.new.derr$ucl)

# create dataframe of results to plot indirect effect risk ratios
plot.table.can.new.ierr <- matrix(NA, length(delta_ipsi)-1, 5)
plot.table.can.new.ierr[,2] <-
melded.can.new.ierr$q.mi[1:length(delta_ipsi)-1]
plot.table.can.new.ierr[,3] <-
melded.can.new.ierr$se.mi[1:length(delta_ipsi)-1]
plot.table.can.new.ierr <- as.data.frame(plot.table.can.new.ierr)
colnames(plot.table.can.new.ierr) <- c("delta", "risk ratio", "se", "lcl",
"ucl")
plot.table.can.new.ierr$delta <- delta_ipsi[-10]
plot.table.can.new.ierr$lcl <- log(plot.table.can.new.ierr$`risk ratio`) -
ci_norm_bounds * plot.table.can.new.ierr$se
plot.table.can.new.ierr$lcl <- exp(plot.table.can.new.ierr$lcl)
```

```
plot.table.can.new.ierr$ucl <- log(plot.table.can.new.ierr$`risk ratio`) +
ci_norm_bounds * plot.table.can.new.ierr$se
plot.table.can.new.ierr$ucl <- exp(plot.table.can.new.ierr$ucl)

# meld results from different imputed data sets to plot direct effects
coef.matrix.can.new.de <- matrix(NA, imps, length(delta_ipsi)-1)
se.matrix.can.new.de <- matrix(NA, imps, length(delta_ipsi)-1)
for(i in 1:imps){for(j in 1:max(length(delta_ipsi)-1)){
  coef.matrix.can.new.de[i,j] <- newout.w4[[i]][[j]][1,2]}}
for(i in 1:imps){for(j in 1:max(length(delta_ipsi)-1)){
  se.matrix.can.new.de[i,j] <- newout.w4[[i]][[j]][1,3]}}
melded.can.new.de <- mi.meld(coef.matrix.can.new.de, se.matrix.can.new.de)

# meld results from different imputed data sets to plot indirect effects
coef.matrix.can.new.ie <- matrix(NA, imps, length(delta_ipsi)-1)
se.matrix.can.new.ie <- matrix(NA, imps, length(delta_ipsi)-1)
for(i in 1:imps){for(j in 1:max(length(delta_ipsi)-1)){
  coef.matrix.can.new.ie[i,j] <- newout.w4[[i]][[j]][2,2]}}
for(i in 1:imps){for(j in 1:max(length(delta_ipsi)-1)){
  se.matrix.can.new.ie[i,j] <- newout.w4[[i]][[j]][2,3]}}
melded.can.new.ie <- mi.meld(coef.matrix.can.new.ie, se.matrix.can.new.ie)

# create dataframe of results to plot direct effects
plot.table.can.new.de <- matrix(NA, length(delta_ipsi)-1, 5)
plot.table.can.new.de[,2] <- melded.can.new.de$q.mi[1:length(delta_ipsi)-1]
plot.table.can.new.de[,3] <- melded.can.new.de$se.mi[1:length(delta_ipsi)-1]
plot.table.can.new.de <- as.data.frame(plot.table.can.new.de)
colnames(plot.table.can.new.de) <- c("delta", "risk difference", "se",
"lcl", "ucl")
plot.table.can.new.de$delta <- delta_ipsi[-10]
plot.table.can.new.de$lcl <- plot.table.can.new.de$`risk difference` -
ci_norm_bounds * plot.table.can.new.de$se
plot.table.can.new.de$ucl <- plot.table.can.new.de$`risk difference` +
ci_norm_bounds * plot.table.can.new.de$se

# create dataframe of results to plot indirect effects
plot.table.can.new.ie <- matrix(NA, length(delta_ipsi)-1, 5)
plot.table.can.new.ie[,2] <- melded.can.new.ie$q.mi[1:length(delta_ipsi)-1]
plot.table.can.new.ie[,3] <- melded.can.new.ie$se.mi[1:length(delta_ipsi)-1]
plot.table.can.new.ie <- as.data.frame(plot.table.can.new.ie)
colnames(plot.table.can.new.ie) <- c("delta", "risk difference", "se",
"lcl", "ucl")
plot.table.can.new.ie$delta <- delta_ipsi[-10]
plot.table.can.new.ie$lcl <- plot.table.can.new.ie$`risk difference` -
ci_norm_bounds * plot.table.can.new.ie$se
```

```
plot.table.can.new.ie$ucl <- plot.table.can.new.ie$`risk difference` +
ci_norm_bounds * plot.table.can.new.ie$se

# plot results of direct effects
#plot.table.can.new.de <-
plot.table.can.new.de[!plot.table.can.new.de$delta == 0.1,]
plot.can.new.derr <- ggplot(plot.table.can.new.derr) +
  geom_ribbon(aes(x = delta, ymin = lcl, ymax = ucl),
              fill = "lightblue", color = "lightblue", alpha = 0.5) +
  geom_line(aes(x = delta, y = `risk ratio`), color = "steelblue") +
  scale_x_continuous(expand = c(0,0), breaks = seq(0,1,.2)) +
  scale_y_continuous(limits = c(0.95, 1.12)) + #c(0.84, 1.14))
  theme_hc() + scale_color_hc()
plot.can.new.derr

plot.can.new.de <- ggplot(plot.table.can.new.de) +
  geom_ribbon(aes(x = delta, ymin = lcl, ymax = ucl),
              fill = "lightblue", color = "lightblue", alpha = 0.5) +
  geom_line(aes(x = delta, y = `risk difference`), color = "steelblue") +
  scale_x_continuous(expand = c(0,0), breaks = seq(0,1,.2)) +
  scale_y_continuous(limits = c(-0.01, 0.03)) + #c(-0.03, 0.02)
  theme_hc() + scale_color_hc()
plot.can.new.de

# plot results of indirect effects
#plot.table.can.new.ie <-
plot.table.can.new.ie[!plot.table.can.new.ie$delta == 0.1,]
plot.can.new.ierr <- ggplot(plot.table.can.new.ierr) +
  geom_ribbon(aes(x = delta, ymin = lcl, ymax = ucl),
              fill = "lightblue", color = "lightblue", alpha = 0.5) +
  geom_line(aes(x = delta, y = `risk ratio`), color = "steelblue") +
  scale_x_continuous(expand = c(0,0), breaks = seq(0,1,.2)) +
  scale_y_continuous(limits = c(0.98, 1.03)) + #c(0.95, 1.1)) +
  theme_hc() + scale_color_hc()
plot.can.new.ierr
plot.can.new.ie <- ggplot(plot.table.can.new.ie) +
  geom_ribbon(aes(x = delta, ymin = lcl, ymax = ucl),
              fill = "lightblue", color = "lightblue", alpha = 0.5) +
  geom_line(aes(x = delta, y = `risk difference`), color = "steelblue") +
  scale_x_continuous(expand = c(0,0), breaks = seq(0,1,.2)) +
  scale_y_continuous(limits = c(-0.01, 0.01)) + #c(-0.01, 0.02)) +
  theme_hc() + scale_color_hc()
plot.can.new.ie

plot.can.new.ie <- ggplot(plot.table.can.new.ie) +
```

```
  geom_ribbon(aes(x = delta, ymin = lcl, ymax = ucl),
              fill = "lightblue", color = "lightblue", alpha = 0.5) +
  geom_line(aes(x = delta, y = `risk difference`), color = "steelblue") +
  scale_x_continuous(expand = c(0,0), breaks = seq(0,1,.2)) +
  scale_y_continuous(limits = c(-0.01, 0.01)) + #c(-0.01, 0.02)) +
  theme_hc() + scale_color_hc()
plot.can.new.ie


# saved 7/15/2022 7:17am - still have to estimate W4 outcome and E-values
#APY12_M3_Y4_can_de <- plot.table.can.new.de #tmle
#APY12_M3_Y4_can_ie <- plot.table.can.new.ie #tmle
#APY12_M3_Y4_can_derr <- plot.table.can.new.derr #tmle
#APY12_M3_Y4_can_ierr <- plot.table.can.new.ierr #tmle
APY123_M4_Y5_can_de <- plot.table.can.new.de #tmle
APY123_M4_Y5_can_ie <- plot.table.can.new.ie #tmle
APY123_M4_Y5_can_derr <- plot.table.can.new.derr #tmle
APY123_M4_Y5_can_ierr <- plot.table.can.new.ierr #tmle

# E-values for Wave 5 outcome
rbind_rr_tables_can_Y5 <- rbind(plot.table.can.new.derr,
plot.table.can.new.ierr)
rbind_rr_tables_can_Y5$lcl <- ifelse(rbind_rr_tables_can_Y5$lcl < 1, 1,
rbind_rr_tables_can_Y5$lcl)
rbind_rr_tables_can_Y5$`risk ratio` <- ifelse(rbind_rr_tables_can_Y5$`risk
ratio` < 1, 1/rbind_rr_tables_can_Y5$`risk ratio`,
rbind_rr_tables_can_Y5$`risk ratio`)
evalue_table_can_Y5 <- cbind(delta = rbind_rr_tables_can_Y5[,1],
rbind_rr_tables_can_Y5[,c(2,4,5)] + sqrt(rbind_rr_tables_can_Y5[,c(2,4,5)] *
rbind_rr_tables_can_Y5[,c(2,4,5)]-1))

# E-values for Wave 4 outcome
#rbind_rr_tables_can_Y4 <- rbind(plot.table.can.new.derr,
plot.table.can.new.ierr)
#rbind_rr_tables_can_Y4$lcl <- ifelse(rbind_rr_tables_can_Y4$lcl < 1, 1,
rbind_rr_tables_can_Y4$lcl)
#rbind_rr_tables_can_Y4$`risk ratio` <-
ifelse(rbind_rr_tables_can_Y4$`risk ratio` < 1,
1/rbind_rr_tables_can_Y4$`risk ratio`, rbind_rr_tables_can_Y4$`risk
ratio`)
#evalue_table_can_Y4 <- cbind(delta = rbind_rr_tables_can_Y4[,1],
rbind_rr_tables_can_Y4[,c(2,4,5)] + sqrt(rbind_rr_tables_can_Y4[,c(2,4,5)]
* rbind_rr_tables_can_Y4[,c(2,4,5)]-1))

save.image("H:/Rudolph_1/Aim3_redo.RData")
```

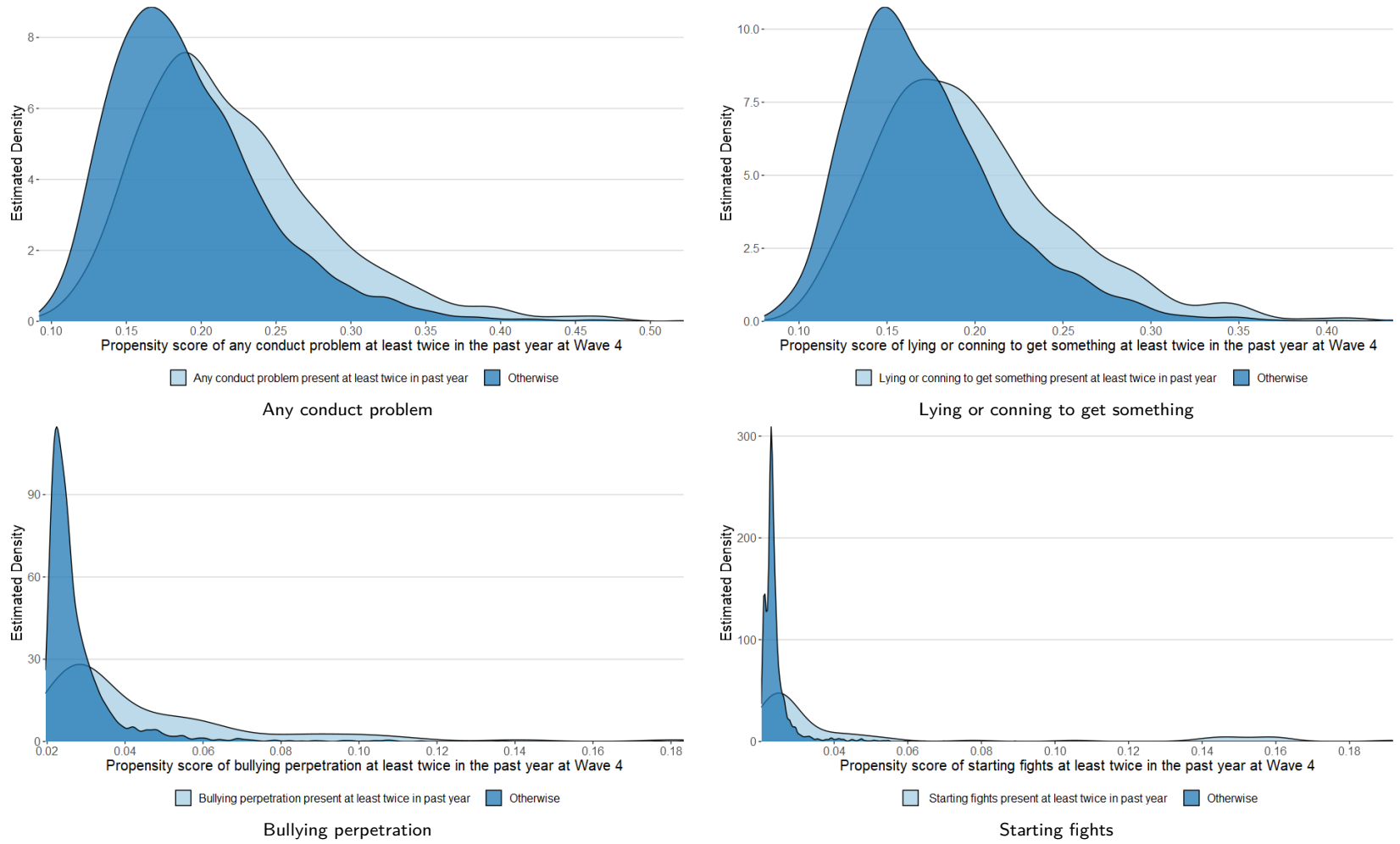

Figure B.5: Plot showing overlap of propensity scores for any conduct problem, lying or conning to get something, bullying perpetration, and starting fights by corresponding observed groupings (i.e., present at least twice in the past year and otherwise) in Wave 4 of PATH, 2,949 baseline participants aged 12 to 16 years old in cannabis outcome cohort

Table B.8: Distribution of baseline confounders among baseline 12- to 16-year-old youth in cannabis initiation (N = 2,949) and other substance initiation (N=2,995) outcome cohorts, Population Assessment of Tobacco and Health

	Cannabis initiation				Other substance initiation			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[a]	Weighted proportion (%)[c]	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Electronic nicotine product use/nicotine vaping								
No	2,888	98.72	98.27	99.16	2,879	96.65	95.90	97.40
Yes	37	1.28	0.84	1.73	93	3.35	2.60	4.10
Racialized groups								
White	2,662	74.88	68.16	81.6	2,151	74.56	67.75	81.37
Other	1,037	25.12	18.4	31.84	844	25.44	18.63	32.25
Sex								
Male	1,562	52.04	49.77	54.32	1,644	53.74	51.55	55.93
Female	1,387	47.96	45.68	50.23	1,351	46.26	44.07	48.45
Baseline age								
12 to 14	696	65.92	64.00	67.83	641	61.84	59.92	63.77
15 to 16	672	34.08	32.17	36.00	644	38.16	36.23	40.08
Household income								
Up to $50,000	1,049	43.20	39.69	46.70	1,042	43.09	39.39	46.79
More than $50,000	945	56.80	53.30	60.31	963	56.91	53.21	60.61
Parent education								
Up to a bachelor's degree	1,833	53.94	50.71	57.16	1,884	54.75	51.58	57.92

Continued on next page

Table B.8: Distribution of baseline confounders among baseline 12- to 16-year-old youth in cannabis initiation (N = 2,949) and other substance initiation (N=2,995) outcome cohorts, Population Assessment of Tobacco and Health) (cont.).

	Cannabis initiation				Other substance initiation			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[a]	Weighted proportion (%)[c]	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Some graduate or professional education or degree	1,095	46.06	42.84	49.29	1,093	45.25	42.08	48.42
Parent currently married								
Currently married	1,985	69.14	66.20	72.07	1,979	67.95	65.06	70.84
Not currently married	949	30.86	27.93	33.80	1,004	32.05	29.16	34.94
A parent lives elsewhere / youth sees them								
No	1,900	65.04	62.10	67.97	1,879	63.57	60.68	66.45
Yes, child sees them	545	19.74	17.79	21.70	578	20.73	18.67	22.80
Yes, child never sees them	486	15.22	13.38	17.05	522	15.70	13.78	17.62
Someone other than youth uses tobacco in the home								
No	2,120	73.94	70.79	77.10	2,108	72.67	69.74	75.61
Yes	791	26.06	22.90	29.21	856	27.33	24.39	30.26
Tobacco prohibited in home								
Yes	2,146	71.84	68.76	74.92	2,156	71.04	67.99	74.09

Continued on next page

Table B.8: Distribution of baseline confounders among baseline 12- to 16-year-old youth in cannabis initiation (N = 2,949) and other substance initiation (N=2,995) outcome cohorts, Population Assessment of Tobacco and Health) (cont.).

	Cannabis initiation				Other substance initiation			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[a]	Weighted proportion (%)[c]	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
No	787	28.16	25.08	31.24	825	28.96	25.91	32.01
Youth has curfew								
No	325	12.02	10.35	13.69	333	12.14	10.43	13.85
Yes	2,612	87.98	86.31	89.65	2,653	87.86	86.15	89.57
Parent's reaction if youth used tobacco								
Very upset	2,808	96.59	95.90	97.27	2,828	95.69	94.90	96.49
Not too upset or no reaction	109	3.41	2.73	4.10	135	4.31	3.51	5.10
Parent talked to child about not using tobbaco in past year								
Yes	1,407	47.39	44.93	49.85	1,465	48.30	45.80	50.80
No	1,494	52.61	50.15	55.07	1,483	51.70	49.20	54.20
Youth school performance (grades)								
Mostly B's or worse	1,023	31.06	28.49	33.64	1,090	32.54	30.01	35.08
Mostly A's and B's or better	1,912	68.94	66.36	71.51	1,894	67.46	64.92	69.99
Youth sexual orientation								
Straight	1,480	97.07	96.19	97.96	1,596	96.45	95.34	97.57

Continued on next page

Table B.8: Distribution of baseline confounders among baseline 12- to 16-year-old youth in cannabis initiation (N = 2,949) and other substance initiation (N=2,995) outcome cohorts, Population Assessment of Tobacco and Health) (cont.).

	Cannabis initiation				Other substance initiation			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[a]	Weighted proportion (%)[c]	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Not straight	54	2.93	2.04	3.81	66	3.55	2.43	4.66
Youth's overall health (parent report)								
Good, very good, or excellent	2,548	90.05	88.72	91.38	2,605	90.54	89.23	91.86
Fair or poor	389	9.95	8.62	11.28	381	9.46	8.14	10.77
Doctor told child they had ADHD (parent report)								
Never	2,599	87.94	86.56	89.32	2,637	88.11	86.65	89.56
Ever	336	12.06	10.68	13.44	345	11.89	10.44	13.35
Youth likes to do frightening things								
No	2,715	93.32	92.24	94.40	2,724	91.92	90.73	93.10
Yes	187	6.68	5.60	7.76	224	8.08	6.90	9.27
Youth likes new and exciting experiences, even if they have to break rules								
No	2,724	93.71	92.76	94.67	2,725	91.84	90.77	92.92
Yes	189	6.29	5.33	7.24	238	8.16	7.08	9.23

Continued on next page

Table B.8: Distribution of baseline confounders among baseline 12- to 16-year-old youth in cannabis initiation (N = 2,949) and other substance initiation (N=2,995) outcome cohorts, Population Assessment of Tobacco and Health) (cont.).

	Cannabis initiation				Other substance initiation			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[a]	Weighted proportion (%)[c]	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Youth prefers friends who are exciting and unpredictable								
No	2,356	82.07	80.52	83.62	2,352	80.59	79.00	82.18
Yes	528	17.93	16.38	19.48	577	19.41	17.82	21.00
Youth felt depressed at least twice in past year								
No	2,398	83.02	81.58	84.45	2,425	82.26	80.66	83.86
Yes	496	16.98	15.55	18.42	517	17.74	16.14	19.34
Youth had trouble sleeping at least twice in past year								
No	2,159	74.22	72.03	76.41	2,210	74.61	72.46	76.77
Yes	748	25.78	23.59	27.97	748	25.39	23.23	27.54
Youth felt anxious at least twice in past year								
No	2,173	73.91	71.80	76.02	2,224	74.52	72.63	76.41
Yes	736	26.09	23.98	28.20	729	25.48	23.59	27.37
Youth felt distressed about past events at least twice in past year								

Continued on next page

Table B.8: Distribution of baseline confounders among baseline 12- to 16-year-old youth in cannabis initiation (N = 2,949) and other substance initiation (N=2,995) outcome cohorts, Population Assessment of Tobacco and Health) (cont.).

	Cannabis initiation				Other substance initiation			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[a]	Weighted proportion (%)[c]	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
No	2,412	83.70	81.98	85.43	2,458	83.79	82.29	85.29
Yes	490	16.30	14.57	18.02	494	16.21	14.71	17.71
Youth had trouble paying attention at school/work/home at least twice in past year								
No	2,145	74.14	71.99	76.28	2,165	73.07	71.10	75.04
Yes	766	25.86	23.72	28.01	789	26.93	24.96	28.90
Youth had trouble listening to instructions at school/work/home at least twice in past year								
No	2,367	81.62	79.84	83.39	2,392	80.70	78.87	82.52
Yes	544	18.38	16.61	20.16	567	19.30	17.48	21.13
Youth tobacco use in past month[d]								
No	2,737	98.82	98.39	99.25	2,741	97.39	96.75	98.03
Yes	29	1.18	0.75	1.61	75	2.61	1.97	3.25
Youth ever used alcohol								
Never	2,500	84.19	82.48	85.90	2,440	80.68	78.83	82.53
Ever	430	15.81	14.10	17.52	534	19.32	17.47	21.17

Continued on next page

Table B.8: Distribution of baseline confounders among baseline 12- to 16-year-old youth in cannabis initiation (N = 2,949) and other substance initiation (N=2,995) outcome cohorts, Population Assessment of Tobacco and Health) (cont.).

	Cannabis initiation				Other substance initiation			
			95% Confidence Interval[c]				95% Confidence Interval[c]	
	N[a]	Weighted proportion (%)[c]	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Youth ever used other substances/Youth ever used other substances								
Never	2,740	96.20	95.30	97.10	1,360	94.02	92.81	95.24
Ever	113	3.80	2.90	4.70	137	5.98	4.76	7.19

[a] N = Unweighted frequency
[b] ADHD = Attention deficit hyperactivity disorder
[c] All proportions are weighted to reflect the complex survey design of the PATH and are a weighted average based on 15 imputed datasets to account for missingness. Frequency sums are different and do not add to the baseline cohort totals of 2,949 and 2,995 for some variables because they are extracted from the original baseline wave that has missingness. Proportions are weighted by the baseline probability of being sampled.
[d] Tobacco was defined as any cigarettes, cigars, cigarillos, filtered cigars, pipes, hookah, or smokeless tobacco in Wave 1.

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Cannabis initiation cohort								
Race								
White	1,364	73.88	66.90	80.86	83	78.60	69.22	87.98
Other	558	26.12	19.14	33.10	29	21.40	12.02	30.78
Age								
12 to 14	1,330	67.47	65.53	69.40	51	43.93	35.41	52.44
15 to 16	592	32.53	30.60	34.47	61	56.07	47.56	64.59
Sexual identity								
Male	1,005	51.39	49.03	53.75	75	61.31	51.83	70.80
Female	917	48.61	46.25	50.97	37	38.69	29.20	48.17
Household income								
Up to $50,000	876	42.43	38.85	46.01	52	54.10	44.29	63.90
More than $50,000	784	57.57	53.99	61.15	36	45.90	36.10	55.71
Parent education								
Up to a bachelor's degree	1,161	53.18	49.93	56.42	74	64.73	56.28	73.17
Some graduate or professional education or degree	749	46.82	43.58	50.07	37	35.27	26.83	43.72

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Parent currently married								
Currently married	1,352	69.49	66.58	72.39	70	64.17	54.52	73.82
Not currently married	563	30.51	27.61	33.42	41	35.83	26.18	45.48
A parent lives elsewhere / youth sees them								
No	1,312	65.76	62.71	68.82	59	54.74	45.92	63.56
Yes, child sees them	304	19.13	17.08	21.17	35	28.46	21.23	35.70
Yes, child never sees them	297	15.11	13.16	17.06	17	16.80	10.10	23.50
Someone other than youth uses tobacco in the home								
No	1,449	75.22	72.20	78.23	57	55.85	46.14	65.55
Yes	445	24.78	21.77	27.80	54	44.15	34.45	53.86
Tobacco prohibited in home								
Yes	1,416	72.56	69.58	75.53	68	61.65	52.11	71.20
No	497	27.44	24.47	30.42	43	38.35	28.80	47.89

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Parent's reaction if youth used tobacco								
Very upset	1,846	97.09	96.37	97.82	97	89.37	84.20	94.53
Not too upset or no reaction	53	2.91	2.18	3.63	12	10.63	5.47	15.80
Parent talked to child about not using tobacco in past year								
Yes	924	47.02	44.43	49.62	59	52.57	43.13	62.01
No	965	52.98	50.38	55.57	52	47.43	37.99	56.87
Youth school performance (grades)								
Mostly B's or worse	613	29.89	27.28	32.51	54	47.61	38.91	56.32
Mostly A's and B's or better	1,304	70.11	67.49	72.72	57	52.39	43.68	61.09
Youth's overall health (parent report)								
Good, very good, or excellent	1,672	90.16	88.78	91.54	97	88.53	82.34	94.72
Fair or poor	245	9.84	8.46	11.22	14	11.47	5.28	17.66

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Doctor told child they had ADHD (parent report)								
Never	1,702	88.44	87.00	89.88	94	80.76	73.39	88.13
Ever	213	11.56	10.12	13.00	17	19.24	11.87	26.61
Youth likes to do frightening things								
No	1,793	93.76	92.70	94.83	100	87.12	80.02	94.23
Yes	99	6.24	5.17	7.30	11	12.88	5.77	19.98
Youth likes new and exciting experiences, even if they have to break rules								
No	1,799	94.40	93.41	95.40	90	83.88	76.40	91.35
Yes	105	5.60	4.60	6.59	20	16.12	8.65	23.60
Youth prefers friends who are exciting and unpredictable								
No	1,586	82.93	81.31	84.55	76	69.86	60.59	79.12
Yes	302	17.07	15.45	18.69	34	30.14	20.88	39.41

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Youth felt depressed at least twice in past year								
No	1,582	83.59	82.13	85.05	85	74.89	67.66	82.12
Yes	309	16.41	14.95	17.87	25	25.11	17.88	32.34
Youth had trouble sleeping at least twice in past year								
No	1,406	74.59	72.34	76.83	75	69.00	61.85	76.15
Yes	495	25.41	23.17	27.66	36	31.00	23.85	38.15
Youth felt anxious at least twice in past year								
No	1,443	74.04	71.81	76.27	80	72.16	63.69	80.62
Yes	460	25.96	23.73	28.19	30	27.84	19.38	36.31
Youth felt distressed about past events at least twice in past year								
No	1,603	84.15	82.41	85.90	81	77.37	70.47	84.26
Yes	294	15.85	14.10	17.59	27	22.63	15.74	29.53

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Youth had trouble paying attention at school/work/home at least twice in past year								
No	1,453	74.93	72.67	77.19	68	62.86	54.62	71.11
Yes	446	25.07	22.81	27.33	44	37.14	28.89	45.38
Youth tobacco use in past month[d]								
No	1,820	99.74	99.50	99.97	88	85.82	80.20	91.45
Yes	5	0.26	0.03	0.50	15	14.18	8.55	19.80
Youth ever used alcohol								
Never	1,671	85.66	83.84	87.48	68	63.29	54.38	72.20
Ever	241	14.34	12.52	16.16	44	36.71	27.80	45.62
Youth ever used prohibited substances								
Never	1,805	96.27	95.31	97.22	99	95.28	91.58	98.98
Ever	64	3.73	2.78	4.69	8	4.72	1.02	8.42
Other substance initiation cohort								
Race								

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
White	1,305	74.31	67.43	81.19	169	76.40	67.77	85.03
Other	512	25.69	18.81	32.57	57	23.60	14.97	32.23
Age								
12 to 14	1,226	65.80	63.76	67.84	75	32.88	26.53	39.23
15 to 16	591	34.20	32.16	36.24	151	67.12	60.77	73.47
Sexual identity								
Male	983	52.95	50.51	55.39	146	59.52	53.27	65.76
Female	834	47.05	44.61	49.49	80	40.48	34.24	46.73
Household income								
Up to $50,000	819	42.18	38.44	45.91	91	49.74	41.59	57.90
More than $50,000	760	57.82	54.09	61.56	74	50.26	42.10	58.41
Parent education								
Up to a bachelor's degree	1,098	53.30	49.93	56.67	158	65.35	58.89	71.82
Some graduate or professional education or degree	706	46.70	43.33	50.07	67	34.65	28.18	41.11
Parent currently married								
Currently married	1,269	68.82	65.90	71.75	130	61.53	54.75	68.32
Not currently married	541	31.18	28.25	34.10	95	38.47	31.68	45.25

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
A parent lives elsewhere / youth sees them								
No	1,226	65.17	62.13	68.22	115	51.82	44.68	58.96
Yes, child sees them	294	19.43	17.26	21.59	65	30.30	24.11	36.49
Yes, child never sees them	288	15.40	13.35	17.45	45	17.88	12.62	23.15
Someone other than youth uses tobacco in the home								
No	1,363	74.87	71.78	77.96	118	56.62	49.74	63.51
Yes	433	25.13	22.04	28.22	105	43.38	36.49	50.26
Tobacco prohibited in home								
Yes	1,339	72.02	68.84	75.20	142	63.87	57.04	70.69
No	469	27.98	24.80	31.16	83	36.13	29.31	42.96
Parent's reaction if youth used tobacco								
Very upset	1,744	96.78	95.99	97.56	190	87.74	83.29	92.18
Not too upset or no reaction	53	3.22	2.44	4.01	32	12.26	7.82	16.71

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Parent talked to child about not using tobacco in past year								
Yes	899	47.97	45.24	50.70	119	50.76	44.29	57.22
No	888	52.03	49.30	54.76	104	49.24	42.78	55.71
Youth school performance (grades)								
Mostly B's or worse	585	30.17	27.47	32.86	120	49.93	43.48	56.37
Mostly A's and B's or better	1,227	69.83	67.14	72.53	105	50.07	43.63	56.52
Youth's overall health (parent report)								
Good, very good, or excellent	1,592	90.50	89.06	91.93	204	90.88	87.03	94.72
Fair or poor	220	9.50	8.07	10.94	21	9.12	5.28	12.97
Doctor told child they had ADHD (parent report)								
Never	1,623	89.01	87.53	90.49	185	81.48	76.36	86.61
Ever	187	10.99	9.51	12.47	38	18.52	13.39	23.64

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Youth likes to do frightening things								
No	1,683	93.11	92.00	94.21	189	83.18	77.41	88.95
Yes	104	6.89	5.79	8.00	36	16.82	11.05	22.59
Youth likes new and exciting experiences, even if they have to break rules								
No	1,696	93.74	92.68	94.80	174	77.96	72.19	83.73
Yes	105	6.26	5.20	7.32	50	22.04	16.27	27.81
Youth prefers friends who are exciting and unpredictable								
No	1,492	82.61	80.88	84.35	146	65.80	59.52	72.07
Yes	291	17.39	15.65	19.12	76	34.20	27.93	40.48
Youth felt depressed at least twice in past year								
No	1,501	83.28	81.68	84.88	171	74.81	69.13	80.49
Yes	290	16.72	15.12	18.32	51	25.19	19.51	30.87

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Youth had trouble sleeping at least twice in past year								
No	1,359	75.50	73.29	77.72	152	68.09	61.84	74.34
Yes	442	24.50	22.28	26.71	73	31.91	25.66	38.16
Youth felt anxious at least twice in past year								
No	1,378	74.53	72.42	76.63	169	74.47	68.63	80.31
Yes	419	25.47	23.37	27.58	53	25.53	19.69	31.37
Youth felt distressed about past events at least twice in past year								
No	1,530	84.74	83.13	86.34	173	76.89	71.10	82.68
Yes	266	15.26	13.66	16.87	49	23.11	17.32	28.90
Youth had trouble paying attention at school/work/home at least twice in past year								
No	1,386	74.79	72.55	77.03	139	60.48	53.99	66.97
Yes	408	25.21	22.97	27.45	86	39.52	33.03	46.01

Continued on next page

Table B.9: Descriptive statistics of baseline 12-16-year-old youth who never used cannabis (N=2,949) or other substances (N=2,995) as of Wave 4 and never reported conduct problems as of Wave 3 from the Population Assessment of Tobacco and Health (cont.).

	No past-year nicotine vaping at Waves 1 through 3				Past-year nicotine vaping at Waves 1 or 2 or 3			
			95% Confidence Interval [c]				95% Confidence Interval [c]	
	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)	N[a]	Weighted proportion (%)	Lower (%)	Upper (%)
Youth tobacco use in past month[d]								
No	1,719	99.35	98.95	99.74	167	83.07	78.37	87.77
Yes	8	0.65	0.26	1.05	46	16.93	12.23	21.63
Youth ever used alcohol								
Never	1,558	84.67	82.70	86.63	111	51.46	44.69	58.23
Ever	248	15.33	13.37	17.30	114	48.54	41.77	55.31
Youth ever used cannabis								
Never	846	97.18	96.19	98.18	98	70.90	64.03	77.77
Ever	18	2.82	1.82	3.81	73	29.10	22.23	35.97

[a] N = Unweighted frequency

[b] ADHD = Attention deficit hyperactivity disorder

[c] All proportions are weighted to reflect the complex survey design of the PATH and are a weighted average based on 15 imputed datasets to account for missingness. Frequency sums are different and do not add to the baseline cohort totals of 2,949 and 2,995 for some variables because they are extracted from the original baseline wave that has missingness. Proportions are weighted by the baseline probability of being sampled.

[d] Tobacco was defined as any cigarettes, cigars, cigarillos, filtered cigars, pipes, hookah, or smokeless tobacco in Wave 1.