Applied Computational Modeling Approaches in Cigarette Smoking Epidemiology: Extending Statistical Associations to Convey Theoretical Pathways

by

Sarah T. Cherng

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2017

Doctoral Committee:

Assistant Professor Rafael Meza, Chair
Professor Ana Diez Roux, Drexel University
Assistant Professor Marisa C. Eisenberg
Professor Paula M. Lantz
Associate Professor David Mendez

Sarah T. Cherng

scherng@umich.edu

ORCID iD: 0000-0002-9000-4202

DEDICATION

To my mother, for being exactly the mother that I've needed in every moment, and to the University of Michigan, for providing an environment that fosters profound personal and intellectual growth.

Forever, Go Blue.

ACKNOWLEDGEMENTS

This dissertation is the result of a culmination of support from various people at the University of Michigan School of Public Health, the Center for the Study of Complex Systems, the Michigan Institute for Computational Discovery and Engineering, the Center for Social Epidemiology and Population Health, and in my daily life. First, I would like to thank my committee chair: Dr. Rafael Meza. I can say with absolute certainty that this dissertation would not have materialized if not for his support. His expertise in research and mentorship are further reinforced by his genuine kindness and care for his students. I have been extremely lucky to work under his patient mentorship, and as a result, I have learned many important scientific (and life!) lessons that I will carry with me into the future. I'd also like to thank my committee members. Dr. Ana Diez Roux, for generously sharing her insights during committee meetings. These perspectives, along with her extensive work in multi-level modeling of neighborhood influences on health have contributed enormously to the ideas presented throughout this dissertation. Dr. Paula Lantz, for her frank life advice throughout the past decade, and for her support of the topics of this thesis. Her emphasis on the practical applications of science and its real-world implications has played a large role in the framing of this research. Dr. Marisa Eisenberg, for her unique kindness, willingness to listen, engage, and share ideas, and most importantly, for her gentle encouragement of my research-related endeavors and beyond. Dr. David Mendez, for offering his expertise in computational modeling methods in tobacco control as part of my committee. His role in the development of impactful tobacco control models has played a large role in the relevance of this dissertation today.

I would also like to thank Dr. Rick (Mario) Riolo, who was directly and independently responsible for my initial interest in computational modeling (;; i forgot ehat i i eshop radiozzzz/ paper wag sour?). In a similar vein, thanks to those (past and present) at the Center for the Study of Complex Systems: Howard Oishi, Mita Gibson, and Dr. Scott Page. There are few academic centers in the world that can support and change a person as much as CSCS has supported and

changed me. Throughout this dissertation process, I have had the support of many wonderful friends. Thank you to Yu-Han Kao and Lisa Lau, for being my best friends and allowing me to live in their house, drink their tea, break their dishes, eat their ramen, and ask (or sometimes bait) a never-ending slew of statistical and epidemiologic methods questions. This dissertation is finally complete because they have been such wonderful supporters of me in life and in science. Thank you to ("Dr.") Josh Evron, for taking care of Kai and Annie when I was shirking my parental responsibilities these past couple of months, for letting me borrow his blue pass, and for his (unique and sometimes awkward brand of) support and kindness. Thank you to Dr. Beth Percha, for providing reliable sanity checks, and for always encouraging me to confront my fears. Additionally, thank you to the coauthors of the papers included in this dissertation, who have provided invaluable intellectual contributions to its contents: Jamie Tam, Drs. Paul Christine, Nancy Fleischer, and Tonatiuh Barrientos Gutiérrez. Thank you, also, to Joanne Chang, Velma Lopez, Sonia Hegde, and Michael Hayashi, for their friendship and sincere support in various forms throughout this dissertation process.

Finally, I'd like to thank my family. My mom, for her gentle heart, and for providing a consistent and secure home that nurtured my (very broad) interests. Thank you to my aunts, uncles, and cousins for always taking care of me and for being a consistent part of my growth. Specifically, thank you, 三姨 and 小姨, for all the emotional support that they have provided to me throughout my life, from adolescence into adulthood. These memories comprise of a shared love that I will forever carry with me and treasure. And thank you, to my father, for teaching me at a very young age to think for myself. Life has a way of evolving to challenge and change us in ways that we would have never imagined. This dissertation process has been no exception. I have experienced enormous growth throughout these past years, all of which would not have been possible without the accumulation of conversation, experience, adaptation, evolution, and interaction. This dissertation, in addition to who I am today, is an emergent property of a complex system of those that have been involved in my life, both included in this acknowledgements section and not. Thank you to all of those who have contributed to my strange loop.

This work was made possible due to generous funding support from the Myron and Isabel Wegman Scholarship fund and the Office of Behavioral and Social Sciences Research / National Institutes of Health Contract HHSN276200800013C. Additional funding was received from the University of Michigan Department of Epidemiology and the Michigan Institute for

Computational Discovery and Engineering for computational resources. A portion of this work was also supported by grants U01CA152956 and U01CA199284 from the Cancer Intervention and Surveillance Modeling Network (CISNET). The contents of this dissertation are solely the responsibility of the author and do not necessarily represent the official views of the National Cancer Institute, CISNET, or the National Institutes of Health.

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for these analyses.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

50 years since the landmark 1964 Surgeon General's report on smoking and health, cigarette smoking remains the leading preventable cause of death and disability in the United States. The success of epidemiology and public health in the study of cigarette smoking, both as an exposure as well as a health outcome, has offered rich datasets and mechanistic discoveries that provide opportunities for the evolution of epidemiologic methods. Specifically, advancing computational science approaches allow for novel applications of methodologies, such as agent-based modeling or networks theory, in the epidemiological sciences to expand on existing knowledge. In this dissertation, we utilize approaches from epidemiology, statistics, computer science, and the philosophy of science to explore a range of hypothesized dynamics of smoking behavior that could contribute to changes in population-level smoking prevalence.

We begin with a computational model that weighs the magnitude of the potential harms and benefits of electronic cigarette (e-cigarette or vaping) use from an adult smoking prevalence standpoint. We find that e-cigarettes can exert a much larger influence on smoking prevalence through routes of smoking cessation, as opposed to smoking initiation, if e-cigarette use remains primarily concentrated among current smokers. Conversely, e-cigarettes would need to behave as extremely effective gateways for smoking initiation, and never smokers would need to become e-cigarette users at substantially higher levels than currently observed, for these products to independently generate increases in population-level smoking prevalence.

Next, we explore how contextual and individual network factors and demographic covariates change the effect of peer influence on smoking behavior in the National Longitudinal Study of Adolescent to Adult Health (Add Health). Using stratified mixed effects models, we find that the magnitude of friendship influence on smoking initiation differs by school social network density. We additionally find that the contextual factors, rather than peer influence, may be stronger predictors of smoking cessation. The effect estimates of these factors on smoking cessation of also differ by network density.

Extending these results, we conclude with an abstract simulation of the hypothesized mechanisms that contribute to the outcomes of the stratified mixed effects model described previously. We find that network structure and peer influence are sufficient in combination to generate substantial differences in smoking prevalence by urbanicity, sex, and race, among US adolescents. These results provide evidence that support the potential for effect modification by network density on the hypothesized pathway between friendship influence and smoking behavior.

While the field of tobacco control has been traditionally amenable to computational modeling approaches, few studies use computational modeling within an epidemiologic framework to provide support for hypothesized causal pathways that contribute to smoking behavior outcomes. Such perspectives are critical as the tobacco landscape continues to change with the introduction of new products, and as we gain a better understanding of the role that social networks play in the propagation of health behaviors. Through the integration of statistics, computational modeling, and epidemiologic methods, this dissertation seeks to provide insights into the potential causal pathways between various risk factors and smoking behavior outcomes. The results and discussions of this dissertation present potential avenues through which computational modeling can contribute added value to epidemiologic methods, in addition to our understanding of smoking behavior, beyond those of projection and evaluation.

CHAPTER 1

Introduction

The Health, Social, and Economic Impacts of Smoking

50 years since the landmark 1964 Surgeon General's report on smoking and health,[1] cigarette smoking remains the leading preventable cause of death and disability in the United States. Despite substantial evidence of the enormous public health burden of smoking behavior, more than 42 million Americans today continue to smoke, with 5.6 million adolescents under the age of 18 projected to die prematurely if smoking trends were to continue at 2014 rates.[2] Although consistent declines in US smoking prevalence have been observed throughout the past two decades, these reductions have been inadequate, in aggregate, to achieve the Health People 2020 goal of 12% smoking prevalence. Instead, experts estimate a more probable goal of 14% smoking prevalence.[2,3] From an individual health perspective, researchers have established strong relationships between cigarette smoking and a variety of chronic diseases, including cancer, cardiovascular disease, and respiratory disease.[2] With smoking behavior concentrated among those that are disabled, uninsured, and live below the poverty line, the social and economic costs of smoking are also immense.[4] A 2015 study estimates that 170 billion dollars of annual healthcare spending in the US is attributed to illnesses associated with cigarette smoking, with approximately 60% of the attributable spending paid for by publically funded programs, such as Medicare, Medicaid, and Veterans Affairs.[5]

Tobacco Control Overview

Continued health, social, and economic detriments of cigarette smoking behavior in the US has prompted numerous proposals since the 1960s to further decrease the public health burden of smoking. Addictive adult smoking behavior is known to begin with adolescent experimentation, with nearly 90% of adult smokers reporting experimenting with cigarettes before age 18, and 99% of adult smokers reporting experimenting with cigarettes before age 26.[6] A recent Centers for

Disease Control and Prevention (CDC) report indicates that adult smoking cessation rates have remained low for decades, with fewer than one in ten smokers reporting quit success in the past year.[7] This figure is slightly lower than the 12% of smokers who report continued abstinence from smoking for two years in a 1997 study.[8] With adolescent smoking experimentation predicting addictive adult smoking behavior and low adult smoking cessation rates, policies narrowly targeted to discourage adolescent initiation of smoking have proven to be most effective at decreasing the public health burden of smoking. Historically, the most effective tobacco control policy has been to increase the cost of cigarettes, [6,9–11] which provides a two pronged benefit of discouraging adolescent smoking initiation, and encouraging smoking cessation among current smokers.[12] Other effective methods targeting smoking initiation have included banning point-of-sale advertising and implementing school-based intervention programs.[9] However, despite reports of historical lows in adolescent smoking initiation, the elimination of smoking experimentation among adolescents remains elusive.[13] With the large body of evidence supporting the progression of smoking experimentation in adolescence to addictive smoking in adulthood, concerted efforts to understand the interacting risk factors of smoking initiation in adolescents remains paramount.

Interventions tailored for individuals have also demonstrated success in encouraging smoking cessation. For instance, managed care coverage of behavioral counseling and the use of pharmacotherapy methods such as nicotine patches, gums, and inhalers have been shown to increase the likelihood of smoking cessation.[14] These efforts are additionally supported by population-based policies such as media campaigns, cigarette price increases, and smoke free indoor air policies.[2,15] Yet while a number of studies report low population-level smoking cessation rates,[15,16] there is groing evidence suggesting that cessation rates have increased since the 1990s,[17] and have increased among younger cohorts since then.[18]

The various risk factors associated with adolescent smoking behavior range from personal, social, and environmental factors that interact to produce complex feedback systems.[6,19–22] Central to the understanding of smoking initiation in adolescence are the effects of friendship selection and peer influence,[23–26] the latter of which has been demonstrated to have greater effects in adolescence than in adulthood.[27] Given that social factors function within a broader context of regional tobacco control policy differences,[28] their potential interaction with other contextual and individual risk factors[29,30] requires a multifaceted and coordinated approach to understand how to best address continued smoking experimentation in adolescence. Additionally, the evolving

tobacco product landscape creates new uncertainties and complexities in tobacco control policy and research. Recent governmental funding cuts that threaten to eliminate effective tobacco prevention programs at the CDC[31] have resulted in an renewed urgency to develop more and narrowly focused endgame strategies, particularly targeted towards youth, in order to sustain continuing declines in adult smoking prevalence now and into the future.

Electronic Cigarettes

A potentially detrimental barrier for adolescent-focused prevention strategies has been the introduction of electronic cigarettes (e-cigarettes) to the US tobacco market. These products are a diverse product class, although they typically contain a battery-operated vaporizing device that produces inhalable nicotine vapors, along with a range of other additives that differ depending on the brand of e-cigarette.[32] One study in January 2014 counted 466 brands in 7,764 unique flavors of e-cigarette products,[33] with approximately 30-50% of e-cigarette sales occurring through online retail.[34] Although e-cigarettes are currently regulated by the Food and Drug Administration (FDA) through age accessibility, marketing, distribution, and manufacturing restrictions, they are not as stringently regulated as traditional cigarettes.[32] The relatively recent introduction of e-cigarettes means that its long-term effects remain unknown. However, the contentious nature of introducing cigarette-like products with addictive properties to the market has prompted numerous studies exploring the potential short-term effects of e-cigarette use on smoking behavior and health.

Important to the topics of this dissertation, inconsistent results have been found regarding whether e-cigarettes can act as harm reduction products,[35–38] which help current smokers stop smoking, or if they may induce more harm by allowing current smokers sustain nicotine addiction or act as a gateway that encourages traditional smoking initiation among individuals that would have otherwise remained non-smokers.[39–41] One key point of agreement across most published studies exploring e-cigarette use patterns has been that their prevalence and incidence are concentrated among current cigarette smokers and adolescents.[42–45] Between the years 2011-2015, e-cigarette use is reported to have increased 900% among high school students.[32] Concerns related to the potential for e-cigarettes to encourage adolescents to begin smoking through nicotine addiction pathways has fueled a rancorous debate within the tobacco control community weighing in on the merits of an expanding body of research studying the health impacts of e-cigarette use.

Computational Models in Tobacco Control

Changing landscapes of nicotine delivery product use patterns, such as the boom in e-cigarette use prevalence, have created additional dimensions of complexity for tobacco control researchers and experts. In recent decades, advancing computational power provides new avenues to better disentangle the cultural, social, and policy factors that contribute to sustained smoking behavior. Dynamical systems and simulation modeling play an important role in the development of tobacco control policy across a broad range of applications. These models traditionally include quantitative assessments of current and new tobacco control policies, along with smoking prevalence and incidence prediction models.[2] Utilizing computational modeling methods allows researchers to incorporate the dynamic and multifaceted complexities associated with smoking behavior, while additionally considering feedback mechanisms.[46] Various smoking models are widely cited and utilized in the development of effective tobacco control policy, given their utility to predict, project, and describe the complex phenomena associated with smoking behavior.[46–48] They are particularly useful when data is incomplete or missing and expensive to obtain.[49]

Among the most notable example within tobacco models is the Mendez-Warner Model.[50] Published first in 1998, the model simulated adult smoking prevalence using data from the National Health Interview Survey (NHIS) and acted both as an exploratory model of the mechanisms contributing to observed adult smoking prevalence, and as a projection model of future smoking prevalence. It suggested that the perceived plateau in smoking prevalence declines in the early 1990s were likely due to measurement error rather than a stall, and that smoking prevalence would likely decrease into the future because of continually decreasing initiation rates relative to the stable cessation rates that had been historically observed. Mendez and Warner validated this 1998 model in a 2004 paper,[51] demonstrating that the original model provided accurate future predictions of smoking prevalence declines, and reiterated the importance of addressing slowing decreases in smoking prevalence.

The Sim Smoke Model[52–55] is also commonly used as a framework to determine the potential long-term outcomes of various policy initiatives focused on decreasing smoking initiation and prevalence. This model simulates the outcomes of a range of tobacco control policies, such as taxation, smoke-free air laws, strategies to reduce youth access to cigarettes, mass media policies, among others. For each policy module, researchers can modify parameters associated with the magnitude and direction of the expected effect change on smoking initiation and prevalence, and

project the long-term smoking initiation and cessation outcomes of these policies.[2] An additional model, the smoking Control Dynamic Model[56] takes a different perspective on tobacco control, and focuses primarily on the prediction of smoking cessation behavior. Finally, the National Cancer Institute supports the Cancer Intervention and Surveillance Modeling Network (CISNET), which has developed a range of computational models addressing a variety of smoking focused research questions. These models estimate the attributable decreases in US lung cancer mortality[57] and overall mortality[57,58] associated with population declines in smoking, and also provide estimations of birth-cohort specific smoking histories.[18] Other smoking models have focused on the diffusion of adolescent smoking behavior in social networks,[59] and the dynamics of pro- and anti- tobacco policy forces on the patterns of tobacco use.[2] However, most of these models ignore the complex effects of social networks and population feedbacks on the dissemination of smoking behaviors, which can contribute to substantial differences in the patterning of smoking across population groups.

A primary benefit in the application of computational modeling and simulation is the ability to modify parameters of the model to determine a range of possible outcomes. The flexibility of modeling allows researchers to explore counterfactual scenarios and examine how and to what extent the projections of smoking behavior and smoking prevalence can change over time in dynamic populations (e.g., accounting for death and birth). Thus, the challenges of tobacco control creates an environment that is, and has been, especially amenable to computational modeling methods. An overarching theme emerges in a broad overview of prior tobacco control models, which is a focus on quantifying the impact of smoking behavior in various contexts, as measured by smoking prevalence, morbidity, mortality, and economic impacts.[49] Additionally, modeling studies in the tobacco control literature generally seek to capitalize on data that can provide reliable and representative measures of a population.[46] Various data-driven statistical analysis techniques are often used within the frameworks of simulation models, such as choosing parameters that minimize the sum squared differences between the model estimates with empirical data, or other goodness of fit statistics.[2,18,50] Advancing computational power, bigger and better data, and new statistical modeling techniques for parameter estimation has allowed tobacco control modeling to solidify its place in prediction and projection models of smoking behavior that account for a variety of topics, including policy, intervention, and demographic changes. However, the tendency for tobacco control models to focus on the benefits of modeling for outcome prediction and estimation

overshadows the additional benefits of modeling. These benefits include the ability of computational models to test the theoretical mechanisms posited by data-driven methods to better understand the pathways connecting complex risk factors with smoking behavior.

Towards an Interdisciplinary Approach

The success of epidemiology and public health in the study of cigarette smoking has presented a unique opportunity in the evolution of epidemiologic methods and understanding of pathways between the risk factors and health outcomes of smoking behavior, particularly when accounting for new technology and methodologies capable of handling large amounts of data and computation. That is, in the context of such rich literature over decades of focused energy on tobacco control, we can quantitatively test hypothesized causal mechanisms contributing to various facets of smoking behavior that contribute to adult smoking addiction and population level smoking prevalence. The intersection of statistical methods, computational modeling, and epidemiologic principles and theory provide a unique and valuable addition to existing models in tobacco research. When combined, these methods provide the resources to identify and disentangle the behavioral, environmental, and social structures that perpetuate smoking initiation, encourage smoking cessation, modulate response to policies and conversely to tobacco industry marketing, and generate changes to population-level smoking patterns. This dissertation seeks to connect the existing tobacco control literature with computational models that add theoretical value beyond prediction and counterfactual effect estimation.

Dissertation Overview

In a 2008 paper, Josh Epstein suggests 16 reasons other than prediction to build models.[60] This dissertation applies a subset of these reasons towards a better understanding of smoking behavior, specifically, to 1) bound outcomes of plausible ranges; 2) illuminate core uncertainties; 3) illuminate core dynamics; and 4) challenge the robustness of prevailing theory through perturbations. The second chapter of this dissertation presents a computational model that weighs the magnitude of the potential harms and benefits of e-cigarette use from an adult smoking prevalence standpoint, bounded by the possibilities of current population demographics. The recent introduction of e-cigarettes to the US marketplace also offers an opportunity to illuminate the core uncertainties that remain in the e-cigarette literature, and what these uncertainties can

mean for future adult smoking prevalence more generally. Chapter 2 concludes by discussing how computational models that have predictive capabilities can also illuminate the core dynamics that generate observed patterns of smoking prevalence. These results emphasize the mechanisms of the model, rather than the absolute magnitude of model outcomes, which pave the path for further mechanistic explorations in the third and fourth chapters of my dissertation.

Chapters 3 and 4 tie together existing epidemiologic analysis methods with computational models to suggest a new way to substantiate hypothesized causal pathways of social influence on smoking behavior. Chapter 3 uses mixed effects and stratified analyses to identify potential effect modification by network density on the hypothesized causal pathway between adolescent peer influence and future adult smoking behavior. The results from Chapter 3 provide a causal hypothesis that we test in Chapter 4 using an agent-based model (ABM). There, we build a sufficient cause framework for the contextually patterned effects of social influence on smoking behavior. In the concluding chapter, we summarize our findings from Chapters 2 through 4 to suggest a different role for which simulation models, along with data-driven statistical counterparts, can coalesce to contribute added value to the epidemiologic methods literature.

Discussion

The content of this dissertation rests at the interface of epidemiologic methods, statistical methods, computational modeling approaches, and the philosophy of science, seeking to generate new insights into a range of hypothesized causal processes contributing to smoking behavior, while additionally providing approaches that seek to answer how best to dampen smoking initiation and promote smoking cessation in the US. Specifically, the goal of these dissertation chapters is not necessarily to add to the existing (large) body of population- or individual- level prediction models of smoking incidence and prevalence, but rather, to illuminate the mechanisms that drive the outcomes of the existing models of smoking behavior. In contrast to data-driven models, where associations can be confounded by theoretically infinite factors, computational models are closed-solution, where all possible outcomes of the model are a result of the explicitly programmed assumptions. However, this doesn't suggest that outcomes of computational models are known *a priori*. Complexity of a system can emerge from simple mechanisms and model assumptions. The outcomes of these models can vary widely depending on the model parameter values, and must also depend on statistical models to identify the best ways to apply scientific theory of causal

pathways to public health practice. This work additionally seeks to highlight the benefits of systems science modeling in epidemiology, and explicitly demonstrates ways in which carefully crafted computational models can support existing epidemiologic theories and frameworks, and reveal abstract, yet testable, pathways between risk factors and smoking behavior that can be obfuscated in the scientific process of data-driven analysis.

Given the social, cultural, environmental, and economic aspects of smoking epidemiology, adolescent smoking behavior and adult smoking behavior are inexorably intertwined, both in the present and in the future. Interventions should not only focus on adolescents and other specific risk groups, but also find ways to target these risk groups while addressing broader social and policy environments that encourage feedbacks to influence other population groups. In other words, understanding and revealing the mechanisms that contribute to the complex landscape of cigarette smoking within certain risk groups can result in broader changes to smoking behavior beyond that risk group. Identifying and understanding how these mechanisms work will not only guide future empirical research, but also provide insights for policymakers seeking to develop effective tobacco control measures that are narrowly focused towards population subsets that can subsequently benefit broader population health through mechanisms of behavioral diffusion.

References

1. United States Public Health Service. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. 1964.
2. United States Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress A Report of the Surgeon General. *A Rep Surg Gen*. 2014:1081.
3. Mendez D, Warner KE. Setting a Challenging yet Realistic Smoking Prevalence Target for Healthy People 2020: Learning from the California Experience. *Am J Public Health*. 2008;98(3):556-559. doi:10.2105/AJPH.2006.107441.
4. Jamal A, King BA, Neff LJ, Whitmill J, Babb SD, Graffunder CM. Current Cigarette Smoking Among Adults - United States, 2005-2015. *MMWR Morb Mortal Wkly Rep*. 2016;65(44):1205-1211. doi:10.15585/mmwr.mm6544a2.
5. Xu X, Bishop EE, Kennedy SM, Simpson SA, Pechacek TF. Annual Healthcare Spending Attributable to Cigarette Smoking: An Update. *Am J Prev Med*. 2015;48(3):326-333. doi:10.1016/j.amepre.2014.10.012.
6. United States Department of Health and Human Services. Preventing Tobacco Use Among Youth and Young Adults. A Report of the Surgeon General. 2012:1395.
7. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting Smoking Among Adults — United States, 2000–2015. *Morb Mortal Wkly Rep*. 2017;65(52):1457-1464.
8. Gilpin E, Pierce J, Farkas A. Duration of Smoking Abstinence and Success in Quitting. *J Natl Cancer Inst*. 1997;89(8):572-576. doi:10.1093/jnci/89.8.572.
9. Lantz P, Jacobson P, Warner K, et al. Investing in Youth Tobacco Control: A Review of Smoking Prevention and Control Strategies. *Tob Control*. 2000;9(1):47-63. doi:10.1136/tc.9.1.47.
10. Chaloupka FJ, Cummings KM, Morley CP, Horan JK. Tax, Price and Cigarette Smoking: Evidence from the Tobacco Documents and Implications for Tobacco Company Marketing Strategies. *Tob Control*. 2002;11 Suppl 1:I62-I72. doi:10.1136/tc.11.suppl_1.i62.
11. Pierce JP, White VM, Emery SL. What Public Health Strategies are Needed to Reduce Smoking Initiation? *Tob Control*. 2012;21(2):258-264. doi:10.1136/tobaccocontrol-2011-050359.
12. Hyland A, Bauer JE, Li Q, et al. Higher Cigarette Prices Influence Cigarette Purchase Patterns. *Tob Control*. 2005;14(2):86-92. doi:10.1136/tc.2004.008730.
13. Johnston LD, O'Malley PM, Miech RA, Bachman JG. *Monitoring the Future National Survey Results on Drug Use, 1975-2016: Overview, Key Findings on Adolescent Drug Use.* Ann Arbor: Institute for Social Research; 2017.
14. Rigotti NA, Quinn VP, Stevens VJ, et al. Tobacco-control Policies in 11 Leading Managed Care Organizations: Progress and Challenges. *Eff Clin Pract*. 2002;5(3):130-136.

15. Zhu SH, Lee M, Zhuang YL, Gamst A, Wolfson T. Interventions to Increase Smoking Cessation at the Population Level: How Much Progress has Been Made in the Last Two Decades? *Tob Control*. 2012;21(2):110-118. doi:10.1136/tobaccocontrol-2011-050371.

16. Zhuang YL, Gamst AC, Cummins SE, Wolfson T, Zhu SH. Comparison of Smoking Cessation Between Education Groups: Findings from 2 US National Surveys Over 2 Decades. *Am J Public Health*. 2015;105(2):373-379. doi:10.2105/AJPH.2014.302222.

17. Méndez D, Tam J, Giovino GA, Tsodikov A, Warner KE. Has Smoking Cessation Increased? An Examination of the US Adult Smoking Cessation Rate 1990–2014. *Nicotine Tob Res*. 2016. doi:10.1093/ntr/ntw239.

18. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort-specific Smoking histories, 1965-2009. *Am J Prev Med*. 2014;46(2). doi:10.1016/j.amepre.2013.10.022.

19. Thornberry TP. Toward an Interactional Theory of Delinquency. *Criminology*. 1987;25(4):863-892. doi:10.1111/j.1745-9125.1987.tb00823.x.

20. Maxwell KA. Friends: The Role of Peer Influence Across Adolescent Risk Behaviors. *J Youth Adolesc*. 2002;31(4):267-277. doi:10.1023/A:1015493316865.

21. Backinger CL, Fagan P, Matthews E, Grana R. Adolescent and Young Adult Tobacco Prevention and Cessation: Current Status and Future Directions. *Tob Control*. 2003;12 Suppl 4:IV46-V53. doi:10.1136/tc.12.suppl.

22. Lakon CM, Wang C, Butts CT, Jose R, Timberlake DS, Hipp JR. A Dynamic Model of Adolescent Friendship Networks, Parental Influences, and Smoking. *J Youth Adolesc*. 2014;44(9):1767-1786. doi:10.1007/s10964-014-0187-7.

23. McGloin JM, Sullivan CJ, Thomas KJ. Peer Influence and Context : The Interdependence of Friendship Groups , Schoolmates and Network Density in Predicting Substance Use. *J Youth Adolesc*. 2014;43:1436-1452. doi:10.1007/s10964-014-0126-7.

24. Ali MM, Dwyer DS. Estimating Peer Effects in Adolescent Smoking Behavior: A Longitudinal Analysis. *J Adolesc Heal*. 2009;45(4):402-408. doi:10.1016/j.jadohealth.2009.02.004.

25. Alexander C, Piazza M, Mekos D, Valente T. Peers, Schools, and Adolescent Cigarette Smoking. *J Adolesc Heal*. 2001;29(1):22-30. doi:10.1016/S1054-139X(01)00210-5.

26. Ennett ST, Bauman KE, Hussong A, et al. The Peer Context of Adolescent Substance Use: Findings from Social Network Analysis. *J Res Adolesc*. 2006;16(2):159-186. doi:10.1111/j.1532-7795.2006.00127.x.

27. Gardner M. Peer Influence on Risk Taking, Risk Preference, and Risky Decision Making in Adolescence and Adulthood: An Experimental Study. *Dev Psychol*. 2005;41(4):625-635. doi:10.1037/a0026993.

28. Forster JL, Murray DM, Wolfson M, Blaine TM, Wagenaar AC, Hennrikus DJ. The effects of community policies to reduce youth access to tobacco. *Am J Public Health*. 1998;88(8):1193-1198. doi:10.2105/AJPH.88.8.1193.

29. Leyden KM. Social Capital and the Built Environment: The Importance of Walkable Neighborhoods. *Am J Public Health*. 2003;93(9):1546-1551. doi:10.2105/AJPH.93.9.1546.

30. Williams CT, Grier SA, Marks AS. "Coming to Town": The Impact of Urbanicity, Cigarette Advertising, and Network Norms on the Smoking Attitudes of Black Women in Cape Town, South Africa. *J Urban Heal*. 2008;85(4):472-485. doi:10.1007/s11524-008-9286-7.

31. 114th Congress: 1st Session. *S. 1695 [Report No. 114-74]*. 2015.

32. United States Department of Health and Human Services. *E-Cigarette Use Among Youth and Young Adults. A Report of the Surgeon General*. 2016.

33. Zhu SH, Sun JY, Bonnevie E, et al. Four Hundred and Sixty Brands of E-cigarettes and Counting: Implications for Product Regulation. *Tob Control*. 2014;23 Suppl 3(January):iii3-9. doi:10.1136/tobaccocontrol-2014-051670.

34. Noel JK, Rees VW, Connolly GN. Electronic Cigarettes: a New "Tobacco" Industry? *Tob Control*. 2011;20(1):81. doi:10.1136/tc.2010.038562.

35. Polosa R, Caponnetto P, Morjaria JB, Papale G, Campagna D, Russo C. Effect of An Electronic Nicotine Delivery Device (e-Cigarette) on Smoking Reduction and Cessation: A Prospective 6-month Pilot Study. *BMC Public Health*. 2011;11(1):786. doi:10.1186/1471-2458-11-786.

36. Brown J, Beard E, Kotz D, Michie S, West R. Real-world Effectiveness of E-cigarettes when Used to Aid Smoking Cessation: A Cross-sectional Population Study. *Addiction*. 2014;109(9):1531-1540. doi:10.1111/add.12623.

37. Bullen C, Howe C, Laugesen M, et al. Electronic Cigarettes for Smoking Cessation: A Randomised Controlled Trial. *Lancet*. 2013;382(9905):1629-1637. doi:10.1016/S0140-6736(13)61842-5.

38. Choi K, Forster JL. Beliefs and Experimentation with Electronic cigarettes: A Prospective Analysis among Young Adults. *Am J Prev Med*. 2014;46(2):175-178. doi:10.1016/j.amepre.2013.10.007.

39. Coleman BN, Apelberg BJ, Ambrose BK, et al. Association Between Electronic Cigarette Use and Openness to Cigarette Smoking Among US Young Adults. *Nicotine Tob Res*. 2015;17(2):212-218. doi:10.1093/ntr/ntu211.

40. Dutra LM, Glantz SA. Electronic Cigarettes and Conventional Cigarette Use Among US Adolescents. *JAMA Pediatr*. 2014;168(7):610. doi:10.1001/jamapediatrics.2013.5488.

41. Bunnell RE, Agaku IT, Arrazola RA, et al. Intentions to Smoke Cigarettes Among Never-Smoking US Middle and High School Electronic Cigarette Users: National Youth Tobacco Survey, 2011-2013. *Nicotine Tob Res*. 2015;17(2):228-235. doi:10.1093/ntr/ntu166.

42. Delnevo CD, Giovenco DP, Steinberg MB, et al. Patterns of Electronic Cigarette Use Among Adults in the United States. *Nicotine Tob Res*. 2015:ntv237. doi:10.1093/ntr/ntv237.

43. Leventhal AM, Strong DR, Kirkpatrick MG, et al. Association of Electronic Cigarette Use With Initiation of Combustible Tobacco Product Smoking in Early Adolescence. *Jama*. 2015;314(7):700. doi:10.1001/jama.2015.8950.

44. Schmidt L, Reidmohr A, Harwell TS, Helgerson SD. Prevalence and Reasons for Initiating Use of Electronic Cigarettes Among Adults in Montana, 2013. 2014:1-4.

45. Durta L, Glantz S. Electronic Cigarettes and Conventional Cigarette use Among US Adolescents. *JAMA Pediatr*. 2014;168(7).

46. United States Department of Health and Human Services. Appendix 15.1 Tobacco Control Simulation Models. In: *The Health Consequences of Smoking—50 Years of Progress A Report of the Surgeon General*. 2014.

47. Institute of Medicine. *Public Health Implications of Raising the Minimum Age of Legal Access to Tobacco Products*. 2015.

48. National Cancer Insitute. Greater Than the Sum. Systems Thinking in Tobacco Control. *Tob Control Monogr*. 2007;18:1-67.

49. Feirman SP, Donaldson E, Glasser AM, et al. Mathematical Modeling in Tobacco Control Research: Initial Results From a Systematic Review. *Nicotine Tob Res*. 2015;05/16(3):1-14. doi:10.1093/ntr/ntv104.

50. Mendez D, Warner KE, Courant PN. Has Smoking Cessation Ceased? Expected Trends in the Prevalence of Smoking in the United States. *Am J Epidemiol*. 1998;148(3):249-258.

51. Mendez D, Warner KE. Adult Cigarette Smoking Prevalence: Declining as Expected (Not as Desired). *Am J Public Health*. 2004;94(2):251-252. doi:10.2105/AJPH.94.2.251.

52. Levy DT, Nikolayev L, Mumford E. Recent Trends in Smoking and the Role of Public Policies: Results from the SimSmoke Tobacco Control Policy Simulation Model. *Addiction*. 2005;100(10):1526-1536. doi:10.1111/j.1360-0443.2005.01205.x.

53. Levy DT, Nikolayev L, Mumford E, Compton C. The Healthy People 2010 Smoking Prevalence and Tobacco Control Objectives: Results from the SimSmoke Tobacco Control Policy Simulation Model (United States). *Cancer Causes Control*. 2005;16(4):359-371. doi:10.1007/s10552-004-7841-4.

54. Levy D, de Almeida LM, Szklo A. The Brazil SimSmoke Policy Simulation Model: The Effect of Strong Tobacco Control Policies on Smoking Prevalence and Smoking-Attributable Deaths in a Middle Income Nation. *PLoS Med*. 2012;9(11). doi:10.1371/journal.pmed.1001336.

55. Levy DT, Meza R, Zhang Y, Holford TR. Gauging the Effect of U.S. Tobacco Control Policies from 1965 Through 2014 Using SimSmoke. *Am J Prev Med*. 2016;50(4):535-542. doi:10.1016/j.amepre.2015.10.001.

56. Levy DT, Chaloupka F, Gitchell J, Mendez D, Warner KE. The use of Simulation Models for the Surveillance, Justification and Understanding of Tobacco Control Policies. *Health Care Manag Sci*. 2002;5(2):113-120. doi:10.1023/A:1014476916361.

57. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of Reduced Tobacco Smoking on Lung Cancer Mortality in the United States during 1975-2000. *J Natl Cancer Inst*. 2012;104(7):541-548. doi:10.1093/jnci/djs136.

58. Holford TR, Meza R, Warner KE, et al. Tobacco Control and the Reduction in Smoking-Related Premature Deaths in the United States, 1964-2012. *JAMA*. 2014;311(2):164-171. doi:10.1001/jama.2013.285112.

59. Schaefer DR, Adams J, Haas SA. Social Networks and Smoking: Exploring the Effects of Peer Influence and Smoker Popularity through Simulations. *Health Educ Behav*. 2013;40(1 Suppl):24S-32S. doi:10.1177/1090198113493091.

60. Epstein J. Why Model? *J Artif Soc Soc Simul*. 2008;11(4):6. doi:10.1080/01969720490426803.

CHAPTER 2

Electronic Cigarettes: Modeling the Implications for Future Adult Smoking Prevalence

Background

Agent-based models (ABMs) have been used across a wide variety of disciplines to explore the ways in which macro-level phenomena can be driven by micro-level interactions between individuals and their environment. They are especially well suited for modeling individual-to-individual or individual-to-environment feedback mechanisms and adaptation given their bottom-up framework. Many have emphasized the utility of ABMs in the practice of epidemiology and public health when examining causal inference.[1–4] A recent Institute of Medicine (IOM) report highlighting the value of ABMs in tobacco research concluded that the use of ABMs has not been fully explored in the tobacco regulatory space,[5] despite their more common application in other public health areas (e.g., obesity[6–8] and infectious disease[9]).

ABMs differ philosophically and programmatically from traditional compartmental models often used in epidemiological research. Philosophically speaking, ABMs seek to explain outcomes from an individual-level and generative perspective (e.g., feedbacks, adaptations, evolution), while compartmental models explain outcomes from an aggregate (i.e., group) perspective. Thus, ABMs can be more amenable to research questions that focus on individual level behavior, particularly when identifying the important mechanisms of individual behavior that generate the patterns of population outcomes. In this chapter, we develop an ABM of traditional cigarette smoking (smoking) and e-cigarette use to examine how, and to what extent, different levels of e-cigarette effects on adult smoking behaviors (i.e., smoking initiation and cessation) generate changes to population-level smoking patterns under a range of hypothetical scenarios. We simulate the potential population-level outcomes generated from individual-level e-cigarette use under combinations of the following scenarios: 1) e-cigarettes inhibit smoking cessation; 2) e-cigarettes support smoking cessation; 3) e-cigarettes encourage smoking initiation; and

4) e-cigarettes discourage smoking initiation. There are two objectives for this chapter, one focuses on the development of a practical model to address a scientific question, and the other focuses on the mechanisms that generates model outcomes. First, practically, the model building process and exploration presented in this chapter seek to determine to what extent e-cigarette effects on traditional smoking behavior can change cigarette smoking prevalence outcomes under a variety of potential scenarios. Second, methodologically, this model is built with complex mechanistic assumptions, which is disentangled in the results of the model, along with the sensitivity analyses, to better understand our model outcomes.

Introduction

E-cigarette use has increased substantially in recent years, from 1.0% of U.S. adults reporting ever using e-cigarettes in 2009 to 13% in 2013.[10,11] The prevalence of current e-cigarette use among U.S. adults has also grown from 0.3% in 2010 to 6.8% in 2013.[10] To date, almost all research examining e-cigarette use by smoking status has shown that current smokers are more likely to currently use, initiate use of, and experiment with e-cigarettes.[11–15] While the majority of current e-cigarette users are also current cigarette smokers, nearly a third are former or never smokers.[10,16] The rapid increase in e-cigarette use is of growing concern to public health authorities as e-cigarettes are being marketed as smoking cessation aids and safe alternatives to traditional cigarettes[17], without consistent scientific evidence supporting these claims. Some experts believe that e-cigarettes have the potential to reduce the toll of cigarette smoking on population health,[18] while others are concerned that e-cigarettes may weaken tobacco control efforts by renormalizing smoking behavior and serving as a gateway for cigarette smoking initiation among young adults.[19]

While public health surveillance efforts continue to monitor rates of e-cigarette trial and experimentation, particularly among youth[20], longitudinal data evaluating the long-term effects of continued or current e-cigarette use on traditional cigarette smoking cessation or initiation in youth and adult populations are not yet available. Two small randomized controlled studies, both lasting less than two years, suggest that e-cigarette use among cigarette smokers increases smoking cessation relative to placebo e-cigarettes, with efficacy comparable to other cessation aids.[21,22] Conversely, one recent longitudinal study found an association between ever use of e-cigarettes and initiation of cigarette smoking among high school students in Los Angeles, suggesting that

gateway effects may indeed exist.[23] However, most evidence of e-cigarette initiation and cessation relies on self-reported smoking behaviors and convenience samples with known limitations.[11,24]

In the absence of robust, longitudinal empirical data, computational models can support decision-making by investigating a range of possible outcomes under different scenarios,[25] and exploring the underlying properties of a system that contribute to the model results. An ABM approach is particularly amenable to questions of smoking and e-cigarette use given the potential for behavioral feedback dynamics to occur as individuals experiment with new products. The likelihood of e-cigarette initiation differs by individual characteristics such as smoking status and propensity to try e-cigarettes as the product becomes more popular, which can drive changes to e-cigarette use prevalence and incidence at the population level.

Methods

This model is developed using an ABM framework that simulates cigarette smoking as well as e-cigarette use. We used data from the National Health Interview (NHIS)[26], the U.S. Census[27], the CDC[20,28], the Cancer Intervention and Surveillance Modeling Network (CISNET)[29,30], in addition to epidemiological, clinical, and modeling studies.[10,11,16,31–33] This model simulates a population of U.S. adults, aged 18 to 85, and their smoking and e-cigarette use status starting in the year 1997 and ending in 2070. The model has four binary (yes/no) nicotine-use states: 1) exclusive e-cigarette user (e-cigarette user), 2) exclusive cigarette smoker (smoker), 3) dual user of both e-cigarettes and cigarettes (dual user), and 4) never user of either product (never user). A model diagram illustrating all possible nicotine use states and transitions is presented in Figure 2.1.

Individuals' smoking status affects their transitions between nicotine-use states and their probability of death each year. Reflecting our input data sources, we define current e-cigarette use in this model as any individual using e-cigarettes every day or some days. We define current smoking as having smoked at least 100 cigarettes in one's lifetime and currently smoking every day or some days. E-cigarette use is assumed to emulate smoking status, with transitions from never user to current user, to former user. Age-specific smoking cessation rates in the model are based on CISNET estimates, which assume that individuals have quit for at least two years with no relapse. We assume no relapse back to e-cigarette use once an individual quits e-cigarette use. Individuals in the model do not initiate smoking after age 30, consistent with evidence showing

almost no smoking initiation beyond that age.[34] It is also assumed no e-cigarette initiation beyond age 30, given dramatically higher rates of e-cigarette initiation and current use among young adults (age ≤ 25) compared to older adults.[10,11,35] This assumption places a conservative bound on the magnitude of decline to population-level smoking prevalence due to positive e-cigarette effects on smoking cessation. The annual probability of quitting e-cigarettes is set to 0.026, reflecting annual population smoking cessation rate estimates in 2009.[36] Consistent with observed patterns of e-cigarette use, never and former smokers in this model are 15 and 6 times less likely than current smokers, respectively, to initiate e-cigarette use, which closely approximates data on reported e-cigarette prevalence.[10,16] We performed additional sensitivity analyses around our assumption of these values, with the outcomes presented in Appendix A4.2. Finally, this chapter focuses on the impact of e-cigarettes on estimated adult smoking prevalence, so no further negative health effects due to e-cigarette use independent of their effects on cigarette smoking behavior are assumed, despite a growing body of literature suggesting that e-cigarettes may be independently less harmful than smoking.[37]

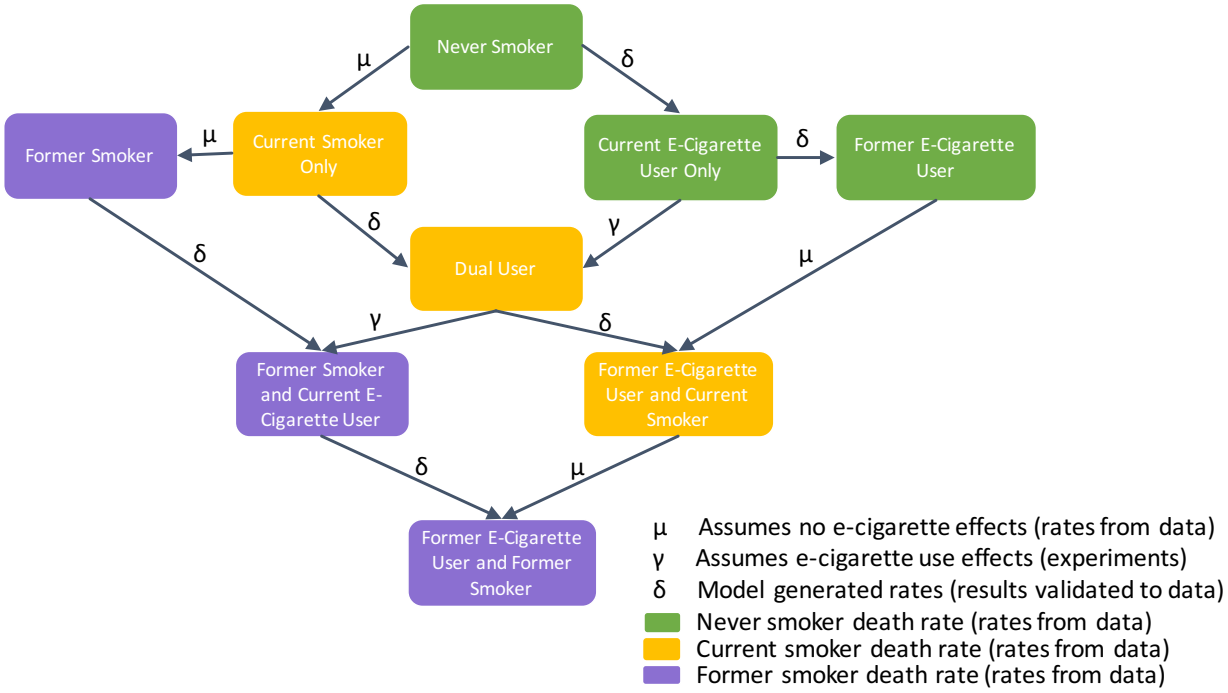


Figure 2.1 Model diagram illustrating all possible nicotine use states (i.e., excludes age, which is considered a separate state), transition pathways, and descriptive transition rates between states. Our model assumes that death rates change only by smoking behavior, and not by e-cigarette use (see model assumptions subsection in the methods section of this chapter).

At each time step, every simulated person ages by one year and either stays in the current nicotine-use state, moves to a different state (e.g., a never smoker starting to smoke), or dies. Movement between nicotine-use states are probabilistic and determined by a combination of: individual age, empirical data on the risk of transition from one state to another (e.g., the annual probability that a 22 year-old non-smoker, current e-cigarette user, initiates smoking), and model parameters representing e-cigarette effects. Every year, new 18 year-olds enter the population through a "birth rate" that reflects the 1997 birth rate as reported by the CDC.[28] In order to account for early youth initiation of e-cigarettes,[10,20] approximately 14% of 18 year-olds enter the population as current e-cigarette users, regardless of smoking status, beginning in 2009. Table 2.1 presents model parameter descriptions used in this chapter, along with the range of values for these parameters that this model explored. The model scenario that generates the lowest smoking prevalence would simulate a 200% increase in smoking cessation rates with a 100% decrease to smoking initiation due to e-cigarettes (i.e., 3 times greater cessation and no smoking initiation). In contrast, the scenario that would generate the highest smoking prevalence would simulate a 100% decrease in smoking cessation rates with a 200% increase in smoking initiation (i.e., no smoking cessation and 3 times greater smoking initiation). Please refer to Appendix A for a complete model description, pseudo-code, as well as a list of all model parameters, equations, and citations for the assumptions and rules governing the model.

To achieve baseline estimates of future smoking prevalence, we calibrated a "smoking only model" to historical and projected smoking estimates in the U.S. using data from NHIS and CISNET.[29,30] For the years 1997-2013, smoking prevalence estimates from this model were within the 95% confidence intervals for smoking prevalence as reported by NHIS for all years except in 2002. Next, to calibrate e-cigarette use prevalence by smoking status to the current literature, we incorporated e-cigarettes in a "baseline" model. The baseline model simulates e-cigarette initiation and cessation using a time-based sigmoidal function representing the rapid uptake of e-cigarettes in the population and eventual plateau due to saturation, analogous to the diffusion of innovations theory often used in systems science research.[38] Using this sigmoid function, the e-cigarette initiation rate among current smokers is low in 2009, two years after the first introduction of e-cigarettes into the U.S. market, and when empirical research of e-cigarette use first begins to emerge.[39] This initiation rate then grows exponentially until 2016, when the initiation rate plateaus to a maximum level that is then held constant until the end of our simulations in 2070 (the timing

and level of the plateau are tunable model parameters). The baseline model assumes no e-cigarette effects on smoking behavior to serve as a platform for experiments examining the independent e-cigarette effects on smoking prevalence. Appendix A also presents full calibration outcomes, discussion of these outcomes, detailed parameter descriptions and values, and additional information about the sigmoidal function referenced in this section.

Table 2.1 Subset of primary model parameters and descriptions.

Parameters†	**Description**
Smoking Cessation Rate	Yearly age-specific smoking cessation rates [1997-2070][28]
Smoking Initiation Rate◊	Yearly age-specific smoking initiation rates [1997-2070]
E-Cigarette Smoking Cessation Effect*	E-cigarette effect on smoking cessation [-100% to 200%]. Effects represent the percentage change relative to baseline, where values less than 0% are decreases in smoking cessation rates (e.g. -50% effect is equal to a reduction of the baseline smoking cessation rate by half); values greater than 0 are increases in smoking cessation rates (e.g. 200% effect is equivalent to 3x the baseline cessation rate)
E-Cigarette Smoking Initiation Effect*	E-cigarette effect on smoking initiation [-100% to 200%]. Effects represent the percentage change relative to baseline, where values less than 0 are decreases in smoking initiation rates (e.g., -50% effect is equal to a reduction of the baseline smoking initiation rate by half); values greater than 0 are increases in smoking initiation rates (e.g., 200% effect is equivalent to 3x the baseline initiation rate)
E-Cigarette Initiation Sigmoid Function◊	E-cigarette initiation rate is derived from a time-based sigmoidal function to mimic the rapid uptake of e-cigarettes with growing use (theory of innovations)
E-cigarette Quit Rate◊	Rate at which e-cigarette users quit using e-cigarettes
Maximum Age of E-Cigarette Initiation◊	Maximum age that an individual can initiate e-cigarette use
Maximum Age of Cigarette Initiation	Maximum age that an individual can initiate traditional smoking behavior[33]
Death Rate◊	Death rates based on smoking status (i.e., relative risk of mortality among former smokers, never smokers, and current smokers)

†All parameter descriptions, citations, and parameter values can be found in the supplementary material
*Experimental parameters
◊Evidence and assumptions derived from a combination of recent peer-reviewed studies

After establishing the baseline model, we performed experiments allowing e-cigarette use to alter the rate of smoking initiation, smoking cessation, or both, to assess the outcomes of the potential harm-inducing and harm-reducing effects of e-cigarettes. While holding the parameters that determine e-cigarette use prevalence by smoking status constant, we modify e-cigarette effects on smoking initiation and cessation with increases of 0% to 200%, or decreases of 0% to 100%, relative to baseline (i.e., no effect) rates. E-cigarette use effects below 100% result in decreases to annual baseline estimates of smoking initiation and cessation, while e-cigarette effects above 100% result in increases to annual baseline estimates of smoking initiation and cessation. For example,

with an e-cigarette cessation effect of 200%, a 30-year-old smoker would increase their annual probability of quitting traditional cigarettes from a baseline of 0.026[29] to 0.078.

To assess the robustness of our results and to examine unexpected changes to the model resulting from the interaction of parameters, we performed sensitivity analyses exploring how the parameters governing the operationalization of our model assumptions changes the model outcomes. We additionally investigated the potential effects of e-cigarettes on adolescent smoking outcomes, and the subsequent implications for adult smoking prevalence. Sensitivity of the model outcomes to parameter settings were explored in scenarios with and without e-cigarettes to identify any potential interaction abnormalities: 1) maximum age at e-cigarette initiation, 2) rate which e-cigarette initiation rates increase over time, 3) maximum probability of e-cigarette initiation, 4) amount of time from e-cigarette introduction to maximum e-cigarette initiation, and 5) smoking cessation rates by birth cohort.

Results

Figure 2.2 shows projected population e-cigarette prevalence by smoking status from 2010 to 2070 using the baseline smoking model (i.e., e-cigarettes have no effect on smoking initiation or cessation), which is fit to match past and present data, projected into the future.[26,31] E-cigarette use outcomes from 2010-2014 are fit to existing e-cigarette use literature by smoking status,[10,16] whereas future projections of e-cigarette use are based on the e-cigarette initiation and cessation parameters generated from the process of model fitting and the assumptions described previously. For all groups, e-cigarette prevalence increases steadily over time except among current smokers whose e-cigarette use prevalence plateaus. Rising population e-cigarette prevalence in the baseline model is primarily driven by current smokers in earlier years and then by former and never smokers in later years. While most e-cigarette users remain former and current smokers, the baseline model also projects a continued rise in e-cigarette use prevalence among never smokers.

Figure 2.3 shows smoking prevalence outcomes resulting from seven hypothetical scenarios of e-cigarette effects on smoking behavior, assuming e-cigarette use patterns described in Figure 2.2. This includes the baseline scenario that assumes no e-cigarette effects on smoking behavior and hypothetical scenarios of e-cigarette effects to the baseline scenario. A 20% decrease in smoking cessation due to the introduction of e-cigarette use (*i.e.*, addiction-sustaining effects), increases smoking prevalence in 2060 by approximately 7.5% compared to baseline smoking

prevalence estimates, bringing smoking prevalence from an initial baseline projection of 13.4% to 14.4%. In the case that e-cigarettes increase smoking initiation by 20% (*i.e.*, addiction inducing "gateway" effects), smoking prevalence would increase by 0.8% in 2060 compared to baseline (13.4% to 13.5%). Under the assumption that e-cigarettes aid smoking cessation, a 20% increase in smoking cessation due to e-cigarettes generates a 6% reduction in smoking prevalence compared to the baseline scenario (13.4% to 12.6%).

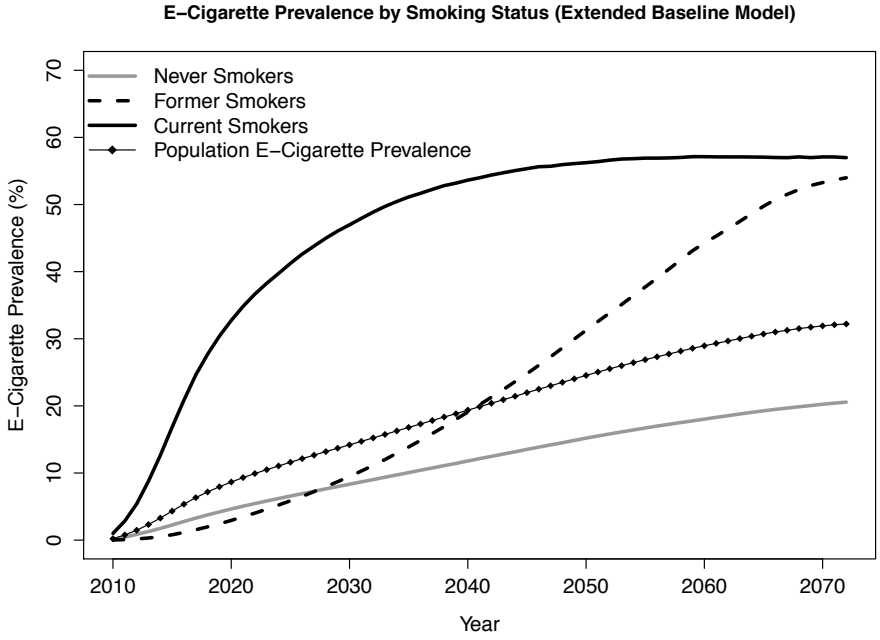


Figure 2.2 E-cigarette use prevalence by smoking status (extended baseline model). Includes e-cigarette initiation and cessation, with no e-cigarette effects on smoking.

In contrast, e-cigarettes would have to increase smoking initiation by over 200% in the absence of any effect on smoking cessation to generate a 6% increase to baseline smoking prevalence. Overall, we observe that e-cigarette effects on smoking cessation, by either increasing or decreasing cessation, generate substantially larger changes to population-level smoking prevalence by 2070 than e-cigarette effects on smoking initiation in this model. Figure 2.4 shows smoking prevalence projections in 2030 and 2060 relative to baseline model outcomes under varying e-cigarette effects on smoking cessation (horizontal dimension) and smoking initiation (vertical dimension). Values above 1.0 (*i.e.*, 100%) are increases to smoking prevalence relative to the baseline scenario, and values below 1.0 are reductions to smoking prevalence relative to baseline. The baseline scenario estimates 14.3% and 13.4% smoking prevalence for the years 2030 and 2060, respectively. There

is an absence of major variation by e-cigarette effects on smoking initiation (along the vertical axis), suggesting that population smoking prevalence is driven primarily by e-cigarette effects on smoking cessation.

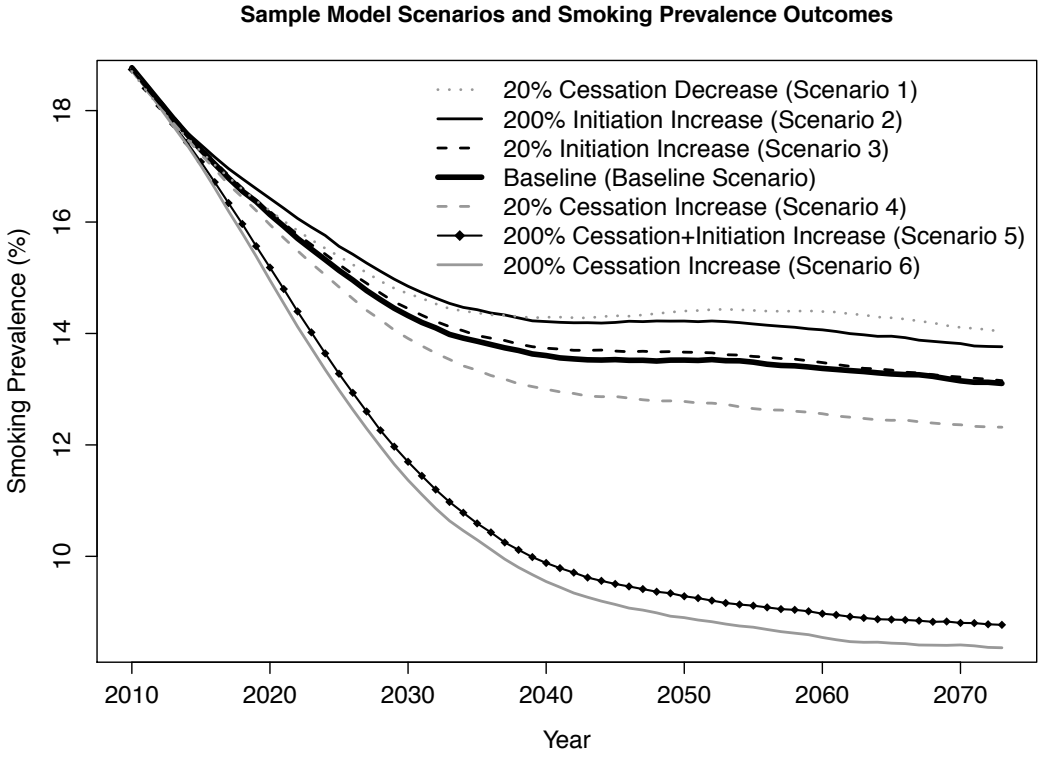


Figure 2.3 Baseline smoking projections and select model scenarios of e-cigarette effects on smoking, 2010 to 2070.

For instance, if e-cigarettes increase both smoking initiation and cessation by 50%, we estimate smoking prevalence to be approximately 90% of baseline estimates in 2060. In other words, despite equal effects on smoking initiation and cessation, e-cigarettes would generate lower smoking prevalence in 2060. If e-cigarettes increase smoking cessation by 100% (i.e. double the likelihood of cessation), assuming baseline smoking initiation values, the smoking prevalence would reduce by 23% relative to baseline. A similar change in e-cigarettes effects on smoking initiation (i.e. they increase smoking initiation by 100%) increases baseline smoking prevalence by 1.03 times, or 3%, when assuming baseline cessation values. The figure also shows greater variation in smoking prevalence in 2060 than in 2030. These patterns emerge because e-cigarette prevalence among never smokers, current smokers, and former smokers changes over time, resulting in changes to the size of exposure groups (i.e., those who are e-cigarette users).

Additionally, projected smoking prevalence declines from 2030 to 2060, thus modifying the relative share of the population that are eligible to quit smoking (smokers) and start smoking (never smokers).

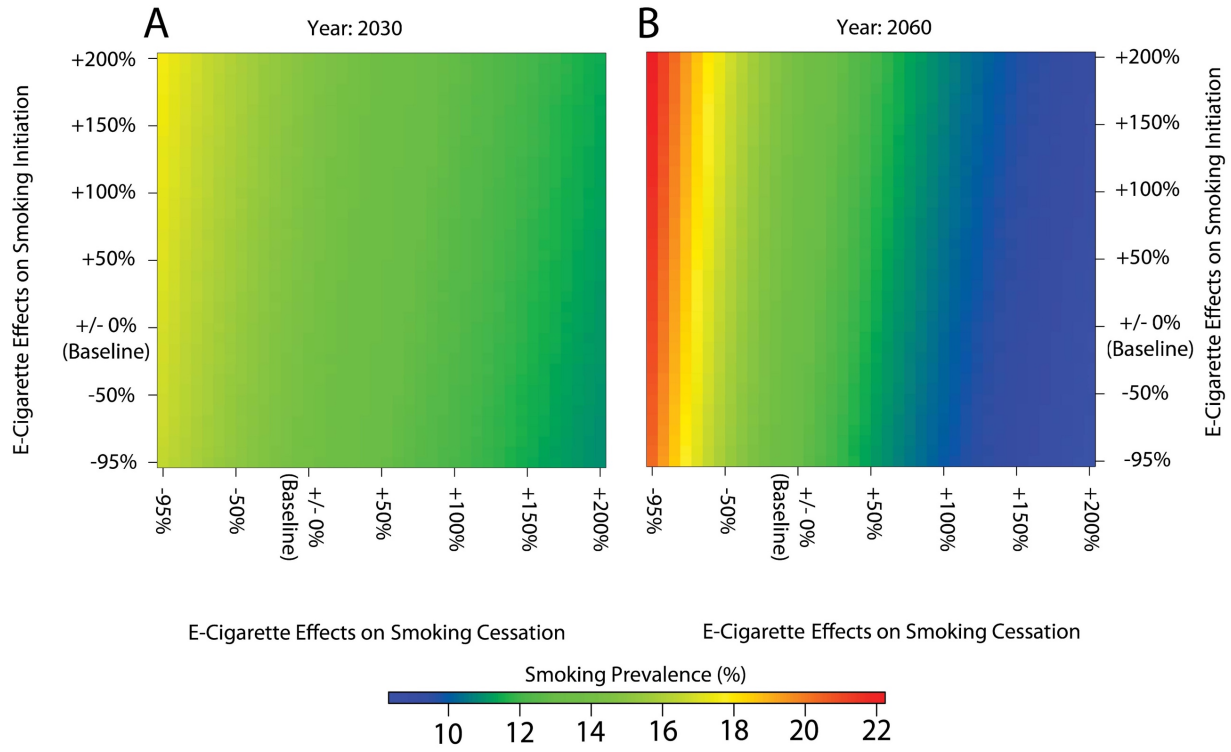


Figure 2.4 Smoking prevalence projections relative to baseline prevalence by e-cigarette effects on smoking initiation and cessation, 2030 and 2060.

Figure 2.5 shows smoking prevalence in 2060 relative to baseline scenario by prevalence of e-cigarette use among never smokers (y-axis) and e-cigarette cessation effects (x-axis). The panels assume 10%, 50%, 100%, and 200% increases to individual smoking initiation rates due to e-cigarette use (i.e. addiction inducing "gateway" effects). Assuming e-cigarette exposure increases the likelihood of smoking initiation, the changing color gradient in the vertical dimension for Figures 2.5A-2.5D shows that projected smoking prevalence is higher relative to baseline with increasing e-cigarette use among never smokers. The panels illustrate that unless e-cigarettes increase smoking initiation by more than 100% (i.e., 2 times baseline smoking initiation), they have almost no noticeable effect on smoking prevalence, even if e-cigarettes prevalence reaches 60% in never smokers. In the case that e-cigarettes increase initiation rates by 200% (Figure 2.5D), approximately 50% of never-smokers would need to use e-cigarettes to increase smoking prevalence by 15% compared to baseline. In this same scenario, if e-cigarette use prevalence is

less than 20% among never smokers, smoking prevalence increases by approximately 4%-6% relative to baseline regardless of the smoking cessation effects of e-cigarettes. E-cigarette effects on smoking cessation (horizontal gradient) become more noticeable in this extreme scenario at higher levels of e-cigarette use prevalence in never smokers, suggesting a feedback between smoking initiation and smoking cessation effects of e-cigarettes, where e-cigarettes can lead to smoking initiation first, and then also to higher rates of smoking cessation.

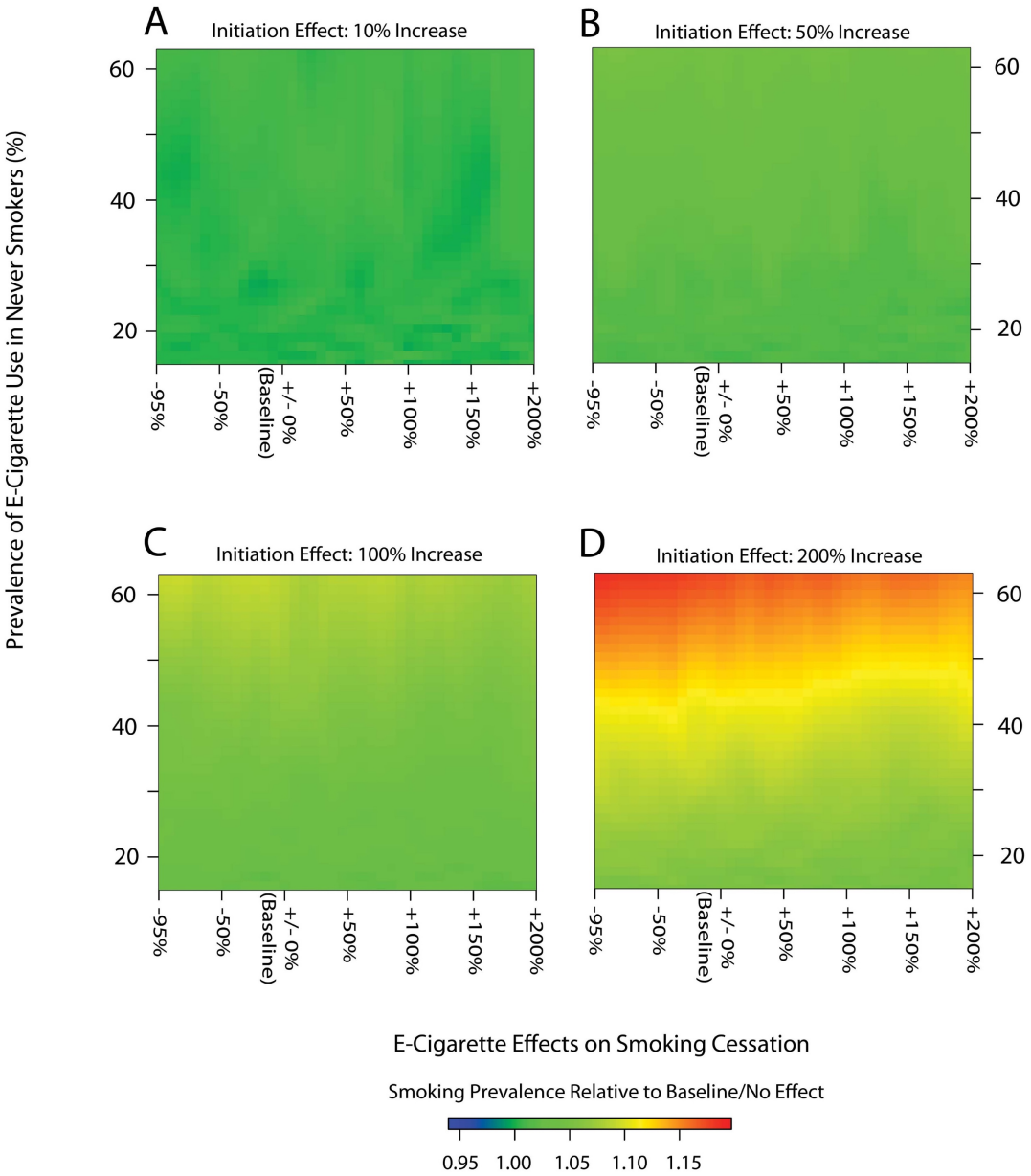


Figure 2.5 Smoking prevalence projections relative to baseline prevalence by e-cigarette prevalence among never smokers and effects on smoking cessation, 2060. Panels A-D show relative changes to smoking prevalence assuming e-cigarettes increase smoking initiation rates by 10%, 50%, 100%, and 200%.

Results from sensitivity analyses across a range of parameters and parameter values suggest that our model assumptions do not substantially change our main results. Given concerns that e-cigarettes behave may act as a gateway for adolescents to begin smoking combustible cigarettes, we ran additional analyses to investigate the range of possible smoking outcomes resulting from e-cigarette use among adolescents before age 18. If we assume that youth e-cigarette use increases the likelihood of smoking initiation before age 18 by 200% (i.e., relative risk of 3.0), compared with never users of e-cigarettes, while having no effect on smoking cessation in the population, our model estimates that adult smoking prevalence would reach 16.6% in 2060. This is a 23.8% increase from the corresponding estimate in our baseline model (13.4%), which assumes no e-cigarette effects on initiation or cessation. Our original adults-only model shows that if e-cigarette use increases the likelihood of smoking initiation by 200% with no impact on cessation, smoking prevalence would reach 14.1% by 2060—a much smaller increase of 4.7% relative to baseline. However, our primary objective was to evaluate population-level outcomes that consider e-cigarette effects on both smoking initiation and cessation. If e-cigarette use increases smoking cessation by 200% among adults, while also increasing smoking initiation by 200% among both adolescents and adults, our model estimates that smoking prevalence would drop to 10.3% by 2060—a 23.1% decrease compared with baseline. Thus, even when the model accounts for large initiation effects whereby youth and adult smoking initiation rates escalate due to e-cigarettes, the cessation effects of e-cigarettes on adult smoking offset these undesirable "gateway" effects. This is consistent with the results of the portion of this study that study focuses primarily on adults.

The largest variation in smoking prevalence outcomes due to e-cigarettes occur when using birth cohorts earlier than the one used in the baseline model (1970). These cohorts were observed to have much lower rates of smoking cessation[30], and thus generated smoking prevalence trends much higher than current and future predictions.[31,34,40] The sensitivity analyses generated similar (relative) results to our main model results when exploring a range of values for parameters governing the maximum age at e-cigarette initiation, rates of adoption of e-cigarettes, the period over which e-cigarette adoption occurs, and the amount of time from e-cigarette introduction to maximum e-cigarette initiation. We also tested the sensitivity of our model to the assumption of decreasing age-specific initiation over time. We began by setting age-specific smoking initiation rates static to 1997 values. Then, we performed parameter sweeps across a range of e-cigarette effects on smoking initiation and smoking cessation. We observed that, relative to decreasing

initiation rates, static and high initiation values could be responsible for a greater range of smoking prevalence values by 2060 due to e-cigarette effects on smoking behavior (Figure A3). However, we continue to observe larger cessation effects of e-cigarettes relative to e-cigarette initiation effects on population smoking prevalence over time. From this figure, we also observe that a higher static initiation rate also results in larger changes to the initiation effects of e-cigarettes. These patterns were dampened when assuming decreasing smoking initiation rates over time (i.e., in the vertical axis of the heatmap). Additional discussion of sensitivity analyses and their results are available in the Appendix A and Figures A3-A10 of this dissertation.

Discussion

Under a variety of hypothetical scenarios of the possible effects of e-cigarettes on smoking behavior, our model shows that smoking prevalence is far more sensitive to e-cigarette effects on smoking cessation than on smoking initiation. Additionally, given current values of population smoking initiation and cessation and the relative prevalence of e-cigarette use between never, current and former smokers, if e-cigarettes induce smoking in never smokers, even small increases in smoking cessation due to e-cigarettes would counteract the potential negative impact on overall population smoking prevalence. Finally, if e-cigarettes decrease smoking cessation by allowing current smokers to continue smoking, population smoking prevalence could increase considerably.

The results from our model are largely due to three main mechanistic factors that arise from our assumptions. First, we assumed relatively high rates of e-cigarette initiation among current smokers compared to never smokers in light of current evidence from the literature.[10,16] Accordingly, there are more smokers than never smokers using e-cigarettes in the simulated population, which means that there are more smokers susceptible to benefit from potential smoking cessation effects of e-cigarettes compared to the number of never smokers that could be affected by their potential effects on smoking initiation. Second, declining smoking initiation rates in the baseline scenario[5,12,29] generate small effect outcomes of e-cigarettes on smoking initiation rates among never smokers. That is, we multiply e-cigarette effects in the model by the underlying smoking initiation and cessation rates (*e.g.* 1.5 times initiation rate). Given declining smoking initiation rates, even a 200% increase in smoking initiation results in relatively small absolute changes to annual age-specific smoking initiation probabilities. Third, smoking prevalence is more sensitive in the short-term to changing smoking cessation rates because there are greater time

delays between an increase in smoking initiation rates and their eventual impact on the number of smokers in the population. Thus, e-cigarettes would have to act as an extremely effective gateway to cigarette smoking to increase smoking prevalence substantially, and never smokers would have to use e-cigarettes much more than the current evidence suggests.

Complex feedbacks occur when investigating e-cigarette effects on smoking behavior, as presented in our results. If e-cigarettes solely undermine smoking cessation efforts, our results suggest that we may experience a substantial increase in smoking prevalence as more never smokers use e-cigarettes, regardless of e-cigarette effects on smoking initiation. That is, if these never smokers eventually become smokers, they would be less likely to quit smoking due to their e-cigarette use, thus raising population smoking prevalence. However, if e-cigarettes both increase smoking initiation and smoking cessation, the effects on smoking initiation would have to be extremely large (i.e., increase over 200%) to offset even small cessation effects on smoking (Figures 2.5A- 2.5D).

The results and discussion of this study should be considered with several limitations in mind. First, our results and interpretation emphasize the general patterns produced by the model, and not the actual values, due to the challenges of quantifying e-cigarette effects, and uncertainty surrounding the sparse longitudinal data that currently exist. Values presented in this paper are outcomes relative to our baseline model and represent model- and parameter- specific relative estimates. Therefore, these results serve only as an educated guess of the potential impact of e-cigarettes on future adult smoking prevalence. Second, our model used a variety of data sources that provide estimates of e-cigarette initiation and cessation values that can range widely across various reports and are challenging to measure accurately. Third, our model does not explore any potential direct and independent health effects of e-cigarette use. Fourth, our results are largely dependent on the low and decreasing smoking initiation rate in the U.S. Our outcomes are not applicable to countries or settings with relatively high, stable, or increasing smoking prevalence. Finally, we made simplifying assumptions about e-cigarette use behavior and smoking behavior to account for high variability in the e-cigarette data, and improve interpretability of our model and its outcomes. While these assumptions do not meaningfully change our primary conclusions, future work should continue to explore and refine these assumptions as further data become available.

Our study contributes to a growing literature that examines systems-level feedbacks and interactions relevant to smoking behavior that could not otherwise be explored using conventional statistical modeling research methods.[26,29–31] In light of the recent IOM report specifically emphasizing the utility of ABMs for decision-making related to tobacco policies,[5] the results of this model may be of interest to those considering how simultaneous e-cigarettes and alternative tobacco product effects for individuals might translate into broader changes to population cigarette smoking patterns. These study results can also provide useful insights given the uncertain regulatory environment surrounding e-cigarettes. The U.S. Food and Drug Administration (FDA) has extended its regulatory authority over additional tobacco products including e-cigarettes[41] to subject their distribution to Tobacco Control Act provisions, including sale restrictions to minors, label requirements to include nicotine warnings, and finally, prohibitions on vending machines sales, except in establishments that never admits minors (e.g., casinos).[42] In the context of our results, these new regulations should further decrease the use of e-cigarettes among never smokers, thereby diminishing the impact that any "gateway" influence that e-cigarettes may have on traditional smoking behavior and also population level smoking prevalence.

In exploring the possible net effects of e-cigarettes on the population by simulating such risks and benefits with respect to smoking status, this model provides useful insights into how the distribution of e-cigarette use behavior can change population smoking prevalence outcomes due to e-cigarette effects on smoking behavior at the individual level. This study also may provide useful information for future tobacco regulation. However, to precisely determine the eventual net impact of e-cigarettes on smoking prevalence, researchers must continue to empirically evaluate the effects of e-cigarettes on smoking initiation and cessation. Robust longitudinal[23] studies that assess the consequences of e-cigarette use for smoking behaviors remain paramount. As these data become available, modeling can serve as a framework to assess the potential impact of e-cigarettes under varying scenarios of use prevalence and their effects on smoking patterns. Given the potential for e-cigarettes to sustain addiction by promoting dual use with cigarettes, the continued pattern of high e-cigarette use among current smokers remains a major public health concern. In this instance, we find that under current patterns of e-cigarette use by smoking status and a range of hypothesized e-cigarette effects on smoking behavior, it is unlikely that the potential gateway effects of e-cigarettes will substantially increase smoking prevalence, unless they also reduce smoking cessation rates.

Conclusions

The primary outcome of this study demonstrates two uses of computational modeling in the study of smoking epidemiology. First, the outcomes of our model quantify the potential effects of e-cigarettes across a spectrum of hypothetical scenarios. These outcomes are generated by interacting mechanisms that contribute to the second focus of the study, which is to illuminate core dynamics of e-cigarette influence on smoking behavior. We find that the patterning of e-cigarette use by smoking status will contribute substantially to their potential effects on smoking prevalence, although our results are skewed towards cessation effects in the most likely scenarios (i.e., given historical trends of smoking prevalence and e-cigarette initiation). Computational modeling to explore a range of potential outcomes of interventions is common in the public health literature. Less common, however, is using observed empirical patterns to hypothesize how assumptions and mechanisms can interact to generate feasible model outcomes. Given the lack of empirical data on e-cigarette initiation and cessation across smoking states, we use aggregate population estimates, drawn from previous studies, as parameter settings in the model. In following chapters, we use empirical data on an adolescent US population to further explore the ways in which computational models can integrate with traditional statistical methods in epidemiologic research to build causal hypotheses.

References

1. Resnicow K, Page SE. Embracing Chaos and Complexity: A Quantum Change for Public Health. *Am J Public Health*. 2008;98(8):1382-1389. doi:10.2105/AJPH.2007.129460.

2. Auchincloss AH, Diez Roux A. A New Tool for Epidemiology: The Usefulness of Dynamic-agent Models in Understanding Place effects on Health. *Am J Epidemiol*. 2008;168(1):1-8. doi:10.1093/aje/kwn118.

3. Diez Roux A. Complex Systems Thinking and Current Impasses in Health Disparities Research. *Am J Public Health*. 2011;101(9):1627-1634. doi:10.2105/AJPH.2011.300149.

4. Marshall BDL, Galea S. Formalizing the Role of Agent-based Modeling in Causal Inference and Epidemiology. *Am J Epidemiol*. 2015;181(2):92-99. doi:10.1093/aje/kwu274.

5. IOM (Institute of Medicine). *Assessing the Use of Agent-Based Models for Tobacco Regulation*. Washington, DC: The National Academies Press. 2015.

6. Zhang D, Giabbanelli PJ, Arah OA., Zimmerman FJ. Impact of Different Policies on Unhealthy Dietary Behaviors in an Urban Adult Population: An Agent-Based Simulation Model. *Am J Public Health*. 2014;104(7):1217-1222. doi:10.2105/AJPH.2014.301934.

7. Orr MG, Galea S, Riddle M, Kaplan GA. Reducing Racial Disparities in Obesity: Simulating the Effects of Improved Education and Social Network Influence on Diet Behavior. *Ann Epidemiol*. 2014;24(8):563-569. doi:10.1016/j.annepidem.2014.05.012.

8. Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux A. An Agent-Based Model of Income Inequalities in Diet in the Context of Residential Segregation. 2013;40(3):303-311. doi:10.1016/j.amepre.2010.10.033.An.

9. Epstein JM, Goedecke DM, Yu F, Morris RJ, Wagener DK, Bobashev G V. Controlling Pandemic Flu: The Value of International Air Travel Restrictions. *PLoS One*. 2007;2(5):e401. doi:10.1371/journal.pone.0000401.

10. Mcmillen RC, Gottlieb MA, Shaefer RM, Winickoff JP, Klein JD. Trends in Electronic Cigarette Use Among U.S. Adults : Use is Increasing in Both Smokers and Nonsmokers. 2014:1-8. doi:10.1093/ntr/ntu213.

11. Pepper JK, Brewer NT. Electronic Nicotine Delivery System (Electronic Cigarette) Awareness, Use, Reactions and Beliefs: A Systematic Review. *Tob Control*. 2013;23(5):1-10. doi:10.1136/tobaccocontrol-2013-051122.

12. Hughes K, Bellis MA, Hardcastle KA, et al. Associations Between E-cigarette Access and Smoking and Drinking Behaviours in Teenagers. *BMC Public Health*. 2015;15(1):1-9. doi:10.1186/s12889-015-1618-4.

13. Grana R, Benowitz N, Glantz SA. E-cigarettes: A Scientific Review. *Circulation*. 2014;129(19):1972-1986. doi:10.1161/CIRCULATIONAHA.114.007667.

14. Delnevo CD, Giovenco DP, Steinberg MB, et al. Patterns of Electronic Cigarette Use Among Adults in the United States. *Nicotine Tob Res*. 2015:ntv237. doi:10.1093/ntr/ntv237.

15. Brose LS, Hitchman SC, Brown J, West R, McNeill A. Is the Use of Electronic Cigarettes while Smoking Associated with Smoking Cessation Attempts, Cessation and Reduced Cigarette Consumption? A Survey with a 1-year Follow-up. *Addiction*. 2015:n/a-n/a. doi:10.1111/add.12917.

16. Zhu S-H, Gamst A, Lee M, Cummins S, Yin L, Zoref L. The Use and Perception of Electronic Cigarettes and Snus Among the U.S. Population. *PLoS One*. 2013;8(10):e79332. doi:10.1371/journal.pone.0079332.

17. Cobb NK, Brookover J, Cobb CO. Forensic Analysis of Online Marketing for Electronic Nicotine Delivery Systems. *Tob Control*. 2013:1-4. doi:10.1136/tobaccocontrol-2013-051185.

18. Abrams D. Promise and Peril of E-Cigarettes: Can Disruptive Technology Make Cigarettes Obsolete? *JAMA*. 2014;311(2):135-136. doi:10.1016/j.addbeh.

19. Dutra LM, Glantz SA. Electronic Cigarettes and Conventional Cigarette Use Among US Adolescents: A Cross-Sectional Study. *JAMA Pediatr*. 2014;94143(7):1-8. doi:10.1001/jamapediatrics.2013.5488.

20. Centers for Disease Control and Prevention. Tobacco Use Among Middle and High School Students -- United States 2013. *Morb Mortal Wkly Rep*. 2014;63(45):1021-1026.

21. Bullen C, Howe C, Laugesen M, et al. Electronic cigarettes for smoking cessation: a randomised controlled trial. *Lancet*. 2013;382(9905):1629-1637. doi:10.1016/S0140-6736(13)61842-5.

22. Caponnetto P, Campagna D, Cibella F, et al. EffiCiency and Safety of an eLectronic cigAreTte (ECLAT) as Tobacco Cigarettes Substitute: A Prospective 12-Month Randomized Control Design Study. *PLoS One*. 2013;8(6):e66317.

23. Leventhal AM, Strong DR, Kirkpatrick MG, et al. Association of Electronic Cigarette Use With Initiation of Combustible Tobacco Product Smoking in Early Adolescence. *Jama*. 2015;314(7):700. doi:10.1001/jama.2015.8950.

24. McRobbie H, Bullen C, Hartmann-Boyce J, Hajek P. Electronic Cigarettes for Smoking Cessation and Reduction. *Cochrane Database Syst Rev*. 2014;12:CD010216. doi:10.1002/14651858.CD010216.pub2.

25. Cerdá M, Tracy M, Ahern J, Galea S. Addressing Population Health and Health Inequalities: The Role of Fundamental Causes. *Am J Public Health*. 2014;104(SUPPL. 4):609-619. doi:10.2105/AJPH.2014.302055.

26. National Center for Health Statistics. *National Health Interview Survey 1997-2013*.; 1997. http://www.cdc.gov/nchs/nhis.htm. Accessed April 27th, 2014.

27. U.S. Census Bureau. *Profiles of General Demographic Characteristics (1997-2010)*.; 2010. http://www.census.gov/data.html. Accessed February 11th, 2014.

28. Centers for Disease Control and Prevention. *Table 1-1, "Live Births, Birth-Rates, and Fertility Rates, by Race: United States, 1909-2000"* 2005. http://www.cdc.gov/nchs/data/statab/t001x01.pdf. Accessed February 11th, 2014.

29. Holford TR, Meza R, Warner KE, et al. Tobacco Control and the Reduction in Smoking-Related Premature Deaths in the United States, 1964-2012. *JAMA*. 2014;311(2):164-171. doi:10.1001/jama.2013.285112.

30. Holford TR, Levy DT, McKay LA, et al. Patterns of Birth Cohort-Specific Smoking Histories, 1965-2009. *Am J Prev Med*. 2014;46(2):e31-7. doi:10.1016/j.amepre.2013.10.022.

31. Vugrin ED, Rostron BL, Verzi SJ, et al. Modeling the Potential Effects of New Tobacco Products and Policies: A Dynamic Population Model for Multiple Product Use and Harm. *PLoS One*. 2015;10(3):e0121008. doi:10.1371/journal.pone.0121008.

32. Bullen C, Howe C, Laugesen M, et al. Electronic Cigarettes for Smoking Cessation: A Randomised Controlled Trial. *Lancet*. 2013;382(9905):1629-1637. doi:10.1016/S0140-6736(13)61842-5.

33. Sanders-Jackson AN, Tan ASL, Bigman CA., Henriksen L. Knowledge About E-Cigarette Constituents and Regulation: Results From a National Survey of U.S. Young Adults. *Nicotine Tob Res*. 2014:1-8. doi:10.1093/ntr/ntu276.

34. United States Department of Health and Human Services. *The Health Consequences of Smoking—50 Years of Progress A Report of the Surgeon General*. 2014.

35. Durta L, Glantz S. Electronic Cigarettes and Conventional Cigarette use Among US Adolescents. *JAMA Pediatr*. 2014;168(7).

36. Pizacani BA, Maher JE, Rohde K, Drach L, Stark MJ. Implementation of a Smoke-Free Policy in Subsidized Multiunit Housing: Effects on Smoking Cessation and Secondhand Smoke Exposure. *Nicotine Tob Res*. 2012;14(9):1027-1034. doi:10.1093/ntr/ntr334.

37. Shahab L, Goniewicz ML, Blount BC, et al. Nicotine, Carcinogen, and Toxin Exposure in Long-Term E-Cigarette and Nicotine Replacement Therapy Users. *Ann Intern Med*. 2017;24(18):442-448. doi:10.7326/M16-1107.

38. Sternman JD. S-Shaped Growth: Epidemics, Innovation Diffusion, and the Growth of New Products. In: *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill Higher Education; 2000.

39. Noel JK, Rees VW, Connolly GN. Electronic Cigarettes: A New "Tobacco" Industry? *Tob Control*. 2011;20(1):81. doi:10.1136/tc.2010.038562.

40. Mendez D, Warner KE, Courant PN. Has Smoking Cessation Ceased? Expected Trends in the Prevalence of Smoking in the United States. *Am J Epidemiol*. 1998;148(3):249-258.

41. Food and Drug Administration. Deeming Tobacco Products To Be Subject to the Federal Food, Drug, and Cosmetic Act, as Amended by the Family Smoking Prevention and Tobacco Control Act; Regulations on the Sale and Distribution of Tobacco Products and Required Warning Statements for Tobac. *Fed Regist*. 2014.

42. United States Department of Health and Human Services. *E-Cigarette Use Among Youth and Young Adults. A Report of the Surgeon General*. 2016.

Appendix A: Supplemental Material for Chapter 2

A1 Summary of Appendix A

A1.1 Content Overview

This appendix is divided into five sections: A1) Summary of the material presented in this appendix; A2) Modeling methodology; A3) Model parameterization; A4) Model Outcomes; and A5) Programming. Section A1 describes the purpose of the supplementary material and provides an overview of its content. Section A2 discusses the use of agent-based modeling (ABM) in our study and how the potential effects of e-cigarettes on smoking behavior present a research problem amenable to an ABM approach. Section A3 details the parameters used to operationalize mortality, smoking initiation, smoking cessation, e-cigarette initiation, e-cigarette cessation, and other processes in the model. Section A4 describes model outcomes, specifically as they relate to validating model output against empirical data, and discusses the results of our sensitivity analyses. Section A5 presents model pseudo code, and additional notes about programming the model. The content presented here is referenced throughout the main text of the dissertation with the appropriate numbered header for each subsection.

A1.2 Purpose

This appendix is written with three main goals in mind: 1) To provide additional information regarding both the modeling methodology and the validation of the model presented in Chapter 2 of the this dissertation; 2) To further discuss the modeling approach, and detail the specifics of model validation not presented in the main text of the chapter; and 3) To provide technical programming guidance and model process equations so that readers are able to reproduce the model and its outcomes as they are presented in the chapter.

A2 Methodology Discussion

A2.1 Why ABM?

An ABM approach is utilized in this chapter for 3 main reasons:

1) The generative and modular nature of this ABM leaves room for further model development examining dynamics at the individual level as more data about e-cigarettes become available. The purpose of this initial model was to demonstrate the variety of possible outcomes dependent upon the effects that e-cigarettes may have on traditional smoking behavior. This model can be extended to incorporate other tobacco products, like smokeless tobacco, snuff, snus, social network and environmental effects on e-cigarette use;

2) The simplicity of a bottom-up approach from an object-oriented perspective provides greater model clarity and modularity. While a traditional compartmental model would have generated similar results, the structural approach typically used in dynamical systems models would involve programming the set of individual-traits and their various levels of heterogeneity, leading to enormous numbers of explicit model states.[1] From a practical standpoint, ABMs are also an aggregate of N compartmental models operating in parallel, where N represents the number of individuals. Specifically, our agent-specific traits, unique to every individual in our model includes: a) smoking status (former, never, current); b) e-cigarette status (former, never, current); c) dual user (former, current); d) never user of e-cigarettes and cigarettes; e) age (18-85); f) probability of smoking initiation; and g) probability of e-cigarette initiation. The object-oriented programming approach in the backend of the ABM method allows us to store these traits and trait histories to follow any model individual's experience trajectory, giving us more insights into the mechanistic assumptions that generate our outcomes;

3) Our goal was to explore how individual-level e-cigarette use changes population-level smoking prevalence. Smoking status, probability of smoking initiation, and probability of e-cigarette initiation are either dynamic and/or heterogeneous outcomes/states across the individuals in our model, a problem well-suited for ABMs.

A3 Model Parameterization

A3.1 E-Cigarette Initiation and Cessation Rate

We assume that the probability of e-cigarette initiation among never, former, and current smokers follows a diffusion of innovations sigmoid function.[2] This function is also described in

detail in Tables A1 and A2, and more broadly in Chapter 2. Parameters for the slope of the curve (ecigNetSlope), the maximum probability of initiation among smokers (eCigProbMax), years until maximum e-cigarette initiation probability is reached (timeToMaxECigInit), and years until e-cigarettes are introduced in the model (modelBurnIn), are used in by the sigmoid function to generate e-cigarette incidence by smoking status (i.e., never, current, and former smoker). This is done in order to approximate current e-cigarette use by smoking status as reported in recent literature.[3–5]

Figure A1 shows the e-cigarette initiation sigmoid curve for former, never, and current smoker initiation rates by year using baseline parameter values. After the year 2016, the probability of e-cigarette initiation by smoking status remains constant through the end of the model in 2070. Current smokers have the highest probability of e-cigarette initiation and serve as the reference group for e-cigarette initiation among former smokers and never smokers. For each time point on the sigmoid curve, if an individual is a never smoker, the probability of initiating e-cigarette use for that individual is the e-cigarette initiation rate on the curve divided by the parameter divECigNeverSmoker (equal to 15 in our baseline model). This divisor may be an overestimate given recent data showing never smokers are approximately 30 times less likely to use e-cigarettes than current smokers.[4] Nonetheless, in the context of our results, this assumption is conservative (i.e., allows for larger negative effects of e-cigarettes on smoking initiation relative to baseline). Similarly, current e-cigarette prevalence among former smokers was reported as 6 times less than that of current smokers – if an individual is a former smoker, the probability of initiating e-cigarette use is the e-cigarette initiation rate for smokers, divided by the parameter: divECigFormerSmoker. For simplicity and because of scarce data on e-cigarette cessation patterns, we assumed e-cigarette cessation rates to emulate traditional smoking cessation rates (e.g., 0.026 at baseline, Table A1).

A3.2 Smoking Cessation and Initiation Rates

We calibrated the model using smoking cessation data from the 1970 birth cohort, the most recent cohort with all of the data available through the Cancer Intervention and Surveillance Modeling Network (CISNET) website.[6] Though we relied on 1970 birth cohort quit rates for the main analysis, we also assess CISNET estimates of age-specific (ages 18-85) smoking cessation rates for those born in the years 1940, 1950, 1960, and 1970, corresponding to the parameter smokeQuitCohort. The year 1940 corresponds with smokeQuitCohort = 1, 1950 with

smokeQuitCohort = 2, etc. Earlier birth cohorts generally have lower smoking cessation rates, so we performed additional simulations to examine smoking prevalence outcomes resulting from these lower smoking cessation rates as part of our sensitivity analyses. We found that the 1970 cohort had cessation rates that generated smoking prevalence levels for 2013-2070 that most closely resembled available projected adult smoking prevalence data.[7–9] Using cessation rates from birth cohorts earlier than 1970 combined with recent smoking initiation estimates generated higher than expected smoking prevalence values than those reported by NHIS[10] and other studies[7–9]. The lower panels in Figures A4 – A9 present sensitivity analyses of birth cohort effects on smoking prevalence (i.e., the parameter "smokeQuitCohort").

We used survival analysis (i.e., the cumulative hazard function) to calculate smoking initiation rates based on reported NHIS smoking prevalence among 18-24 year olds from 1997 to 2013. Our calculation assumes that the smoking initiation rate for those ages 0-12 years old is zero. This is consistent with research indicating that smoking uptake can occur as early as age 12, with most initiation occurring by age 18.[11–13] The survival and hazard equations are as follows:

$$S(t) = e^{-H(t)}, \text{ where } H(t) = \int_0^t \lambda(u)du, \text{ and } t \text{ denotes age}$$

$$h(t) = -\frac{S'(t)}{S(t)}$$

We assume a smoking initiation of 0 from age 0 to 12, and a constant rate from that age forward (λ) until age 30, when smoking initiation is disallowed, such that:

$$H(t) = \lambda(t - 12)$$

$$S(t) = e^{-\lambda(t-12)}$$

Therefore:

$$\lambda = \frac{-\log(\mathrm{S(t)})}{t - 12}$$

We then performed simple linear regression on these values to project smoking initiation rates into the year 2027 (30 years after model initialization), after which initiation rates remain constant through 2070. These decisions were made to best approximate historical estimates and future projections[6,8,14] of smoking prevalence in the US. Figure A2 presents the regression line,

along with the corresponding slope and intercept values used to project future smoking initiation rates.

A3.3 Death and "Birth" Rates

Mortality rates for smokers and former smokers are determined by using reported values of relative risk of death among current (smokerDeathRiskRelative) and former smokers (formerSmokerDeathRiskRelative) compared to never smokers. We used data from the US Census, the Human Mortality Database, and the Lee-Carter method to calculate age- and year-specific death rates among never smokers[15–18]. In order to achieve population equilibrium and to approximate observed prevalence for current and former smoking, we used an annual birth rate of 14.2 per 1,000 persons[19] (i.e., replacement of 18 year olds in the model) and examined a range of values for relative risks of all-cause mortality for current and former smokers. We determined the relative risk for all-cause mortality to be 2.9 for current smokers and 1.5 for former smokers. For example, in our model, a 19 year never smoker in 1998 has a 0.000895 probability of death. For a 19-year-old never smoker, this probability of death is 0.0025955.[15] These values are within 95% confidence bounds of all-cause mortality for smokers and former smokers reported by Freedman et al.[20], and Lynch et. al.[21], respectively.

A3.4 E-Cigarette Effects

The magnitude of the effect of e-cigarettes on smoking cessation and smoking initiation are driven by parameter values dualUseQuitMultiplier and ecigSmokeInitInc which range from 0.0 to 3.0. These values are multipliers applied to baseline annual probabilities of smoking cessation or initiation. Values above 1 increase smoking initiation or smoking cessation probabilities relative to baseline values. Likewise, values below 1 decrease smoking initiation or smoking cessation probabilities relative to baseline. Baseline probabilities were taken directly from CISNET data by cohort.[22] These values are presented as percentages in the main paper for simplicity (i.e. 100% decrease in value to 200% increase in value). For example, a value of 1 is equivalent no decrease or increase to smoking initiation or cessation, a value of 1.5 is equivalent to a 50% increase to smoking initiation or cessation, and a value of 0.85 is a 15% decrease to smoking initiation or cessation. Finally, Table A2 presents all parameters described in Appendix Section A3, and includes details of how these parameters are operationalized in our model.

A4 Model Outcomes

A4.1 Model Validation

Our study objective was to estimate the effects of e-cigarettes on smoking prevalence relative to baseline projections of smoking prevalence in the absence of e-cigarettes. All aspects of this model are approximations of potential effects given the current state of knowledge on smoking and e-cigarette use. We reviewed a range of studies on patterns of e-cigarette use and validated model outcomes against e-cigarette prevalence estimates between 2010 and 2014. A large proportion of our estimates fell within 95% confidence intervals reported in one or more studies.[3,4,23,24] Table A3 provides model generated outcomes compared to data sources available at the time of model creation. Most our baseline smoking estimates are within the 95% confidence intervals for NHIS reported smoking prevalence from 1997 to 2013, except for 2002. Our model prevalence estimates for 2002 were slightly lower than that reported by NHIS. We speculate that this lower estimate is primarily due to increases in tobacco product marketing occurring between 2001 and 2002 that this model does not account for. However, our smoking prevalence estimates return to the NHIS 95% confidence interval bounds after 2002. We projected a smoking prevalence of 12.8% by 2070—consistent with an upper level projection of a recent IOM report on raising the minimum age of cigarette smoking[17], and comparable to projected smoking prevalence by 2050 as estimated by Vugrin et. al.[8]

To illustrate the extreme worst- and best-case scenarios of e-cigarette use effects on smoking prevalence, we projected estimates of e-cigarette use prevalence such that baseline estimates of e-cigarette effects (harm-reducing or harm-inducing) are likely overestimates (Figure 2.2). Pending additional observational data, e-cigarette use may increase beyond the scenario extremes examined in our study, though they may also fall short of our e-cigarette prevalence estimates. Given the possible overestimation of e-cigarette prevalence, we explored how e-cigarette use prevalence could ultimately affect our outcomes (Figure 2.5). In addition, we mapped initiation effects of e-cigarettes on smoking behavior against the prevalence of e-cigarette use among never smokers in the Results section of Chapter 2.

A4.2 Sensitivity Analyses

In Figures A4 – A9, we show sensitivity analyses for different e-cigarette initiation and cessation effect levels varying the following parameters: maximum age of e-cigarette initiation

(ageStopECigInit), years until maximum e-cigarette initiation is reached (timeToMaxECigInit), annual probability of e-cigarette cessation (eCigProbMax), the slope of the sigmoid function for e-cigarette initiation (ecigNetSlope), e-cigarette initiation rate divisor for never smokers relative to current smokers (divECigNeverSmoker), and smoking cessation rates across different birth cohorts (smokeQuitCohort). The colored bar legend represents smoking prevalence, and axes are labeled according to the parameters used within the model (See Table A1). The largest differences in smoking prevalence are generated by experiments examining age- and year- specific smoking cessation probabilities by birth cohort. This property is discussed further in Chapter 2. The remaining differences in smoking prevalence generated from the entire set of sensitivity analyses range from a 0.01 to 0.6 absolute difference in smoking prevalence. Due to the comparative nature of the model (i.e., e-cigarette effects on smoking behavior compared to the baseline model), the differences in absolute smoking prevalence from these sensitivity analyses are negligible in the context of our study, and do not affect our conclusions. In short, we are more concerned with the directionality and magnitude of the relative outcomes, and not the absolute differences of these outcomes.

To better understand the range of outcomes that our model generates resulting from our assumptions, we explore the sensitivity of smoking prevalence in our model at the extreme values of all combinations of the following model parameters (Figure A10): ecigNetSlope, timeToMaxECigInit, eCigProbMax, and their interaction with the maximum age of e-cigarette initiation allowed in the model on the horizontal axis (ageStopECigInit). From this analysis, we find that extreme variations and combinations of our parameters do not change the smoking prevalence values substantially by ageStopECigInit at baseline or when e-cigarettes only increase smoking initiation rates. However, we observe a slight downward trend in smoking prevalence by ageStopECigInit, and greater minimum and maximum smoking prevalence differences when e-cigarettes result in greater smoking cessation rates. The results of this analysis suggest that our baseline model outcomes and results are conservative estimates of the relative advantage of e-cigarette cessation effects over e-cigarette initiation effects. In other words, the downward trend by ageStopECigInit indicates that we would observe even lower smoking prevalence (even at the extreme values of other parameters) than we report in our results when allowing individuals over the age of 30 to initiate e-cigarettes. Although we believe it is likely that older adults over the age of 30 could use e-cigarettes as a smoking cessation aid, the results from this sensitivity analysis

indicate that allowing individuals in our model to initiate e-cigarettes after the age of 30 would further reinforce our findings of the relative strength of e-cigarette cessation effects compared to e-cigarette initiation effects on smoking behavior.

A5 Programming

A5.1 Programming Notes

- Model was programmed using Python
- Model output was analyzed using Python-Pandas and R
- Primary assumptions: No relapse of smoking or e-cigarette use after quitting
- No smoking initiation after age 30
- No e-cigarette initiation after age 30
- Variable names used here are described in the Table A1 and Table A2

A5.2 Model Pseudo Code

```
OBJECT agent;
       age;
       current_smoking_status;
       smoking_history;
       current_electronic_cigarette_status;
       electronic_cigarette_history;
       alive_or_dead;
PROGRAM smoking_model;
     initialize agents to 1997 US population age and smoking status demographics;
     for every year from 1997 to 2075:
        repeat for all agents in the model:
                if (age >= 85) or probability of death by age, smoking status, history:
                   die;
                if smoker:
                   if not e-cigarette user:
                       start e-cigarettes at P(ecigInitSmoker);
                       quit smoking at P(smokingCess);
                    if e-cigarette user:
                         quit smoking at P(smokingCess) * dualUseQuitMultiplier;
                         quit e-cigarettes at P(ecigQuitProb);
                if former smoker:
                    if e-cigarette user:
                         quit e-cigarettes at P(ecigQuitProb);
```

```
            if not e-cigarette user:
                start e-cigarettes at P(ecigInitSmoker)/divECigFormerSmoker;
        if never smoker:
            if e-cigarette user:
                start smoking at P(somkingInit)*ecigSmokeInitInc;
                quit e-cigarettes at P(ecigQuitProb);
            if not e-cigarette user:
                start e-cigarettes at P(ecigInitSmoker)/divECigNeverSmoker;
    increment age;
birth new 18 year olds at set birth rate to maintain stable population counts;
calculate model and agent statistics;
write model and agent statistics to outputs;
clear model and agent statistics for this step;
```

*Table A1 Baseline Model Parameters.**

Parameters	Values	Description
ecigNetSlope	0.20	Slope of sigmoid function governing e-cigarette initiation
dualUseQuitMultiplier	1.0	Multiplier applied to baseline smoking quit rates for dual users. While this is set to 1.0 (i.e., no effect) in the baseline model, this is an experimental parameter that takes a range of values. Figure 2.4, Figure 2.5, and Section A4.2 of this appendix provide results for these experiments.
smokerDeathRiskRelative	2.9	Relative risk of mortality for smokers compared to never smokers
formerSmokerDeathRiskRelative	1.5	Relative risk of mortality for former smokers compared to never smokers
eCigProbMax	0.23	Maximum probability of e-cigarette initiation of current smokers
modelBurnIn	12.0	Number of years until e-cigarettes are introduced (1997-2009)
ecigQuitProb	0.026	Annual probability of e-cigarette cessation
ecigSmokeInitInc	1.0	Multiplier on baseline smoking initiation rates among e-cigarette users. While this is set to 1.0 (i.e., no effect) in the baseline model, this is an experimental parameter that takes a range of values. Figure 2.4, Figure 2.5, and Section A4.2 of this supplement provide results for these experiments.
smokeQuitCohort	4.0	CISNET smoking cessation rates based on cohort.† Here, cohort 4.0 is equivalent to the 1970 birth cohort
divECigNeverSmoker	15.0	E-cigarette initiation rate divisor for never smokers
divECigFormerSmoker	6.0	E-cigarette initiation rate divisor for former smokers
timeToMaxECigInit	7.0	Time (in years) to maximum e-cigarette initiation rate after their introduction
ageStopECigInit	30.0	Maximum age of e-cigarette initiation

*Parameter values are those used in the baseline model calibrated to e-cigarette use prevalence among never, former, and current sokers in 2010 and 2013, and adult smoking prevalence from 1997 to 2013. Note: We use parameter variable names in the model pseudo code.
†See Section A3.2 Smoking Cessation and Initiation Rates for additional information

*Table A2 Model equations and parameters.**

Variable and Calculation Method	Function	Description
smokingInit	-0.0006*MIN(time,25) + 0.024	The probability of becoming a smoker dependent on the model time step (year)
Survival rates by prevalence (calculated from historical prevalence and then fitted to linear function)		
smokingCess	Table	The probability of quitting dependent on the model time step (year)
CISNET annual smoking cessation probabilities from the 1970 male and female cohort (averaged)[6,14]		
deathRate	Table	The probability of death dependent on the model time step (year)
US death rates table (past, present, and future projections) using Census data[16], the Human Mortality Database[15], and the Lee-Carter method[17,18]		
birthrate	14.2 per 1000 persons	Stable birth rate in the model
Reported by the CDC in 1997[19]		
smokingPrevAtInit	Table	Age-specific population level smoking prevalence in 1997 -- Model individuals are initialized at these age-specific levels
NHIS smoking prevalence values by age groups 18-24, 25-44, 45-64, and 65+ [10]		
ecigInitSmoker	$\frac{\text{eCigProbMax}}{1+e^{(-\text{timeVal}*\text{ecigNetSlope})}}$ where: $\text{timeVal} = (\text{time} - \text{modelBurnIn}) * \left(\frac{20}{\text{timeToMaxECigInit}}\right) - 10$	The probability of a current smoker becoming an e-cigarette user dependent on the model time step (year)
Sigmoidal function from the theory of innovations, based in time[2]		

*Model parameters are used within equations that govern transition probabilities between nicotine use states and life-cycle states (i.e., age, alive/dead). Note: These equation variable names are also used in the pseudo code; some equation variables are generated from parameter values.

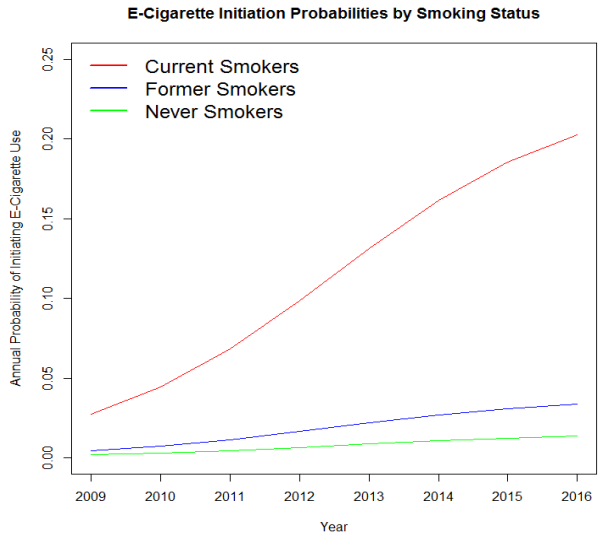
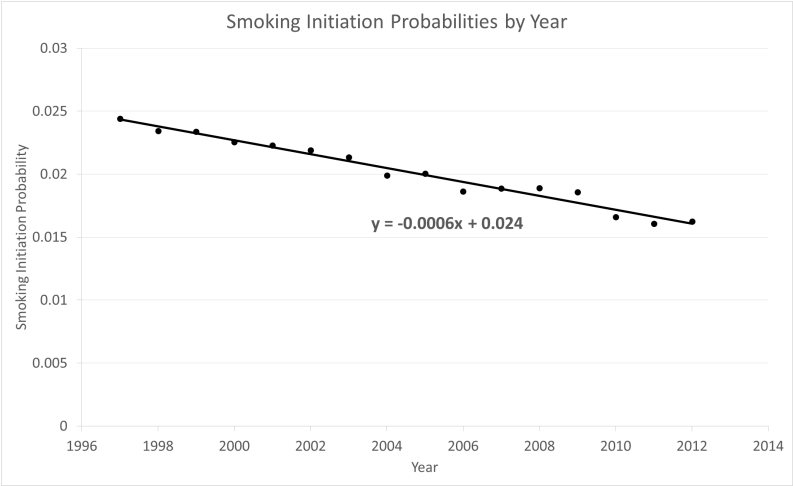


Figure A1 Sigmoid functions determining the probability of e-cigarette initiation by smoking status.



F*igure A2 Smoking initiation probabilities generated by the cumulative hazard function, using linear regression to project initiation rates by year.*

*Table A3 Model generated outcomes compared to data sources.**

Current E-Cigarette Use Prevalence by Smoking Status (%)				
Year	2010	2011	2012	2013
Model Population Prevalence	0.8	1.5	2.4	3.4
McMillen et al. Population Prevalence[4]	0.3	0.8	2.6	6.8
Schmidt et al. Population Prevalence[3]	NA	NA	NA	1.3
Zhu et al. Population Prevalence[23]	NA	NA	1.4	NA
Model Current Smokers	0.1	0.1	8.7	12.6
McMillen et al. Current Smokers	1.4	5.0	10.8	30.3
Zhu et al. Current Smokers	NA	NA	6.3	NA
Model Former Smokers	0.1	0.2	0.3	1.0
McMillen et al. Former Smokers	0.3	0.1	1.1	5.4
Zhu et al. Long Term Former Smokers	NA	NA	0.2	NA
Zhu et al. Recent Former Smokers	NA	NA	6.1	NA
Model Never Smokers	0.5	1.0	1.5	2.0
McMillen et al. Never Smokers	0.1	0.1	0.1	1.4
Zhu et al. Never Smokers	NA	NA	0.04	NA
Adult Smoking Prevalence (%)				
Year	1997	2002	2007	2012
Model Smoking Prevalence	24.2	21.1	19.1	17.6
NHIS Smoking Prevalence	24.7	22.5	19.7	18.1
NHIS (95% CI)	(24.1,25.3)	(21.9, 23.1)	(19.0, 20.6)	(17.5, 18.7)

*McMillen et al. and Schmidt et al. define current use as "everyday" or "someday" use of e-cigarettes. Zhu et al. define current use as "use of e-cigarettes in the past 30 days."

*Note: For Figures A5-A9, colored bars indicate population-level smoking prevalence, with corresponding estimates for the range of colors provided to the right of every figure.

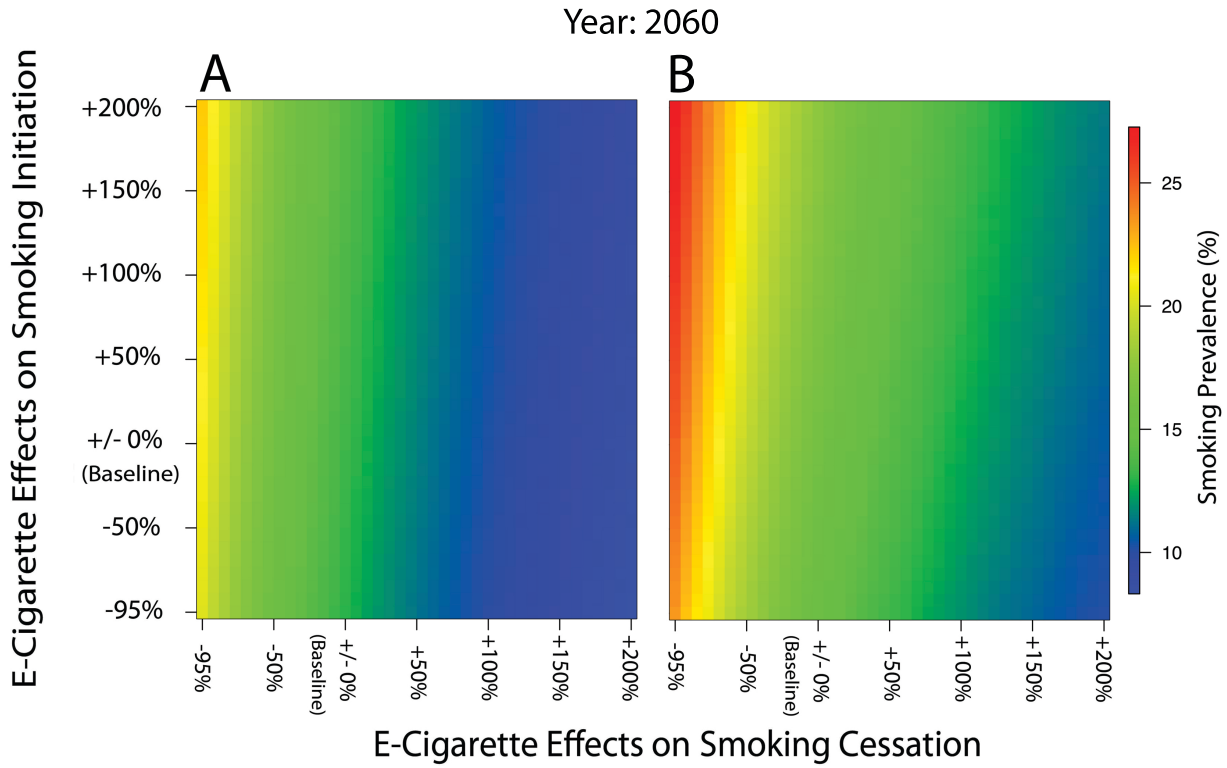


Figure A3 Smoking prevalence outcomes in 2060 by e-cigarette effects on smoking initiation and cessation. Panel A illustrates model outcomes of the baseline model with decreasing smoking initiation. Panel B illustrates model outcomes of a sensitivity analysis with static smoking initiation at 1997 rates.

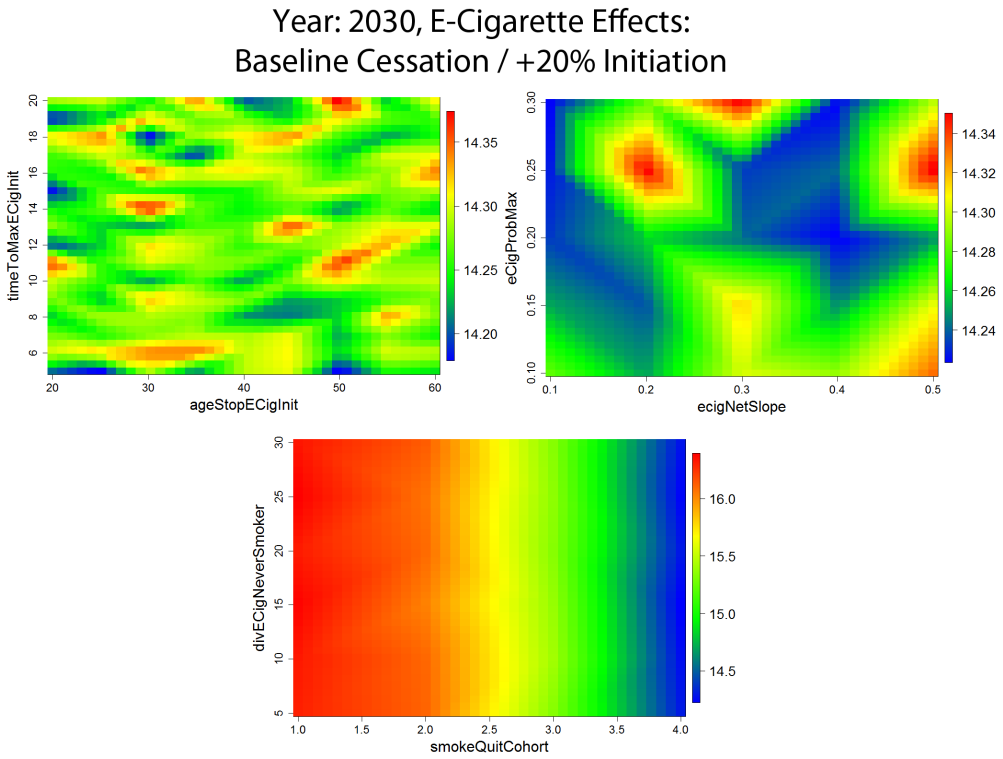


Figure A4 Smoking prevalence outcomes in 2030: e-cigarette use results in a 20% increase to individual-level smoking initiation probability and does not affect smoking cessation, relative to baseline.

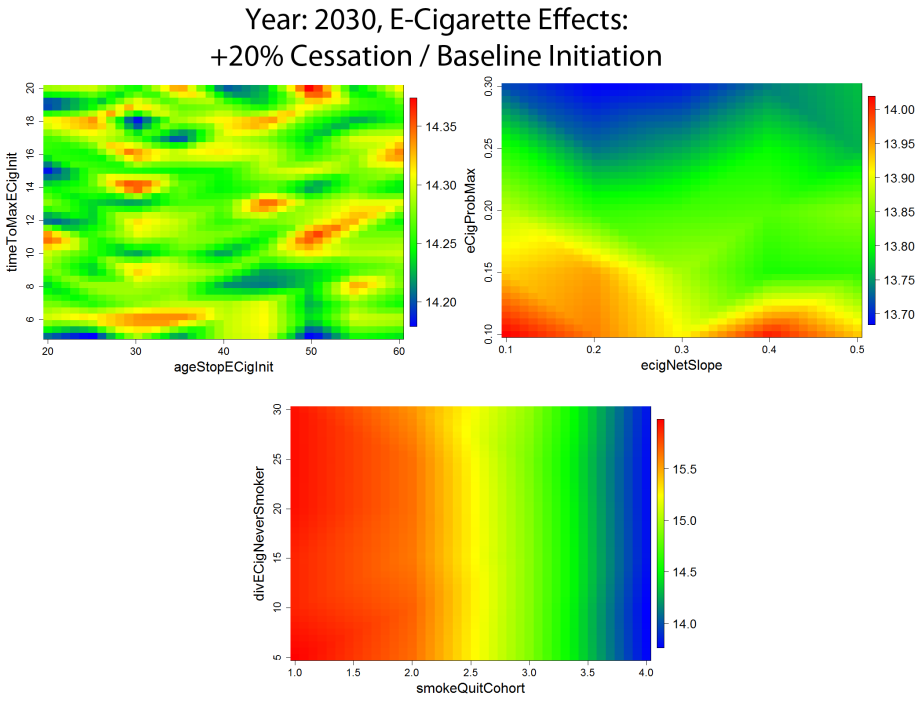


Figure A5 Smoking prevalence outcomes in 2030: e-cigarette use results in a 20% increase to individual-level smoking cessation probability and does not affect smoking initiation, relative to baseline.

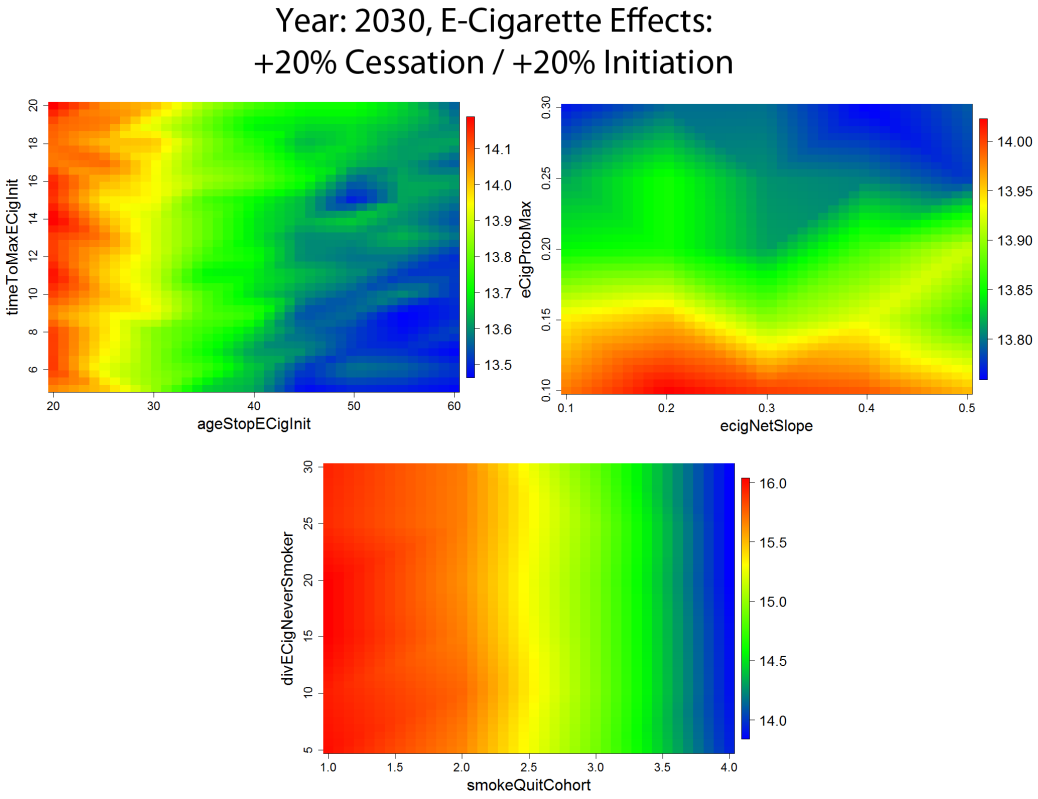


Figure A6 Smoking prevalence outcomes in 2030: e-cigarette use results in a 20% increase to both individual-level smoking cessation probability and individual-level smoking initiation probability, relative to baseline.

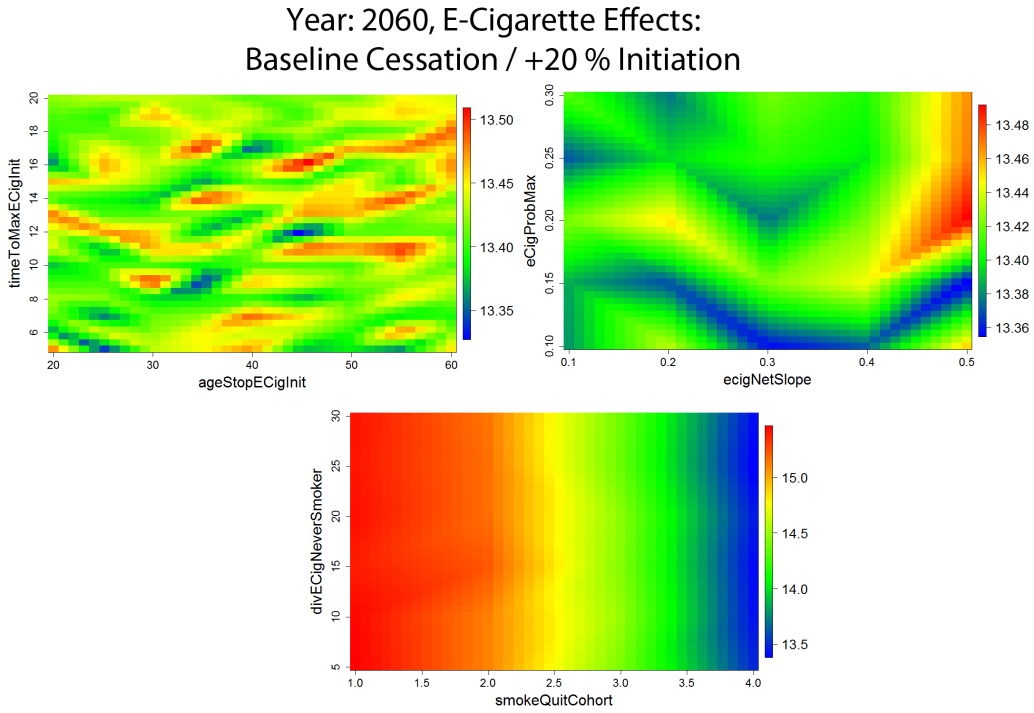


Figure A7 Smoking prevalence outcomes in 2060: e-cigarette use results in a 20% increase to individual-level smoking initiation probability, and does not affect smoking cessation, relative to baseline.

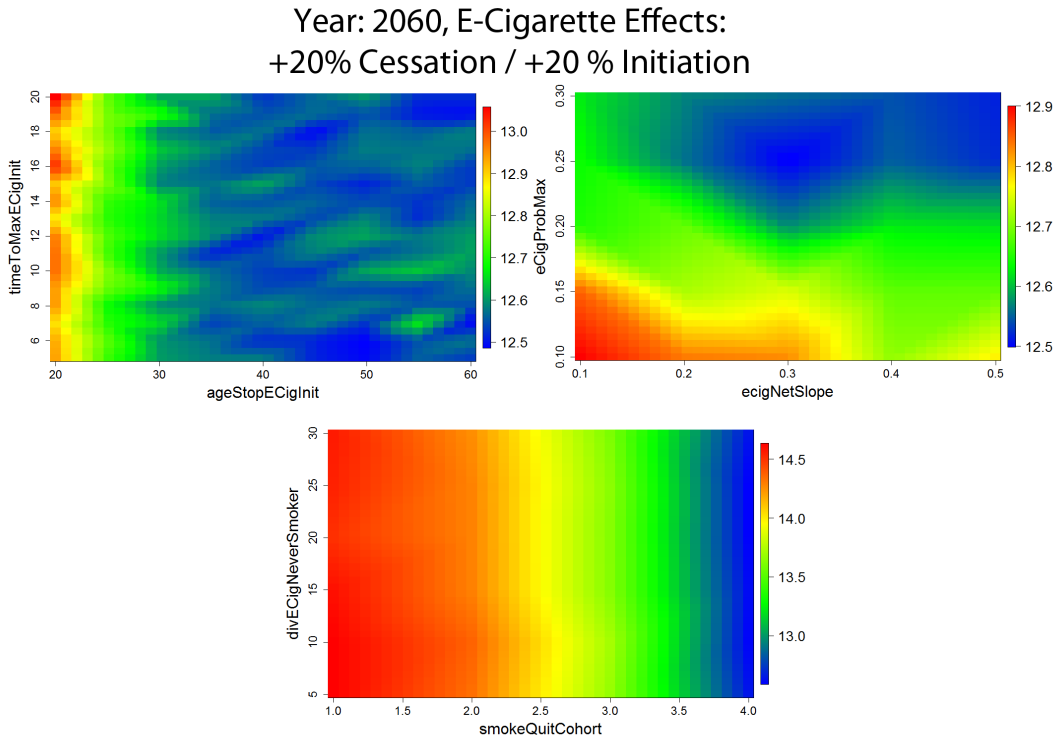


Figure A8 Smoking prevalence outcomes in 2060: e-cigarette use results in a 20% increase to both individual-level smoking cessation probability and smoking initiation probability, relative to baseline.

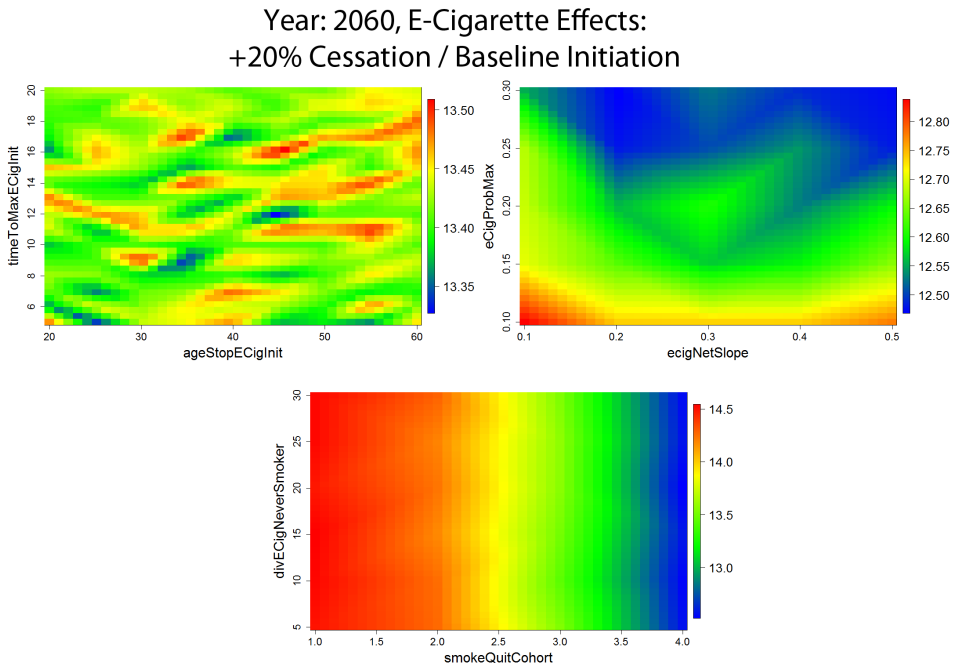


Figure A9 Smoking prevalence outcomes in 2060: e-cigarette use results in a 20% increase to individual-level smoking cessation probability and does not affect smoking initiation, relative to baseline.

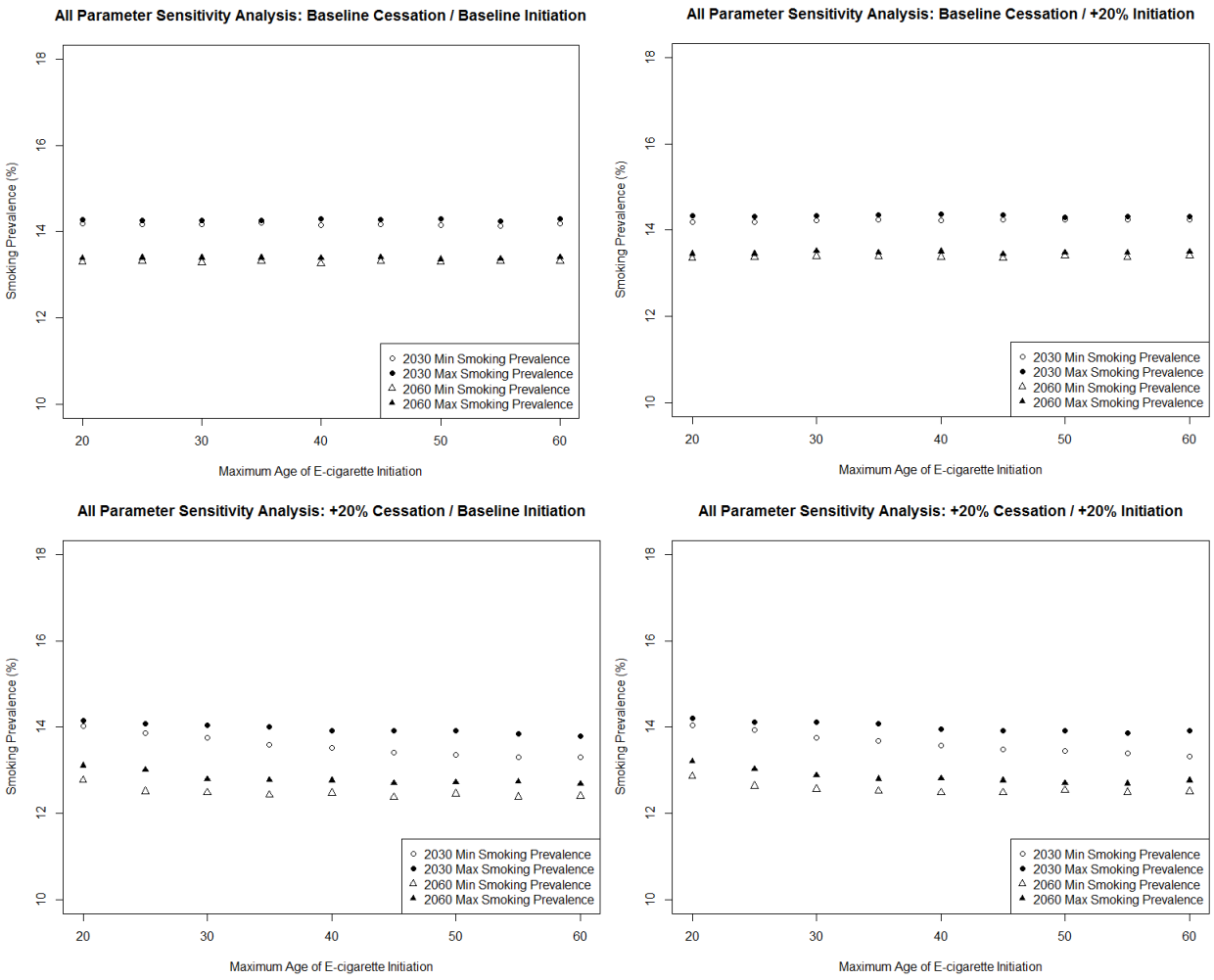


Figure A10 Smoking prevalence outcomes in 2030 and 2060 using extreme ranges of all parameters by a range of maximum age at e-cigarette initiation values (x-axes). Upper Left: Baseline Model. Upper Right: e-cigarette use results in a 20% increase to individual-level smoking initiation probability and does not affect smoking cessation. Bottom Left: e-cigarette use results in a 20% increase to individual-level smoking cessation probability and does not affect smoking initiation. Bottom Right: e-cigarette use results in a 20% increase to both individual-level smoking cessation and individual-level smoking initiation probabilities.

References

1. Siebert U, Alagoz O, Bayoumi AM, Jahn B. State-Transition Modeling : A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. 2012;15:812-820.
2. Sternman JD. S-Shaped Growth: Epidemics, Innovation Diffusion, and the Growth of New Products. In: *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill Higher Education; 2000.
3. Schmidt L, Reidmohr A, Harwell TS, Helgerson SD. Prevalence and Reasons for Initiating Use of Electronic Cigarettes Among Adults in Montana, 2013. 2014:1-4.
4. Mcmillen RC, Gottlieb MA, Shaefer RM, Winickoff JP, Klein JD. Trends in Electronic Cigarette Use Among U.S. Adults: Use is Increasing in Both Smokers and Nonsmokers. 2014:1-8. doi:10.1093/ntr/ntu213.
5. Pepper JK, Brewer NT. Electronic Nicotine Delivery System (Electronic Cigarette) Awareness, Use, Reactions and Beliefs: a Systematic Review. *Tob Control*. 2013;23(5):1-10. doi:10.1136/tobaccocontrol-2013-051122.
6. Holford TR, Meza R, Warner KE, et al. Tobacco Control and the Reduction in Smoking-related Premature Deaths in the United States, 1964-2012. *JAMA*. 2014;311(2):164-171. doi:10.1001/jama.2013.285112.
7. Mendez D, Warner KE, Courant PN. Has Smoking Cessation Ceased? Expected Trends in the Prevalence of Smoking in the United States. *Am J Epidemiol*. 1998;148(3):249-258.
8. Vugrin ED, Rostron BL, Verzi SJ, et al. Modeling the Potential Effects of New Tobacco Products and Policies: A Dynamic Population Model for Multiple Product Use and Harm. *PLoS One*. 2015;10(3):e0121008. doi:10.1371/journal.pone.0121008.
9. Jeon J, Meza R, Krapcho M, Clarke L, Bryne J, Levy DT. Actual and Counterfactual Smoking Prevalence Rates in the US population via Micro-simulation. 2012;32. doi:10.1111/j.1600-6143.2008.02497.x.Plasma.
10. National Center for Health Statistics. *National Health Interview Survey 1997-2013*.; 1997. http://www.cdc.gov/nchs/nhis.htm.
11. Anderson C, Burns D. Patterns of Adolescent Smoking Initiation Rates by Ethnicity and Sex. *Tob Control*. 2000;9 Suppl 2(Suppl II):II4-I8. doi:10.1136/tc.9.suppl_2.ii4.
12. Pierce JP, White VM, Emery SL. What Public Health Strategies are Needed to Reduce Smoking Initiation? *Tob Control*. 2012;21(2):258-264. doi:10.1136/tobaccocontrol-2011-050359.
13. Huxley RR, Yatsuya H, Lutsey PL, Woodward M, Alonso A, Folsom AR. Impact of Age at Smoking Initiation, Dosage, and Time Since Quitting on Cardiovascular Disease in African Americans and Whites: The Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2012;175(8):816-826. doi:10.1093/aje/kwr391.

14. Cancer Intervention and Surveillance Modeling Network. Publication support and modeling resources. https://resources.cisnet.cancer.gov/projects/#shg/tce/summary. Accessed April 14th, 2014.

15. University of California Berkeley (USA), Max Planck Institute for Demographic Research (Germany). *The Human Mortality Database*. www.mortality.org or www.humanmortality.de.

16. U.S. Census Bureau. *Profiles of General Demographic Characteristics (1997-2010)*.; 2010. http://www.census.gov/data.html. Accessed February 11th, 2014.

17. Bonnie RJ, Stratton K, Kwan LY, Practice H. *Public Health Implications of Raising the Minimum Age of Legal Access to Tobacco Products*. (Bonnie R, Stratton K, Kwan L, eds.).; 2015.

18. Janssen F, van Wissen LJG, Kunst AE. Including the Smoking Epidemic in Internationally Coherent Mortality Projections. *Demography*. 2013;50(4):1341-1362. doi:10.1007/s13524-012-0185-x.

19. Centers for Disease Control and Prevention. *Table 1-1, "Live Births, Birth-Rates, and Fertility Rates, by Race: United States, 1909-2000."*; 2005. http://www.cdc.gov/nchs/data/statab/t001x01.pdf. Accessed February 11th, 2014.

20. Freedman ND, Ph D, Hartge P, et al. Smoking and Mortality — Beyond Established Causes. 2015:631-640. doi:10.1056/NEJMsa1407211.

21. Lynch JW, Kaplan GA, Cohen RD, Tuomilehto J, Salonen JT. Do Cardiovascular Risk Factors Explain the Relation Between Socioeconomic Status, Risk of All-cause Mortality, Cardiovascular Mortality, and Acute Myocardial Infarction? *Am J Epidemiol*. 1996;144(10):934-942. doi:10.1093/oxfordjournals.aje.a008863.

22. Holford TR, Levy DT, McKay LA, et al. Patterns of birth cohort-specific smoking histories, 1965-2009. *Am J Prev Med*. 2014;46(2):e31-7. doi:10.1016/j.amepre.2013.10.022.

23. Zhu SH, Gamst A, Lee M, Cummins S, Yin L, Zoref L. The Use and Perception of Electronic Cigarettes and Snus among the U.S. Population. *PLoS One*. 2013;8(10):e79332. doi:10.1371/journal.pone.0079332.

24. Durta L, Glantz S. Electronic Cigarettes and Conventional Cigarette use Among US Adolescents. *JAMA Pediatr*. 2014;168(7).

CHAPTER 3

Contextualizing Social Influence on Adolescent Smoking Behavior

Background

Social influence has long been hypothesized play a role in both harmful and beneficial adolescent behavior.[1,2] Among the first authors to bring interpersonal behavioral feedback to the foreground of adolescent social science is Travis Hirschi, who utilizes control theory to better understand the causes of delinquency.[3] In his seminal chapter "A Control Theory of Delinquency," Hirschi suggests that all humans would be delinquents, if only we dared.[1] Hirschi goes on to propose that because delinquency is intrinsic to human nature, explaining why humans conform to norms in social circles may be a more salient question.[4] In recent decades, the growing evidence supporting the association between peer influence and health behaviors has revealed the importance of understanding interpersonal dynamics across a range public health priorities such as obesity,[5] diabetes management,[6] and relevant to this dissertation, smoking behavior.[2,7–13]

In this study, we hypothesize that the role of social influence on smoking behavior differs by school network density, a contextual variable that some adolescent psychologists and social network experts propose can sway the nature of peer influence.[13,14] From these studies, we believe that the observed differences in social influence on behavioral outcomes by network density could be due to the interdependencies of macrosocial network properties (i.e., school or community networks characteristics) with local friendship networks (i.e., directly nominated friends) and the availability, and subsequent strength, of individual social ties. This chapter serves as an application of general epidemiologic methods as a first step in the development of a causal framework that explains adolescent smoking behavior from the perspective of peer influence in broader contexts, such as school-level network characteristics, and urbanicity.

Introduction

Despite reports of historic lows in adolescent smoking prevalence, recent studies suggest that adolescent smoking prevalence declines have slowed.[15] Moreover, even if smoking prevalence were to persist at 2014 rates, 5.6 million Americans adolescents are projected to die early from smoking-related illnesses.[16] The epidemiology of tobacco use suggests that effective prevention strategies must target young populations, with nearly nine out of ten adult smokers reporting first trying cigarettes before age 18.[16] In reflection of this science, tobacco control measures have developed narrowly focused endgame strategies to continue reducing smoking initiation among adolescents.[17] For example, renewed efforts to increase the minimum legal age for purchasing tobacco products from 18 to 21 years old have materialized at both local and state levels.[18,19] With adolescent populations experiencing greater susceptibility to peer influence than adults,[20,21] and with the changing tobacco use landscape via electronic cigarettes,[22] the development of narrowly tailored policy to suppress smoking initiation in youth remains paramount. A better understanding of the dynamics and mechanisms of the social pathways that connect broader contexts to smoking and other substance use behavior among adolescents can thus help in the development of policies that effectively decrease future adult smoking prevalence and its associated health burdens.

Prior studies have proposed the importance of social influence and context when considering various complex and behavior-driven public health priorities such as obesity,[5,23–25] violence,[26–28] and substance use.[11,29,30] Additional explorations in sociology[14,31,32] and network science[33–35] have also shown how friendship networks are developed within the features of rich social contexts, wherein individual behavior can be influenced by interpersonal and group dynamics. While public health researchers have found large and independent associations between primary friendship network characteristics with substance use behaviors among adolescents,[11,29,30,36,37] there exists a relative dearth of public health studies exploring the relationship between broader network structures with local friendship networks of individuals, and how these factors can interact to change the effects of social influence on behavioral outcomes.

Social influence on adolescent smoking behavior has been extensively explored in a large body of prior research. [38,39] Many studies focus on the concept of popularity[8,10,28–30] with some studies find that popularity is positively associated with smoking initiation when adjusting for the number of friends that smoke.[8] However, not only do measures of popularity vary greatly across studies, measures of the number of friends that smoke may be confounded by popularity, due to

popular adolescents being more likely to be friends with a greater number of smokers than less popular adolescents. Among analyses that adjust for friendship smoking influence as a proportion or percentage of friends, popularity is not found to be independently associated with smoking initiation.[12] Many studies also demonstrate associations between social connectedness and influence with contextual factors such as urbanicity and school size.[7,40] However, studies exploring social influence on individual smoking behavior often exclude contextual factors from their analyses.[10,12,13] Finally, while two studies have examined the importance of considering network density when determining the association between social influence and smoking behavior,[7,14] these studies do not thoroughly explore the potential effect modification by network density on the association between peer influence and smoking behavior.

The main objective of this study is to conduct a comprehensive analysis to identify whether network density changes the relationship between friendship influence and adolescent smoking behavior while adjusting for broader contextual factors commonly associated with smoking behavior, such as school-level smoking prevalence, urbanicity, and relevant individual characteristics. This allows us to identify potential non-linear relationships of social influence on smoking initiation and cessation across density strata, providing insights into critical factors that could be the target of network-based tobacco control policy in adolescents. We additionally investigate the extent of the difference in pathways through which social and environmental factors influence smoking initiation versus smoking cessation among adolescents.

Methods

Data Overview

We use data from the National Longitudinal Study of Adolescent Health (Add Health), a school-based longitudinal study of a nationally representative sample of adolescents in the United States attending grades 7-12 in 1994.[41] High schools were systematically selected in the first wave of Add Health to be representative of US schools with respect to region in the country, urbanicity, size, type, and ethnic composition. In-school questionnaires were administered to all consenting students in each participating school. A subset of participants from each school was selected to complete an in-home questionnaire. These adolescents were then followed over time across five waves of interviews, although for the purposes of this study, we only explore up through wave 4. We drop from our sample schools that are missing administrator surveys, that do not include

grades, or that are designated as specialized schools. To provide accurate network measures that are representative of a full school-level measure, Add Health only calculated network-level characteristics for individuals attending schools with at least 50% participation in the in-school questionnaire.[41] Thus, we only consider schools with greater than 50% participation, and include only observations with complete data for our network measures of interest. Observations with missing values for other independent variables, sample weights, and self-report of smoking in the past 30 days are also dropped. In this study, we use data from waves 1, 2, 3, and 4, conducted in 1994-1995, 1996, 2001-2002, and 2008, respectively. All measures that are not outcome measures in this study are obtained from wave 1 of the in-school or in-home interview. The primary outcome of interest in this study is smoking prevalence at wave 3, based on the self-reported number of days smoked in the past 30 days during the wave 3 in-home interview. A value of 0 denotes self-report of never smoking or of smoking 0 days in the past 30 days (i.e., non-smokers), while a value of 1 represents self-report of smoking 1 or more days in the past 30 days (i.e., current smokers).

Individual Characteristics

Sociodemographic individual characteristics include: self-report of race/ethnicity, age, sex, and frequency of feeling depressed in the past month. Adolescents self-reported as either white, black or African American, Asian or Pacific Islander, American Indian or Native American, or other during the in-school Add Health interviews. A separate ethnicity question determined whether an individual was of Hispanic or Spanish origin. For this study, individuals who selected only white or non-hispanic black are categorized as such, while mixed race, and all other races and ethnicities are categorized into an "other" category. Age is calculated by subtracting self-reported birthdate from the month and year of the in-home interview date at wave 1, and centered on the average age of all observations in wave 1. Sex is categorized as male or female. Finally, the frequency of feeling depressed in the past month was asked as "in the past month, how often did you feel depressed or blue?" Responses were categorized in the Add Health questionnaire as: "never," "rarely," "occasionally," "often," and "every day". We retain these categories in our study as categorical variables. We additionally include a binary (yes/no) measure of whether cigarettes are easily available at home, which is obtained from wave 1 of the in-home interview. Observations for any category with multiple responses are dropped.

Individual Network Measures

The variable "in connections" is a summary count measure of the number of other individuals that select an adolescent of interest as a friend, while the variable "out connections" is defined as the number of other individuals that an adolescent of interest nominates as their friend. The dependent variable of interest in this study is an individual-level measure of the percentage of nominated friends (i.e., "out connections") that self-report smoking in the past 12 months in the in-school interview. Participants at the baseline in-school interview at wave 1 were asked to select five of their closest male friends and five of their closest female friends. Every participant was limited to nominating a maximum of ten friends. Of these friends, only those that participated in the in-school questionnaire in wave 1 are included in the denominator of the percentage of friends that smoke calculation for this study. Individual network measures include influence domain and age heterogeneity. Influence domain, which can be understood as popularity, is an individual-level summary measure of network centrality, and quantifies popularity by measuring the number of other adolescents (i.e., "alters") can reach the adolescent of interest (i.e., "ego"). To account for the potential effects of older peers providing access to adolescents younger than the legal age of cigarette purchase, we include age heterogeneity as an individual-level summary measure to quantify the amount of variability in age of an adolescent's social network. Here, the age heterogeneity calculation includes both in and out connections.

School Characteristics

At the school level, demographic characteristics include: urbanicity, size, type, and regional location. Urbanicity (suburban, rural or urban) and regional location (west, midwest, south, and northeast) were determined by Quality Education Data, with school administrators reporting on school size (small: 1-400 students, medium: 401-1000 students, large: 1001-4000 students) and type (public or private). We calculate a school-level smoking prevalence measure using individual self-report of smoking in the past 12 months in the baseline in-school questionnaire for each school.

School Network Measure

To investigate macro-level social influence factors, we include school-level relative network density (density) in our analyses. This measure is calculated by dividing the number of

ties in the global network by the number of possible ties in the total friendship network, relative to the nomination limit of friendships (i.e., 10 total). A summary glossary of school and individual-level network terms is available for reference in Table 3.1. School-level smoking prevalence, relative density, age heterogeneity, and influence domain are categorized based on their percentile distribution (below the 25th percentile, 25th-50th percentile, 50th-75th percentile, and above the 75th percentile) for ease of interpretation.

Table 3.1 Summary glossary of network terms used throughout this chapter.

Network Term	Definition
Ego	The focal adolescent of interest
Alter	Adolescents in the network that are not the focal adolescent of interest
Tie	Representation of a relationship, specifically represents friendships in this study
Out Connection	A directed tie that runs from the ego to a nominated alter
In Connection	A directed tie that runs from an alter to the ego (i.e., the alter nominated the ego)
Network Density	The total number of connections divided by the the total number of possible connections in a network
Individual Centrality	A broad reference measure of an ego's position in the network relative to alters
Individual Influence Domain	A summary measure of the number of alters that can reach the ego through undirected friendship ties
School Relative Density	The number of actual ties between adolescents in a school divided by the the maximum possible connections in the school, normalized by the maximum possible out connection nominations (i.e., 10 in this study)

Analyses

To better understand the rates of adolescent smoking behavior changes between waves, we subset our observations by smoking status (i.e., non-smoker and smoker) in waves 1, 2, and 3. Then, for each future wave, we determine the weighted proportion of individuals that remain in the same smoking state, or change their smoking state relative to the wave that is considered as baseline. For example, when considering wave 1 as baseline, we determine the proportion of individuals transitioning between smoking states from wave 1 to wave 2, wave 1 to wave 3, and wave 1 to wave 4. Similar calculations are performed when considering wave 2 or wave 3 as the baseline wave.

Weighted descriptive statistics are estimated for all variables. We then determine the univariate associations between smoking initiation and smoking cessation with school urbanicity, school smoking prevalence, child age heterogeneity, child popularity, and percentage of friends that smoke. Next, we examine the association between the percentage of friends that smoke (i.e., peer influence) with smoking initiation by wave 3 in an unstratified mixed-effects regression model, adjusting for all variables explored in the univariate analysis, in addition to age, sex, race/ethnicity, self-report of depression, school size, and the availability of cigarettes at home. In models with smoking initiation as the outcome, we subset our data to only include non-smokers in

wave 1. We perform analogous analyses of smoking cessation using only adolescents that report current smoking in wave 1. Mixed effects regression is used to account for the hierarchical nature of the data and to examine the association of school-level characteristics on individual-level smoking outcomes. Next, we stratify our initiation and cessation models by network density, first for the univariate analysis, and then with the adjusted models. We additionally perform stratified analyses by sex to identify potential differences in peer influence effects by this factor. A modified Poisson approach is used to estimate the risk ratio (RR).[42] All analyses are multi-level weighted and performed using PROC GLIMMIX in SAS v9.4

Results

In total, we retain 5,404 adolescents that participated in the in-school and in-home questionnaires from wave 1 and in wave 3. Overall, the exclusion of observations does not substantially change the distribution of the covariates of interest in our analyzed sample relative to the complete Add Health population. Flow charts of the data exclusion process (Figure B1), a conceptual model of the associations that are tested in this study (Figure B2), and mean comparisons between the data used in this study relative to the full data of the Add Health study (Table B1), are presented in Appendix B. In this chapter, we focus primarily on the potential differences of peer influence by network density in our stratified models. However, we also find statistical associations between network density with urbanicity (Table B2). Rural schools (RR=1.03, 95% CI: 0.84, 1.26) in the Add Health dataset report higher density levels than suburban (reference) and urban schools (RR=0.78, 95% CI: 0.64, 0.97) on average. The details and implications of these outcomes will be discussed in Chapter 4, although some contrasts by urbanicity are presented here to emphasize the importance of context when exploring the potential effects of peer influence on smoking behavior.

Stratified weighted descriptive statistics of our study population are presented in Table 3.2. A greater proportion of smokers attending schools in the highest quartile of network density (i.e., 75th percentile) also report the highest quartile of age heterogeneity and percentage of friends that smoke relative to adolescents attending schools in the lowest quartile of network density. Additionally, we observe that the proportion of white smokers increases relative to other races as network density increases. Adolescents in the 4th quartile of popularity are primarily concentrated in schools below the 75th percentile of network density. Among adolescents in the 4th quartile of

popularity (i.e., "influence domain") attending schools below the 75th percentile of network density, a greater proportion are non-smokers than current smokers. Above the 75th percentile of network density, a slightly greater percentage comprise of smokers than non-smokers. We additionally observe a large percentage of rural schools concentrated in the 4th quartile of network density, as compared to suburban and urban schools, which comprise of all schools in the 1st quartile of network density.

Table 3.3 presents weighted transition percentages of non-smoker to smokers and vice-versa between four waves of Add Health in-home interviews. This table includes transition percentages for the study population and stratified transition percentages by network density. For the entire study population, we observe that adolescents tend to maintain smoking states between wave 3 and wave 4, compared to larger percentages of state transitions (i.e., from non-smoker to smoker or from smoker to non-smoker) between wave 1, 2, and 3. Lower rates of transitions between smoking states from wave 3 to wave 4 indicate that sustained smoking behavior solidified before interviews were administered in wave 4. In contrast, wave 2 data were collected one year after wave 1 interviews, which may produce skewed outcomes due to measurements reflecting short-term smoking experimentation. Thus, our models use wave 1 risk factors to predict smoking outcome measurements in wave 3.

From the stratified transition table, we observe that greater percentage of current smokers remain current smokers (66.49%, 95% CI: 60.36%, 72.63%) between wave 1 and wave 3 in the 4th quartile of network density (Q4), than current smokers in other schools (Q1, Q2, and Q3). We additionally observe that a greater percentage of non-smokers remain non-smokers in Q1 (80.12%, 95% CI: 76.30%, 83.93%) and Q4 (75.20%, 95% CI: 71.72%, 78.69%) relative to Q2 (69.92%, 95% CI: 66.25%, 73.59%) and Q3 (72.40, 95% CI: 68.79%, 76.00%) between waves 1 and 3.

To further examine how smoking transitions may differ by context, we explore smoking transitions stratified by urbanicity in Table 3.4. Here, we also observe differences in smoking trends by urbanicity, where a large proportion of the population maintains smoking states between wave 3 and wave 4 relative to transitions from earlier waves to wave 3 and wave 4. We additionally observe higher rates of non-smokers remaining non-smokers in rural (76.17%, 95% CI: 72.37%, 79.96%) areas compared to suburban (72.79%, 95% CI: 70.26%, 75.33%) and urban (74.20%, 95% CI: 70.51%, 77.90%) areas between wave 1 and wave 3. The percentage of smokers that

remain smokers is also higher in rural areas (71.02%, 95% CI: 64.77%, 77.27%) than in suburban (65.40%, 95% CI: 60.87%, 69.93%) and urban (57.76%, 95% CI: 48.86%, 66.65%) areas.

Table 3.6 reports adjusted risk ratios of risk factors related to smoking cessation in wave 3 among current smokers in wave 1. Like Table 3.5, this table reports two unstratified models, one adjusting for, and another excluding school network density, and presents stratified model results by school network density. For current smokers at wave 1, we observe an inverse association between school smoking prevalence and smoking cessation (RR=0.80, 95% CI: 0.69, 0.92) in the full network density adjusted model. This inverse association increases slightly when removing network density from the model (RR= 0.83, 95% CI: 0.72, 0.95). The magnitude of the association between school smoking prevalence and smoking cessation in the model that adjusts for network density is much larger than that of the percentage of friends that smoke (RR=0.98, 95% CI: 0.94, 1.02) in the same model. Although the estimates do not reach significance, the magnitude and direction of the associations between school smoking prevalence and smoking cessation found in both unstratified models persist across all levels of school network density, except in the 3rd quartile (RR=1.08, 95% CI: 0.64, 1.82). Stratified analyses indicate urbanicity is also associated with smoking cessation in certain contexts, with adolescent smokers attending rural (RR=0.51, 95% CI: 0.29, 0.89) and urban schools (RR=0.42, 95% CI:0.21, 0.84) in the 4th quartile of network density being much less likely to quit than their suburban counterparts. Conversely, we observe that adolescents attending rural schools are more likely to quit if they attend schools in the 2nd (RR=1.37, 95% CI: 0.84, 2.25) or 3rd (RR=1.06, 95% CI: 0.75, 1.51) quartile of network density, though these estimates are not statistically significant.

We observe an inverse association between the percentage of friends that smoke with the likelihood of smoking cessation in both unstratified models, with a higher percentage of friends being negatively associated with cessation across all density strata. However, this association only reaches statistical significance in the 4th quartile of network density (RR=0.90, 95% CI: 0.83, 0.97). Finally, we observe that adolescents in the 2nd quartile of popularity are also less likely to quit smoking if they attend schools in the 1st (RR=0.22, 95% CI: 0.09, 0.51) and 4th (RR=0.34, 95% CI: 0.16, 0.73) quartile of network density relative to the least popular adolescents. Unadjusted risk ratios of smoking initiation and smoking cessation by network density for the variables of interest are also discussed in Appendix B and presented in Table B3. Results presenting all variables adjusted for in the network density stratified models exploring their estimated effects on

smoking initiation and cessation can be found in Table B4. Table B5 presents smoking initiation and smoking cessation relative risk estimates stratified by sex. We observe substantially larger effect in the percentage of friends that smoke for smoking initiation among females (RR=1.10, 95% CI: 1.04, 1.16) than in males (RR=1.05, 95% CI: 0.99, 0.11). Multiplicative statistical interaction between sex and the percentage of friends that smoke in predicting smoking initiation is also found to be significant, and higher in females than in males (RR=1.01, 95% CI: 1.00, 1.01). Finally, we find that males are less likely to stop smoking with increasing school smoking prevalence (RR=0.71, 95% CI: 0.51, 0.98) compared to females (RR=0.92, 95% CI: 0.74, 1.13).

Given the evidence of non-linear effect modification by network density on the relationship between peer influence and smoking initiation, we test for quadratic statistical interaction of this relationship. We find that the effect estimate of the quadratic interaction term is statistically significant (RR=1.23, 95% CI: 1.08, 1.40) in predicting smoking initiation. We additionally assess the presence of statistical interaction between other potential effect modifiers and percentage of friends that smoke for smoking cessation, and do not observe significance on the interaction term for sex or network density. Figures B3 and B4 illustrate the observed heterogeneous effects of network density and sex, respectively, on the relationship between peer influence and smoking initiation.

Discussion

While social network and contextual factors have been commonly identified as determinants of smoking behavior among adolescents, the mechanisms driving the associations between smoking and social factors remain largely unexplained when considering the contextual aspects that give rise to individual social networks. To our knowledge, this is the first study that explores in depth the effects of social network density on the effect of peer influence on smoking initiation and cessation. In providing insights into the mechanisms of contextual factors on the pathway between social influence and smoking behavior, our results suggest properties of schools that may be amenable for social network-based tobacco control intervention, such as empowering the opinions of non-smokers in socially dense schools through media campaigns and school-based programs.[43]

The inclusion of a broad range of confounders in our adjusted and unstratified models does not dramatically change the univariate effect estimates of the percentage of friends that smoke on

smoking initiation and smoking cessation. However, we observe differences regarding the potential social pathways between adolescents that initiate versus quit smoking. We find that primary friendship networks are a stronger predictor than school smoking prevalence for smoking initiation, whereas school smoking prevalence is a stronger predictor than primary friendship networks for smoking cessation. Consistent with existing research, these results suggest that smoking initiation among adolescents has a strong individual social component,[7,9,11] especially among females and adolescents attending schools in the 4th quartile of network density, while the decision to quit may be more highly associated with individual traits and contextual factors,[44,45] rather than through close friendship influence.

Additionally, we find that school network density acts as an effect modifier on primary friendship network influence on smoking behavior. In the stratified models, we observe that primary friendship influence on smoking initiation is highest at schools in the 1st and 4th quartile of network density. Notably, rural adolescents attending schools in the 4th quartile of network density are less likely to initiate smoking and more likely to quit smoking than urban adolescents with comparable school network densities. Our interaction analysis additionally shows that there is evidence for non-linear heterogeneity in the effect of the percentage of friends that smoke and smoking initiation as a function of network density.

Although consistent with other studies suggesting that rural schools have higher smoking prevalence than urban or suburban schools,[39] weighted transition tables between wave 1 and wave 3 show that a greater percentage of non-smokers in rural schools remain non-smokers, and current smokers are less likely to quit, compared to urban and suburban schools. This suggests that smoking uptake may occur at younger ages in rural areas, given high smoking prevalence at baseline. Additionally, low rates of cessation once an adolescent begins smoking in rural areas may contribute to disparate smoking prevalence patterns by urbanicity. Given the association between urban designation of schools and network density,[46,47] these results suggest potential mechanistic differences in smoking behavior between adolescents by urbanicity that may be explained by network factors.

The results presented in this chapter corroborate existing theories in adolescent psychology of differences in friendship influence by contextual factors that affect network characteristics. Social bonds have long been hypothesized to contribute to conforming behavior,[3,4,14] with recent network theory suggesting that stronger social bonds can be quantified through measures of

centrality, such as "triadic closure," such that stronger social ties exist when an adolescent's friends are also friends with each other.[48] These explanations primarily seek to explain the emergence of homiphily by race, sex, and behavior (e.g., sexual, substance use) in adolescent friendship networks, with a large body of literature suggesting that friendship influence can differ in varying contexts by which the friendship manifests. In other words, friendships occur within rich and complex social (e.g., availability of friendship) and physical (e.g., urbanicity) constructs that may contribute to differences in the strength of friendship influence on friendship behavior. Moreover, prior studies in sociology and psychology have suggested that network structures and the subsequent strength of friendship ties differ substantially by race, and these factors can also differ in terms of the distribution of race from the community context, such as urbanicity.[49,50]

Here, the results indicate that the impact of the percentage of friends that smoke on future smoking behavior is context dependent, and the magnitude of this effect varies between socially dense settings (e.g., dense schools in rural settings, where all the students are closely connected to one another) and less dense settings (e.g., urban schools, where disconnected cliques are more likely to form). These factors may be further exacerbated by differences in network characteristics by race, particularly when accounting for disparities in racial distribution by urbanicity. The observed interaction between the impact of friendship influence and network density on smoking behavior may be due to a variety of factors, such as how contextual factors affect the meaning of friendships. For instance, differences in strength of ties may be responsible for differences in the influence that friends might have on smoking behaviors across contextual dimensions. That is, the strength of friendship ties may be largely associated with the availability of schoolmates from which to choose friends (e.g., indirectly measured by network density). If friendship tie strength is consistent across adolescents within school networks, but different across school networks, they may be responsible for generating the observed differences in peer influence on smoking behavior by network density.

The main strength of these analyses is inherent to the Add Health study, which is longitudinal and population representative, allowing us to present generalizable and time-dependent results. Additionally, the rich dataset provides sufficient power to find statistical significance when performing stratified analyses. The data also allows us to match friend identifiers with self-reports of smoking by the friends themselves, as compared to measures of perceived friend smoking behavior used in other literature. The main weakness of this study is also

inherent to the data. Wave 1 of data collection in the Add Health study began in 1994, when smoking incidence and prevalence among adolescents was much higher than it has been in recent years. Importantly, electronic cigarette use has increased rapidly among adolescents, with 30-day use prevalence exceeding that of traditional cigarette use prevalence in 2014.[22,51] With the onset of new products, continued declines to and reports of historic lows in adolescent smoking prevalence,[45] the results of this study may not be generalizable to adolescents today. Additionally, the differences by density on associations between risk factors and adolescent smoking behavior may be largely explained by parent and household factors such as SES and parental smoking, which we exclude in this model to preserve power, although we use the availability of cigarettes at home to account for these factors indirectly. Given that the differences in friendship observed by network density are potential reflections of differences in the conceptualization of friendship in different contexts, more detailed measures of friendship and closeness may be required to gain a better understanding of specific nature of friendship influence on smoking behavior.

Our results bring new insights into youth tobacco usage patterns and social epidemiology right as we are on the cusp of additional legislation that limits access to tobacco (i.e., increasing the minimum age of purchase from 18 to 21 years-old).[52] The findings presented here suggest that tailoring tobacco control measures based on higher-level factors, such as network density of schools, can be effective and efficient at curtailing adolescent smoking initiation, and encouraging smoking cessation among current adolescent smokers. The results of this study also provide potential mechanistic insights into the social influences of smoking behavior that could also translate into other substance use behavior.[14,29] Finally, this study seeks to present a potential causal framework for which higher-level network structures social influence can function through contextual pathways to affect smoking behavior. While many studies have explored social influence and context as independent risk factors for smoking behavior,[7,8,38,39] this is the first study to our knowledge that presents an epidemiologic framework connecting these concepts.

Conclusion

In this chapter, we use mixed effects models to assess how the pathway of peer influence on smoking behavior changes based on school network density in US adolescents. After adjustment for school level (smoking prevalence, urbanicity), individual (age, gender, race, depression history), and other network covariates, we found that with every 10% increase in the

percentage of close friends that smoke, adolescents are 1.07 (95% CI: 1.03, 1.13) times more likely to initiate smoking in the future. This association differs in magnitude by levels of school network density. Adolescents attending schools in the 4th quartile of network density are more likely to initiate smoking by 1.15 (95% CI: 1.07, 1.24) times for every 10% increase in the percentage of nominated friends that smoke. Among current smokers, future smoking cessation is inversely associated with percentage of close friends that smoke in wave 1, with adolescent smokers attending schools in the 4th quartile of network density being 0.90 (95% CI: 0.83, 0.97) times less likely to stop smoking for every 10% increase in percentage of friends that smoke. Given the potential for network structures to modify the association between peer influence and smoking behavior, we develop a computational model using Add Health data to examine how much of these differences can be explained by network structure alone. In the next chapter, we present additional mixed effects analyses by urbanicity, and apply a computational model as a tool to validate the conceptual model of the pathways between contexts (i.e., urbanicity and school network density), social influence, and smoking behavior that is presented here.

Table 3.2 Weighted school- and individual-level characteristic distribution of current smokers and non-smokers stratified by quartiles of school relative network density, Add Health, 2001-2002.

	Total Population						Network Density Q1 [Range: 0.19, 0.38]				Network Density Q2 [Range: 0.38, 0.47]				Network Density Q3 [Range: 0.47, 0.52]				Network Density Q4 [Range: 0.52, 0.83]			
	N	Weighted %	Proportion Smokers (N=1725)	95% CI	Proportion Non-Smokers (N=3679)	95% CI	Proportion Smokers (N=305)	95% CI	Proportion Non-Smokers (N=1025)	95% CI	Proportion Smokers (N=428)	95% CI	Proportion Non-Smokers (N=945)	95% CI	Proportion Smokers (N=460)	95% CI	Proportion Non-Smokers (N=863)	95% CI	Proportion Smokers (N=532)	95% CI	Proportion Non-Smokers (N=846)	95% CI
Sex																						
Male	2290	45.72	49.53	46.49, 52.57	43.56	41.39, 45.74	47.41	39.08, 55.74	43.41	38.7, 48.12	49.48	43.56, 55.40	41.19	37.03, 45.36	48.40	42.83, 53.97	45.21	40.95, 49.48	51.82	46.31, 57.32	44.52	40.21, 48.82
Female	3114	54.28	50.47	47.43, 53.52	56.44	54.26, 58.61	52.59	44.26, 60.92	56.59	51.88, 61.30	50.52	44.6, 56.44	58.81	54.64, 62.97	51.60	46.03, 57.17	54.79	50.52, 59.05	48.18	42.68, 53.69	55.48	51.18, 59.79
Race																						
Non-Hispanic Black	996	12.90	6.05	4.76, 7.34	16.78	15.31, 18.24	15.00	9.06, 20.93	27.16	23.15, 31.16	4.92	2.94, 6.90	21.09	18.03, 24.16	5.76	3.59, 7.92	11.77	9.35, 14.19	3.17	1.47, 4.88	9.42	7.26, 11.58
White	3218	70.37	77.53	75.05, 80.01	66.33	64.35, 68.30	56.45	48.36, 64.53	41.75	37.02, 46.47	74.61	69.41, 79.82	62.50	58.55, 66.45	80.26	76.17, 84.35	72.31	68.73, 75.88	88.09	84.60, 91.59	83.00	80.00, 85.99
Other	1190	16.73	16.42	14.17, 18.67	16.9	15.32, 18.48	28.55	21.54, 35.57	31.10	26.87, 35.32	20.47	15.45, 25.49	16.41	13.25, 19.57	13.98	10.36, 17.6	15.92	13, 18.85	8.73	5.60, 11.87	7.58	5.34, 9.83
Self-Report of Depressive Feelings																						
Never	1878	35.63	33.1	30.18, 36.03	37.06	34.96, 39.17	35.50	27.31, 43.69	40.96	36.32, 45.60	35.86	30.11, 41.62	38.07	34.00, 42.14	31.24	25.97, 36.51	34.47	30.41, 38.52	30.96	25.75, 36.17	35.72	31.57, 39.87
Rarely	1640	30.42	30.03	27.25, 32.81	30.64	28.64, 32.63	22.85	16.1, 29.59	30.71	26.4, 35.02	31.57	26.05, 37.09	29.76	25.94, 33.57	30.29	25.17, 35.4	31.20	27.26, 35.13	31.64	26.54, 36.74	30.95	27.00, 34.89
Occasionally	1014	17.92	17.72	15.49, 19.96	18.04	16.40, 19.68	16.31	10.26, 22.36	14.48	11.2, 17.76	16.41	12.17, 20.65	18.70	15.47, 21.93	18.62	14.52, 22.72	18.93	15.65, 22.22	18.86	14.67, 23.06	19.09	15.86, 22.32
Often	595	10.61	11.84	9.94, 13.74	9.92	8.61, 11.22	16.75	11.07, 22.44	9.42	6.71, 12.13	11.48	7.86, 15.11	9.58	7.06, 12.09	12.12	8.46, 15.78	10.83	8.2, 13.45	9.51	6.38, 12.64	9.71	7.15, 12.27
Everyday	277	5.41	7.31	5.74, 8.87	4.35	3.43, 5.26	8.59	3.94, 13.24	4.42	2.66, 6.19	4.68	2.26, 7.09	3.89	2.14, 5.65	7.73	4.8, 10.67	4.58	2.62, 6.53	9.03	5.76, 12.30	4.54	2.80, 6.28
Age																						
11-15	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
16-20	1487	35.53	36.63	33.61, 39.65	34.91	32.78, 37.04	38.22	30.04, 46.4	33.43	28.84, 38.02	49.71	43.79, 55.64	45.88	41.66, 50.10	30.99	25.62, 36.37	29.53	25.46, 33.6	27.84	22.71, 32.98	29.79	25.81, 33.77
21-25	3912	64.35	63.17	60.14, 66.19	65.02	62.89, 67.16	61.78	53.6, 69.96	66.57	61.98, 71.16	49.90	43.98, 55.82	54.09	49.87, 58.31	68.69	63.31, 74.08	70.27	66.19, 74.34	72.16	67.02, 77.29	70.21	66.23, 74.19
26-30	5	0.12	0.21	0.00, 0.50	0.06	0.00, 0.17	---	---	---	---	0.39	0.00, 1.16	0.03	0.00, 0.10	0.31	0.00, 0.93	0.20	0.00, 0.58	---	---	---	---
31-35	---	---	---	---	---	---	---	---	---	---												
Age Heterogeneity																						
1st Quartile [0, 0.40]	1337	24.28	21.97	19.44, 24.50	25.59	23.68, 27.51	33.33	25.33, 41.34	30.30	25.88, 34.73	26.66	21.41, 31.9	29.95	26.19, 33.72	16.64	12.52, 20.76	21.48	17.87, 25.1	17.04	13.06, 21.02	21.57	17.95, 25.20
2nd Quartile [0.40, 0.50]	1240	24.88	25.07	22.36, 27.79	24.77	22.86, 26.67	29.13	21.45, 36.81	20.92	17.09, 24.75	28.37	22.83, 33.9	28.07	24.19, 31.95	23.60	18.75, 28.45	23.96	20.31, 27.61	21.12	16.57, 25.67	24.96	21.23, 28.68
3rd Quartile [0.50, 0.61]	1473	27.54	27.86	25.14, 30.58	27.36	25.43, 29.29	17.27	11.61, 22.94	27.32	23.2, 31.44	25.71	20.66, 30.77	24.84	21.21, 28.47	33.75	28.44, 39.05	28.95	25.12, 32.78	29.11	23.93, 34.28	28.45	24.55, 32.35
4th Quartile [0.61, 0.84]	1354	23.30	25.1	22.55, 27.65	22.29	20.54, 24.03	20.26	13.83, 26.70	21.46	17.72, 25.19	19.27	14.78, 23.75	17.14	14.05, 20.23	26.01	21.29, 30.73	25.61	21.97, 29.25	32.73	27.63, 37.84	25.02	21.50, 28.54
Percentage of Out Friends Smoke																						
1st Quartile [0, 0]	1339	21.81	16.29	14.03, 18.54	24.94	23.10, 26.78	26.67	19.21, 34.13	35.94	31.50, 40.37	21.83	16.92, 26.73	29.65	25.89, 33.41	12.96	9.25, 16.68	18.59	15.24, 21.94	8.79	5.92, 11.66	18.11	14.93, 21.29
2nd Quartile [10,16.7]	1192	20.95	18.75	16.37, 21.13	22.2	20.43, 23.97	22.79	16.24, 29.34	24.96	20.99, 28.92	21.27	16.4, 26.14	22.68	19.22, 26.14	18.99	14.42, 23.57	21.65	18.14, 25.15	13.82	10.19, 17.45	20.14	16.83, 23.44
3rd Quartile [20, 30]	1331	25.72	24.57	21.97, 27.17	26.38	24.44, 28.32	20.38	13.86, 26.89	20.94	17.05, 24.83	24.58	19.52, 29.64	26.70	22.85, 30.54	26.22	21.35, 31.09	26.66	23, 30.32	24.86	20.09, 29.64	29.89	25.89, 33.90
4th Quartile [30, 100]	1542	31.51	40.39	37.41, 43.38	26.79	24.54, 28.43	30.17	22.35, 37.98	18.17	14.33, 22.00	32.32	26.78, 37.86	20.98	17.58, 24.38	41.83	36.37, 47.28	33.10	29.04, 37.16	52.52	47.02, 58.02	31.86	27.84, 35.88
Influence Domain																						
1st Quartile [1, 292]	1339	29.47	30.83	27.98, 33.69	28.7	26.71, 30.69	37.04	28.95, 45.13	33.55	28.94, 38.15	37.02	31.17, 42.87	33.96	29.9, 38.02	21.26	16.91, 25.6	16.12	13.03, 19.21	31.34	26.12, 36.56	32.57	28.59, 36.55
2nd Quartile [292, 534]	1317	28.50	29.44	26.64, 32.25	27.97	26.01, 29.93	22.54	15.76, 29.32	20.30	16.55, 24.05	32.72	27.05, 38.38	28.88	25.01, 32.76	35.53	30.1, 40.95	35.42	31.32, 39.52	22.88	18.44, 27.32	24.96	21.36, 28.57
3rd Quartile [534, 762]	1392	24.81	25.11	22.55, 27.66	24.65	22.79, 26.50	15.41	9.79, 21.02	18.79	15.19, 22.38	12.85	9.11, 16.58	10.75	8.24, 13.26	25.98	21.28, 30.69	29.07	25.33, 32.82	42.04	36.58, 47.50	39.60	35.36, 43.83
4th Quartile [793, 1705]	1356	17.21	14.62	12.50, 16.73	18.68	17.04, 20.32	25.01	17.74, 32.28	27.37	23.40, 31.34	17.42	13.49, 21.34	26.40	22.91, 29.9	17.23	12.81, 21.66	19.39	15.87, 22.91	3.74	1.63, 5.84	2.87	1.52, 4.21
School Urbanicity																						
Suburban	2861	58.7	59.5	56.58, 62.43	58.25	46.14, 60.37	57.61	49.54, 65.67	53.23	48.54, 57.91	71.13	66.07, 76.19	62.8	58.83, 66.78	75.46	70.89, 80.03	74.16	70.65, 77.68	31.07	25.83, 36.31	40.29	36.05, 44.53
Urban	1282	19.43	16.79	14.66, 18.92	20.92	19.23, 22.62	42.39	34.33, 50.46	46.77	42.09, 51.46	21.35	16.7, 25.99	25.99	22.33, 29.64	7.49	5.14, 9.84	7.82	6.09, 9.54	9.22	6.55, 11.89	9.48	7.35, 11.61
Rural	1261	21.86	23.71	21.21, 26.20	20.82	19.09, 22.56	---	---	---	---	7.53	5, 10.05	11.21	8.9, 13.52	17.05	12.9, 21.21	18.02	14.8, 21.24	59.71	54.3, 65.12	50.23	45.95, 54.50
Region																						
Midwest	944	11.63	9.69	8.04, 11.34	12.73	11.42, 14.04	30.78	22.96, 38.60	24.59	20.21, 28.97	35.35	29.54, 41.17	26.77	22.96, 30.59	27.14	22.05, 32.22	24.99	21.25, 28.74	44.98	39.37, 50.6	40.22	35.85, 44.59
Northeast	1519	31.35	34.98	31.99, 37.98	29.3	27.23, 31.37	9.32	5.00, 13.65	8.12	5.55, 10.7	11.72	8.14, 15.3	12.54	9.65, 15.44	17.21	12.97, 21.45	14.16	11.18, 17.13	20.74	16.79, 24.70	23.48	20.13, 26.83
South	801	15.12	15.5	13.45, 17.55	14.91	13.40, 16.42	22.98	15.55, 30.42	32.06	27.72, 36.4	41.98	36.15, 47.81	45.60	41.44, 49.77	50.71	45.15, 56.28	54.79	50.56, 59.02	34.28	29.21, 39.34	36.30	32.33, 40.27
West	2140	41.89	39.84	36.87, 42.80	43.06	40.92, 45.20	36.91	29.17, 44.66	35.23	30.91, 39.54	10.95	7.59, 14.32	15.09	12.26, 17.91	4.94	3.08, 6.80	6.06	4.46, 7.65	---	---	---	---
Size																						
Large	970	20.68	21.75	19.16, 24.34	20.08	18.34, 21.82	49.10	40.84, 57.36	54.24	49.51, 58.96	18.30	14.27, 22.33	26.50	23, 30	27.59	22.57, 32.62	26.04	22.22, 29.86	30.93	25.53, 36.34	22.80	18.89, 26.72
Medium	2281	49.24	49.56	46.53, 52.60	49.06	46.89, 51.23	40.71	32.47, 48.95	35.80	31.13, 40.47	49.22	43.30, 55.14	44.80	40.61, 48.99	61.28	55.88, 66.68	63.13	59.06, 67.20	41.83	36.55, 47.12	48.97	44.71, 53.23
Small	2153	30.08	28.69	25.96, 31.42	30.86	28.87, 32.85	10.18	5.05, 15.32	9.96	7.14, 12.79	32.48	26.72, 38.25	28.70	24.81, 32.59	11.13	7.89, 14.36	10.83	8.57, 13.09	27.23	22.21, 32.26	28.23	24.34, 32.11
School Smoking Prevalence																						
1st Quartile [0.05,0.26]	1349	26.76	23.44	20.73, 26.16	28.63	26.66, 30.60	41.29	33.05, 49.54	45.31	40.63, 49.99	48.91	42.98, 54.85	53.00	48.83, 57.17	3.60	1.41, 5.79	5.71	3.75, 7.66	8.53	5.47, 11.58	13.56	10.69, 16.43
2nd Quartile [0.26, 0.33]	1309	19.55	19.47	17.10, 21.84	19.59	17.90, 21.28	18.86	13.37, 24.36	19.86	16.48, 23.25	1.51	0.47, 2.55	3.50	2.27, 4.74	33.28	27.96, 38.61	31.18	27.27, 35.1	24.30	19.41, 29.19	24.64	20.83, 28.46
3rd Quartile [0.33, 0.41]	1352	26.68	26.23	23.48, 28.78	27	25.08, 28.91	29.34	22.12, 36.57	27.50	23.26, 31.74	36.70	30.98, 42.42	30.35	26.44, 34.25	27.46	22.56, 32.35	31.24	27.29, 35.19	11.85	8.59, 15.11	18.46	15.53, 21.39
4th Quartile [0.41, 0.54]	1394	27.01	30.96	28.21, 33.71	24.78	22.89, 26.68	10.50	4.00, 16.99	7.32	4.52, 10.12	12.88	9.7, 16.06	13.15	10.57, 15.73	35.66	30.37, 40.95	31.87	27.88, 35.86	55.33	49.84, 60.82	43.34	39.06, 47.61
Cigarettes Easily Accessible at Home																						
No	1640	31.14	37.45	34.52, 40.37	27.58	25.63, 29.53	61.39	53.23, 69.56	76.61	72.58, 80.64	67.81	62.32, 73.29	73.70	70, 77.39	62.36	56.98, 67.74	69.66	65.71, 73.6	57.71	52.3, 63.12	70.74	66.84, 74.63
Yes	3764	68.86	62.55	59.63, 65.48	72.42	70.47, 74.37	38.61	30.44, 46.77	23.39	19.36, 27.42	32.19	26.71, 37.68	26.30	22.61, 30.00	37.64	32.26, 43.02	30.34	26.4, 34.29	42.29	36.88, 47.70	29.26	25.37, 33.16
School Relative Density																						
1st Quartile [0.19, 0.38]	1330	17.40	13.56	11.48, 15.64	19.57	19.90, 21.23																
2nd Quartile [0.38, 0.47]	1373	28.45	29.51	26.64, 32.37	27.85	25.88, 29.82																
3rd Quartile [0.47, 0.52]	1323	27.87	29.32	26.57, 32.07	27.04	25.09, 28.99																
4th Quartile [0.52, 0.83]	1378	26.29	27.61	24.97, 30.25	25.54	23.65, 27.43																

Table 3.3 Weighted transition probabilities of smoking status between waves, stratified by school relative network density, Add Health, 1994-2008. 95% confidence intervals of transition probabilities are described between square brackets.

	N	Weighted Prevalence	95% CI	Wave II		Wave III		Wave IV	
Full Study Population				Non Smoker	Current Smoker	Non Smoker	Current Smoker	Non Smoker	Current Smoker
Wave I									
Non Smoker	6825	72.64	71.37, 73.91	79.97 [78.42,81.53]	20.03 [18.48,21.58]	73.69 [71.84, 75.53]	26.31 [24.47, 28.16]	71.59 [69.57, 73.61]	28.41 [26.40, 30.43]
Current	2345	27.36	26.09, 28.63	23.09 [20.35,25.84]	76.91 [74.16,79.65]	33.98 [30.59,37.37]	66.02 [62.63, 69.41]	35.90 [32.28, 39.52]	64.10 [60.48, 67.72]
Wave II									
Non Smoker	4519	65.27	63.70, 66.84		--	78.96 [77.15, 80.78]	21.04 [19.22, 22.85]	74.70 [72.60, 76.80]	25.30 [23.20,27.40]
Current	2081	34.73	33.16, 36.30	--		34.55 [31.66, 37.44]	65.45 [62.56, 68.34]	39.80 [36.63, 42.98]	60.20 [57.02, 63.37]
Wave III									
Non Smoker	3679	63.89	62.16, 65.61	--	--		--	84.26 [82.50, 86.03]	15.74 [13.97, 17.51]
Current	1725	36.12	34.39, 37.84	--	--	--		24.93 [22.19, 27.66]	75.02 [72.34, 77.81]
Wave IV									
Non Smoker	3182	62.72	60.88, 64.57	--	--	--	--	--	--
Current	1578	37.38	35.43, 39.12	--	--	--	--	--	--
School Relative Density Q1									
Wave I									
Non Smoker	1815	76.19	73.27, 79.12	85.30 [82.21, 88.38]	14.7 [11.62, 17.79]	80.12 [76.30, 83.93]	19.88 [16.07, 23.70]	73.62 [69.03, 78.21]	26.38 [21.79, 30.97]
Current	455	23.81	20.88, 26.73	30.71 [23.41, 38.01]	69.29 [61.99, 76.79]	41.46 [32.40, 50.52]	58.54 [49.48, 67.60]	34.15 [24.81, 43.49]	65.85 [56.51, 75.19]
Wave II									
Non Smoker	1265	72.98	69.61, 76.34	--	--	83.77 [80.08, 87.46]	16.23 [12.54, 19.92]	75.59 [70.91, 80.27]	24.41 [19.73, 29.09]
Current	385	27.02	23.66, 30.39	--	--	38.42 [30.64, 46.20]	61.58 [53.80, 69.36]	35.75 [27.57, 43.94]	64.25 [56.06, 72.43]
Wave III									
Non Smoker	1025	71.85	68.01, 75.68	--	--	--	--	80.71 [76.36, 85.06]	19.29 [14.94, 23.64]
Current	305	28.15	24.32, 31.99	--	--	--	--	25.52 [17.96, 33.07]	74.48 [66.93, 82.04]
Wave IV									
Non Smoker	808	64.74	60.40, 69.08	--	--	--	--	--	--
Current	319	35.22	30.92, 39.60	--	--	--	--	--	--
School Relative Density Q2									
Wave I									
Non Smoker	1690	74.82	72.29, 77.34	81.01 [78.18, 83.83]	18.99 [16.17, 21.82]	69.92 [66.25, 73.59]	30.08 [26.41, 33.75]	68.38 [64.44, 72.33]	31.61 [27.67, 35.56]
Current	494	25.18	22.66, 27.71	22.32 [16.50, 28.15]	77.68 [71.85, 83.50]	35.02 [27.87, 42.17]	64.98 [57.83, 72.13]	36.78 [29.22, 44.34]	63.22 [55.66, 70.78]
Wave II									
Non Smoker	1143	67.45	64.44, 70.45	--	--	76.93 [73.27, 80.58]	23.07 [19.42, 26.73]	72.08 [67.92, 76.24]	27.92 [23.76, 32.08]
Current	493	32.55	29.55, 35.56	--	--	32.02 [26.50, 37.54]	67.98 [62.46, 73.50]	39.60 [33.35, 45.85]	60.40 [54.15, 66.65]
Wave III									
Non Smoker	945	62.54	59.15, 65.93	--	--	--	--	85.14 [81.79, 88.49]	14.86 [11.51, 18.21]
Current	428	37.46	34.07, 40.85	--	--	--	--	22.03 [16.99, 27.08]	77.97 [72.92, 83.01]
Wave IV									
Non Smoker	820	61.87	58.27, 65.46	--	--	--	--	--	--
Current	408	38.13	34.54, 41.73	--	--	--	--	--	--
School Relative Density Q3									
Wave I									
Non Smoker	1718	72.56	70.21, 74.91	77.13 [74.04, 80.22]	22.87 [19.78, 25.96]	72.40 [68.79, 76.00]	27.60 [24.00, 31.21]	72.66 [68.92, 76.40]	27.34 [23.60, 31.08]
Current	637	27.44	25.09, 29.79	21.33 [16.54, 26.13]	78.67 [73.87, 83.46]	33.51 [27.37, 39.64]	66.49 [60.36, 72.63]	32.60 [26.33, 38.88]	67.40 [61.12, 73.67]
Wave II									
Non Smoker	1065	62.63	59.63, 65.63	--	--	76.61 [72.88, 80.34]	23.39 [19.66, 27.12]	76.02 [72.08, 79.95]	23.98 [20.05, 27.92]
Current	553	37.37	34.37, 40.37	--	--	38.12 [32.77, 43.47]	61.89 [56.53, 67.23]	40.18 [34.52, 45.83]	59.82 [54.17, 65.48]
Wave III									
Non Smoker	863	62.00	58.70, 65.29	--	--	--	--	84.92 [81.65, 88.19]	15.08 [11.81, 18.35]
Current	460	38.00	34.71, 41.30	--	--	--	--	26.39 [21.35, 31.43]	73.61 [68.57, 78.65]
Wave IV									
Non Smoker	783	61.84	58.40, 65.27	--	--	--	--	--	--
Current	426	38.16	34.73, 41.60	--	--	--	--	--	--
School Relative Density Q4									
Wave I									
Non Smoker	1602	68.82	66.38, 71.25	78.23 [74.95, 81.51]	21.77 [18.49, 25.05]	75.20 [71.72, 78.69]	24.80 [21.31, 28.28]	72.94 [69.04, 76.85]	27.06 [23.15, 30.96]
Current	759	31.19	28.75, 33.62	21.69 [16.82, 26.55]	78.31 [73.45, 83.18]	30.06 [24.07, 36.04]	69.94 [63.96, 75.93]	39.51 [32.73, 46.28]	60.49 [53.72, 67.27]
Wave II									
Non Smoker	1046	60.85	57.76, 63.95	--	--	80.12 [76.85, 83.39]	19.88 [16.61, 23.15]	76.03 [72.07, 79.99]	23.97 [20.01, 27.93]
Current	650	39.15	36.05, 42.24	--	--	31.55 [26.21, 36.89]	68.45 [63.11, 73.79]	42.03 [35.93, 48.14]	57.97 [51.86, 64.07]
Wave III									
Non Smoker	846	62.07	58.78, 65.35	--	--	--	--	85.13 [81.77, 88.49]	14.87 [11.51, 18.23]
Current	532	37.93	34.65, 41.22	--	--	--	--	25.88 [20.70, 31.05]	74.12 [68.95, 79.30]
Wave IV									
Non Smoker	771	63.31	59.78, 6.85	--	--	--	--	--	--
Current	425	36.69	33.15, 40.22	--	--	--	--	--	--

Table 3.4 Weighted transition probabilities of smoking status between waves, stratified by school urbanicity, Add Health, 1994-2008. 95% confidence intervals of transition probabilities are described between square brackets.

	N	Weighted Prevalence	95% CI	Wave II		Wave III		Wave IV	
Urban				Non Smoker	Current Smoker	Non Smoker	Current Smoker	Non Smoker	Current Smoker
Wave I									
Non Smoker	1799	79.01	76.68, 81.34	84.41 [81.62, 87.21]	15.59 [12.79, 18.38]	74.20 [70.51, 77.90]	25.80 [22.10, 29.49]	70.88 [66.66, 75.10]	29.12 [24.90, 33.34]
Current Smoker	447	20.99	18.66, 23.32	27.01 [20.34, 33.68]	72.99 [66.32, 79.66]	42.24 [33.35, 51.14]	57.76 [48.86, 66.65]	38.86 [29.59, 48.14]	61.34 [51.86, 70.41]
Wave II									
Non Smoker	1200	73.91	70.95, 76.88	--	--	81.80 [78.43, 85.17]	18.20 [14.83, 21.57]	75.73 [71.42, 80.03]	24.27 [19.97, 28.58]
Current Smoker	383	26.09	23.12, 29.05	--	--	31.91 [25.15, 38.67]	68.09 [61.33, 74.85]	35.57 [27.86, 43.28]	64.43 [56.72, 72.14]
Wave III									
Non Smoker	932	68.79	65.29, 72.30	--	--	--	--	81.73 [77.77, 85.68]	18.27 [14.32, 22.23]
Current Smoker	350	31.21	27.70, 34.71	--	--	--	--	28.16 [21.35, 34.96]	71.84 [65.04, 78.65]
Wave IV									
Non Smoker	786	65.35	61.39, 69.31	--	--	--	--	--	--
Current Smoker	326	34.65	30.69, 38.61	--	--	--	--	--	--
Suburban									
Wave I									
Non Smoker	3736	72.68	71.00, 74.37	79.05 [76.97, 81.13]	20.95 [18.87, 23.03]	72.79 [70.26, 75.33]	27.21 [24.67, 29.74]	71.85 [69.14, 74.55]	28.15 [25.45, 30.86]
Current Smoker	1269	27.32	25.63, 29.00	22.03 [18.48, 25.57]	77.97 [74.43, 81.52]	34.60 [30.07, 39.13]	65.40 [60.87, 69.93]	36.68 [31.80, 41.56]	63.32 [58.44, 68.20]
Wave II									
Non Smoker	2420	64.49	62.40, 66.59	--	--	78.20 [75.67, 80.84]	21.80 [19.26, 24.33]	75.03 [72.19, 77.86]	24.97 [22.14, 27.81]
Current Smoker	1122	35.51	33.41, 37.60	--	--	35.72 [31.89, 39.54]	64.28 [60.46, 68.11]	41.24 [37.04, 45.44]	58.79 [54.56, 62.96]
Wave III									
Non Smoker	1969	63.39	61.06, 65.73	--	--	--	--	85.64 [83.36, 87.91]	14.36 [12.09, 16.64]
Current Smoker	892	36.61	34.27, 38.94	--	--	--	--	24.31 [20.70, 27.91]	75.69 [72.09, 79.30]
Wave IV									
Non Smoker	1700	63.15	60.69, 65.62	--	--	--	--	--	--
Current Smoker	818	36.85	34.38, 39.31	--	--	--	--	--	--
Rural									
Wave I									
Non Smoker	1290	66.77	63.83, 69.70	78.35 [74.66, 82.04]	21.65 [17.96, 25.34]	76.17 [72.37, 79.96]	23.83 [20.04, 27.63]	72.12 [67.80, 76.44]	27.88 [23.56, 32.20]
Current Smoker	629	33.23	30.30, 36.17	23.55 [17.97, 29.14]	76.45 [70.86, 82.03]	28.98 [22.73, 35.23]	71.02 [64.77, 77.27]	32.98 [26.30, 39.66]	67.02 [60.34, 73.70]
Wave II									
Non Smoker	899	59.99	56.52, 63.46	--	--	78.13 [74.31, 81.94]	21.87 [18.06, 25.69]	72.73 [68.25, 77.22]	27.27 [22.78, 31.75]
Current Smoker	576	40.01	36.54, 43.48	--	--	33.67 [27.85, 39.49]	66.33 [60.51, 72.15]	39.30 [33.03, 45.57]	60.70 [54.43, 66.97]
Wave III									
Non Smoker	778	60.84	57.25, 64.43	--	--	--	--	82.94 [78.99, 86.89]	17.06 [13.11, 21.01]
Current Smoker	483	39.16	35.57, 42.75	--	--	--	--	24.59 [19.21, 29.78]	75.41 [70.02, 80.79]
Wave IV									
Non Smoker	696	59.14	55.29, 62.99	--	--	--	--	--	--
Current Smoker	434	40.86	37.01, 44.71	--	--	--	--	--	--

Table 3.5 Adjusted risk ratios for smoking initiation by selected risk factors, stratified by school relative network density, Add Health, 2001-2002.

		School Relative Density																	
		Full Model (N=5,404)			No Density Model (N=5,404)			1st Quartile (N=1,330)			2nd Quartile (N=1,373)			3rd Quartile (N=1,323)			4th Quartile (N=1,378)		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Smoking Initiation																			
Region	South (REF)																		
	Midwest	1.24	0.95	1.62	1.19	0.90	1.59	**3.73†**	**1.56**	**8.92**	1.25	0.55	2.83	1.05	0.82	1.35	**2.08†**	**1.31**	**3.31**
	Northeast	1.05	0.79	1.41	1.03	0.75	1.41	1.98	0.78	5.01	0.98	0.53	1.81	0.98	0.66	1.45	1.16	0.85	1.58
	West*	0.86	0.61	1.23	0.89	0.65	1.21	1.95	0.89	4.26	0.58	0.27	1.26	1.06	0.74	1.51			
Urbanicity	Suburban (REF)																		
	Rural*	0.90	0.71	1.14	0.87	0.70	1.09	1.00	1.00	1.00	0.98	0.60	1.58	0.82	0.57	1.18	**1.99†**	**1.22**	**3.25**
	Urban	1.01	0.78	1.29	1.04	0.81	1.33	**0.48†**	**0.29**	**0.80**	1.25	0.62	2.51	0.77	0.50	1.16	**2.34†**	**1.53**	**3.59**
Percentage of Nominated Friends that Smoke ‡		**1.07†**	**1.03**	**1.13**	**1.07†**	**1.03**	**1.12**	1.12	0.98	1.27	1.07	0.97	1.18	1.01	0.95	1.08	**1.15†**	**1.07**	**1.24**
School Smoking Prevalence‡		1.05	0.94	1.18	1.05	0.93	1.17	0.78	0.47	1.27	1.02	0.83	1.25	1.01	0.84	1.23	1.05	0.87	1.27
Influence Domain (i.e., "popularity")	1st Quartile (REF)																		
	2nd Quartile	1.02	0.75	1.39	**1.02**	**0.75**	**1.38**	1.39	0.86	2.24	1.01	0.38	2.64	1.05	0.59	1.88	1.48	0.87	2.53
	3rd Quartile	0.93	0.65	1.32	0.93	0.65	1.33	0.71	0.40	1.23	1.02	0.39	2.63	0.98	0.55	1.73	1.09	0.63	1.88
	4th Quartile	0.51	0.19	1.38	0.53	0.20	1.42	0.38	0.09	1.50	**0.05†**	**0.01**	**0.18**	**0.54†**	**0.35**	**0.86**	**7.01†**	**3.02**	**16.27**
Age Heterogeneity	1st Quartile (REF)																		
	2nd Quartile	1.16	0.93	1.45	1.16	0.94	1.45	**1.66†**	**1.11**	**2.49**	0.92	0.70	1.21	1.11	0.73	1.69	1.18	0.64	2.18
	3rd Quartile	1.23	0.97	1.56	1.23	0.97	1.55	1.00	0.60	1.66	1.15	0.90	1.47	1.23	0.87	1.75	1.31	0.66	2.58
	4th Quartile	1.09	0.83	1.43	1.09	0.83	1.43	1.49	0.75	2.96	0.90	0.59	1.38	1.05	0.66	1.66	1.22	0.66	2.28
School Relative Density*	1st Quartile (REF)																		
	2nd Quartile	1.15	0.80	1.64															
	3rd Quartile	1.03	0.68	1.56															
	4th Quartile	0.91	0.61	1.35															

* There are no rural schools in the 1st quartile of relative density, therefore, the 2nd quartile is used as the referent in that stratum. There are also no schools located in the western region of the US in the 4th quartile of relative density

† p < 0.05

‡ Estimate is related to a 10% increase in the predictor

All models adjust for age, sex, race/ethnicity, self-report of depression frequency, school size, and the availability of cigarettes at home. Full models including estimates not presented here are shown in Table 3B

Table 3.6 Adjusted risk ratios for smoking cessation by selected risk factors, stratified by school relative network density, Add Health, 2001-2002.

		School Relative Density																	
		Full Model (N=5,404)			No Density Model (N=5,404)			1st Quartile (N=1,330)			2nd Quartile (N=1,373)			3rd Quartile (N=1,323)			4th Quartile (N=1,378)		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Smoking Cessation																			
Region	South (REF)																		
	Midwest	0.75	0.53	1.05	0.74	0.53	1.05	0.55	0.17	1.82	**0.44†**	**0.21**	**0.92**	0.73	0.44	1.19	0.93	0.32	2.69
	Northeast	0.70	0.48	1.01	0.71	0.50	1.00	1.11	0.28	4.40	0.64	0.27	1.53	0.75	0.31	1.78	0.50	0.27	0.94
	West*	1.25	0.82	1.90	1.25	0.83	1.88	1.15	0.40	3.28	1.47	0.82	2.66	1.14	0.65	2.00	1.00		
Urbanicity	Suburban (REF)																		
	Rural*	0.76	0.51	1.12	0.78	0.56	1.10	---	---	---	1.37	0.84	2.25	1.06	0.75	1.51	**0.51†**	**0.29**	**0.89**
	Urban	1.18	0.87	1.61	1.14	0.84	1.54	0.85	0.32	2.27	1.13	0.75	1.69	1.18	0.63	2.21	**0.42†**	**0.21**	**0.84**
Percentage of Nominated Friends that Smoke‡		0.98	0.94	1.02	0.98	0.94	1.02	1.01	0.90	1.12	0.98	0.90	1.08	1.04	0.95	1.14	**0.90†**	**0.83**	**0.97**
School Smoking Prevalence‡		**0.80†**	**0.69**	**0.92**	**0.83†**	**0.72**	**0.95**	0.74	0.52	1.04	0.86	0.69	1.07	1.08	0.64	1.82	0.75	0.50	1.14
Influence Domain (i.e., "popularity")	1st Quartile (REF)																		
	2nd Quartile	0.66	0.42	1.02	0.67	0.42	1.05	**0.22†**	**0.09**	**0.51**	0.90	0.48	1.69	1.88	0.88	4.01	**0.34†**	**0.16**	**0.73**
	3rd Quartile	1.28	0.78	2.08	1.28	0.77	2.12	0.78	0.24	2.49	1.30	0.54	3.11	2.20	1.08	4.48	1.57	0.57	4.31
	4th Quartile	1.16	0.66	2.04	1.18	0.66	2.09	1.03	0.55	1.95	3.69	0.66	20.71	1.37	0.25	7.45	1.29	0.16	10.58
Age Heterogeneity	1st Quartile (REF)																		
	2nd Quartile	0.85	0.63	1.15	0.86	0.63	1.16	0.86	0.54	1.37	0.89	0.44	1.80	0.70	0.46	1.05	0.97	0.54	1.77
	3rd Quartile	0.99	0.74	1.32	1.00	0.75	1.34	1.20	0.85	1.70	1.03	0.53	1.98	0.60	0.36	1.02	1.44	1.03	2.02
	4th Quartile	**0.66†**	**0.50**	**0.88**	**0.67†**	**0.51**	**0.89**	0.51	0.20	1.31	0.64	0.38	1.07	**0.56†**	**0.35**	**0.89**	0.80	0.49	1.31
School Relative Density*	1st Quartile (REF)																		
	2nd Quartile	1.12	0.72	1.74															
	3rd Quartile	1.24	0.76	2.02															
	4th Quartile	1.20	0.65	2.20															

* There are no rural schools in the 1st quartile of relative density, therefore, the 2nd quartile is used as the referent in that stratum. There are also no schools located in the western region of the US in the 4th quartile of relative density

† p < 0.05

‡ Estimate is related to a 10% increase in the predictor

All models adjust for age, sex, race/ethnicity, self-report of depression frequency, school size, and the availability of cigarettes at home. Full models including estimates not presented here are shown in Table 3B

References

1. Thornberry TP. Toward an Interactional Theory of Delinquency. *Criminology*. 1987;25(4):863-892. doi:10.1111/j.1745-9125.1987.tb00823.x.
2. Maxwell KA. Friends: The Role of Peer Influence Across Adolescent Risk Behaviors. *J Youth Adolesc*. 2002;31(4):267-277. doi:10.1023/A:1015493316865.
3. Hirschi T. A Control Theory of Delinquency. *Causes Delinq*. 1969:16-26.
4. Wiatrowski MD, Griswold DB, Roberts MK. Social Control Theory and Delinquency. *Am Sociol Rev*. 1981;46(5):525. doi:10.2307/2094936.
5. Christakis NA, Fowler JH. The Spread of Obesity in a Large Social Network Over 32 Years. *N Engl J Med*. 2007;357(4):370-379. doi:10.1056/NEJMsa066082.
6. Simmons D, Cohn S, Bunn C, et al. Testing a Peer Support Intervention for People with Type 2 Diabetes: a Pilot for a Randomised Controlled Trial. *BMC Fam Pract*. 2013;14(1):1. doi:10.1186/1471-2296-14-5.
7. Lakon CM, Hipp JR, Timberlake DS. The Social Context of Adolescent Smoking : A Systems Perspective. *Am J Public Health*. 2010;100(7). doi:10.2105/AJPH.2009.167973.
8. Lakon CM, Wang C, Butts CT, Jose R, Timberlake DS, Hipp JR. A Dynamic Model of Adolescent Friendship Networks, Parental Influences, and Smoking. *J Youth Adolesc*. 2014;44(9):1767-1786. doi:10.1007/s10964-014-0187-7.
9. Mercken L, Snijders TAB, Steglich C, Vartiainen E, de Vries H. Dynamics of Adolescent Friendship Networks and Smoking Behavior. *Soc Networks*. 2010;32(1):72-81. doi:10.1016/j.socnet.2009.02.005.
10. Schaefer DR, Adams J, Haas SA. Social Networks and Smoking: Exploring the Effects of Peer Influence and Smoker Popularity through Simulations. *Health Educ Behav*. 2013;40(1 Suppl):24S-32S. doi:10.1177/1090198113493091.
11. Kobus K. Peers and Adolescent Smoking. *Addiction*. 2003;98(Suppl 1):37-55.
12. Alexander C, Piazza M, Mekos D, Valente T. Peers, Schools, and Adolescent Cigarette Smoking. *J Adolesc Heal*. 2001;29(1):22-30. doi:10.1016/S1054-139X(01)00210-5.
13. Ali MM, Dwyer DS. Estimating Peer Effects in Adolescent Smoking Behavior: A Longitudinal Analysis. *J Adolesc Heal*. 2009;45(4):402-408. doi:10.1016/j.jadohealth.2009.02.004.
14. McGloin JM, Sullivan CJ, Thomas KJ. Peer Influence and Context : The Interdependence of Friendship Groups , Schoolmates and Network Density in Predicting Substance Use. *J Youth Adolesc*. 2014;43:1436-1452. doi:10.1007/s10964-014-0126-7.

15. Singh T, Arrazola R, Corey CG. Tobacco Use Among Middle and High School Students — United States, 2011–2015. *MMWR Morb Mortal Wkly Rep*. 2016;65(361):367. doi:http://dx.doi.org/10.15585/mmwr.mm6514a1.

16. U S Department of Health and Human Services. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*.; 2014. doi:NBK179276.

17. Lantz P, Jacobson P, Warner K, et al. Investing in Youth Tobacco Control: a review of Smoking Prevention and Control Strategies. *Tob Control*. 2000;9(1):47-63. doi:10.1136/tc.9.1.47.

18. United States Department of Health and Human Services. *Study on Raising the Minimum Age to Purchase Tobacco Products*.; 2016.

19. King BA, Jama AO, Marynak KL, Promoff GR. Attitudes Toward Raising the Minimum Age of Sale for Tobacco Among U.S. Adults. *Am J Prev Med*. 2015;49(4):583-588. doi:10.1016/j.amepre.2015.05.012.

20. Steinberg L, Monahan KC. Age Differences in Resistance to Peer Influence. *Dev Psychol*. 2009;43(6):1531-1543. doi:10.1037/0012-1649.43.6.1531.Age.

21. Gardner M. Peer Influence on Risk Taking, Risk Preference, and Risky Decision Making in Adolescence and Adulthood: An Experimental Study. *Dev Psychol*. 2005;41(4):625-635. doi:10.1037/a0026993.

22. United States Department of Health and Human Services. *E-Cigarette Use Among Youth and Young Adults. A Report of the Surgeon General*.; 2016.

23. Hammond RA. Social Influence and Obesity. *Curr Opin Endocrinol Diabetes Obes*. 2010;17(716):467-471. doi:10.1097/MED.0b013e32833d4687.

24. Valente TW, Fujimoto K, Chou CP, Spruijt-Metz D. Adolescent Affiliations and Adiposity: A Social Network Analysis of Friendships and Obesity. *J Adolesc Heal*. 2010;45(2):202-204. doi:10.1016/j.jadohealth.2009.01.007.Adolescent.

25. Voorhees CC, Murray D, Welk G, et al. The Role of Peer Social Network Factors and Physical Activity in Adolescent Girls. *Am J Health Behav*. 2014;29(2):183-190. doi:10.1016/j.biotechadv.2011.08.021.Secreted.

26. Farrell AD, Thompson EL, Mehari KR. Dimensions of Peer Influences and Their Relationship to Adolescents' Aggression, Other Problem Behaviors and Prosocial Behavior. *J Youth Adolesc*. 2016:1-19. doi:10.1007/s10964-016-0601-4.

27. Hiday VA. The Social Context of Mental Illness and Violence. *J Health Soc Behav*. 1995;36:122-137.

28. Schreck CJ, Fisher BS, Miller JM. The Social Context of Violent Victimization: A Study of the Delinquent Peer Effect. *Justice Q*. 2004;21(1):23-47. doi:10.1080/07418820400095731.

29. Santor D, Messervey D, Kusumakar V. Measuring Peer Pressure, Popularity, and Conformity in Adolescent Boys and Girls: Predicting School Performance, Sexual Attitudes, and Substance Abuse. *Youth Adolesc*. 2000;29(2):163-182. doi:10.1023/a:1005152515264.

30. Ennett ST, Bauman KE, Hussong A, et al. The Peer Context of Adolescent Substance Use: Findings from Social Network Analysis. *J Res Adolesc*. 2006;16(2):159-186. doi:10.1111/j.1532-7795.2006.00127.x.

31. Smångs M. Delinquency , Social Skills and the Structure of Peer Relations : Assessing Criminological Theories by Social Network Theory. *Soc Forces*. 2017;89(2):609-631.

32. Haynie DL. Delinquent Peers Revisited: Does Network Structure Matter? *Am J Sociol*. 2001;106(4):1013-1057. doi:10.1086/320298.

33. Bramoullé Y, Djebbari H, Fortin B. Identification of Peer Effects Through Social Networks. *J Econom*. 2009;150(1):41-55. doi:10.1016/j.jeconom.2008.12.021.

34. Eckles D, Kizilcec RF, Bakshy E. Estimating Peer Effects in Networks with Peer Encouragement Designs. *Proc Natl Acad Sci*. 2016;113(27):7316-7322. doi:10.1073/pnas.1511201113.

35. Manski CF. Identification of Endogenous Social Effects: The Reflection Problem. *Rev Econ Stud*. 1993;60(3):531-542. doi:10.2307/2298123.

36. Borsari B, Carey KB. How the Quality of Peer Relationships Influences College Alcohol Use. *Drug Alcohol Rev*. 2006;25(4):361-370. doi:10.1080/09595230600741339.How.

37. Duncan GJ, Boisjoly J, Kremer M, Levy DM, Eccles J. Peer Effects in Drug Use and Sex Among College Students. *J Abnorm Child Psychol*. 2005;33(3):375-385. doi:10.1007/s10802-005-3576-2.

38. Eberhardt MS , Ingram DD, Makuc DM et al. *Urban and Rural Health Chartbook, Health United States*. 2001.

39. Doescher MP, Jackson JE, Jerant A, Gary Hart L. Prevalence and Trends in Smoking: A National Rural Study. *J Rural Heal*. 2006;22(2):112-118. doi:10.1111/j.1748-0361.2006.00018.x.

40. McNeely CA, Nonnemaker JM, Blum RW. Promoting School Connectedness: Evidence from the National Longitudinal Study of Adolescent Health. *J Sch Health*. 2002.

41. Harris KM. The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009 [machine-readable data file and documentation]. *Chapel Hill, NC Carolina Popul Center, Univ North Carolina Chapel Hill*. 2009. doi:DOI: 10.3886/ICPSR27021.v9.

42. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am J Epidemiol*. 2004;159(7):702-706. doi:10.1093/aje/kwh090.

43. Valente TW. Network Interventions. *Science (80)*. 2012;337(6090):49-53. doi:10.1126/science.1217330.

44. Backinger CL, Fagan P, Matthews E, Grana R. Adolescent and Young Adult Tobacco Prevention and Cessation: Current Status and Future Directions. *Tob Control*. 2003;12 Suppl 4:IV46-V53. doi:10.1136/tc.12.suppl.

45. United States Department of Health and Human Services. Preventing Tobacco Use Among Youth and Young Adults. A Report of the Surgeon General. 2012:1395.

46. Herrera-Yagüe C, Schneider CMM, Couronné T, et al. The Anatomy of Urban Social Networks and its Implications in the Searchability Problem. *Sci Rep*. 2015:1-27. doi:10.1038/srep10265.

47. Leyden KM. Social Capital and the Built Environment: The Importance of Walkable Neighborhoods. *Am J Public Health*. 2003;93(9):1546-1551. doi:10.2105/AJPH.93.9.1546.

48. Louch H. Personal Network Integration: Transitivity and Homophily in Strong-tie Relations. *Soc Networks*. 2000;22(1):45-64. doi:10.1016/S0378-8733(00)00015-0.

49. Hirsch BJ, Mickus M, Boerger R. Ties to Influential Adults among Black and White Adolescents: Culture, Social Class, and Family Networks. *Am J Community Psychol*. 2002;30(2):289-303. doi:10.1023/A:1014689129369.

50. Gaudin JM, Davis KB. Social Networks of Black and White Rural Families: A Research Report. *J Marriage Fam*. 1985;47(4):1015-1021. doi:10.2307/352345.

51. Johnston LD, O'Malley PM, Miech RA, Bachman JG. *Monitoring the Future National Survey Results on Drug Use, 1975-2016: Overview, Key Findings on Adolescent Drug Use.* Ann Arbor: Institute for Social Research; 2017.

52. IOM (Institute of Medicine). *Public Health Implications of Raising the Minimum Age of Legal Access to Tobacco Products*. Washington, DC; 2015.

Appendix B: Supplemental Material for Chapter 3

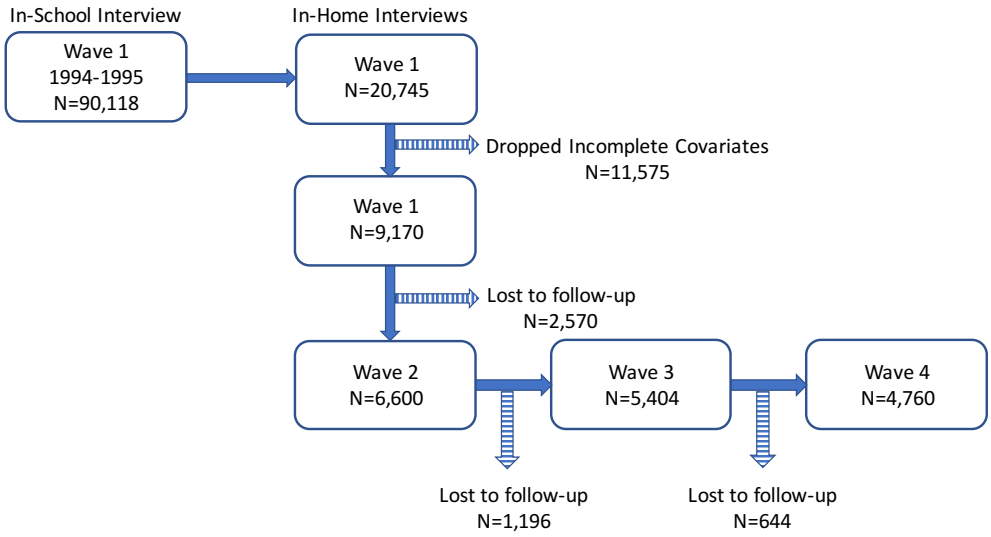


Figure B1 Observation selection process from in-school interviews in wave 1 to in-home interviews in wave 4.

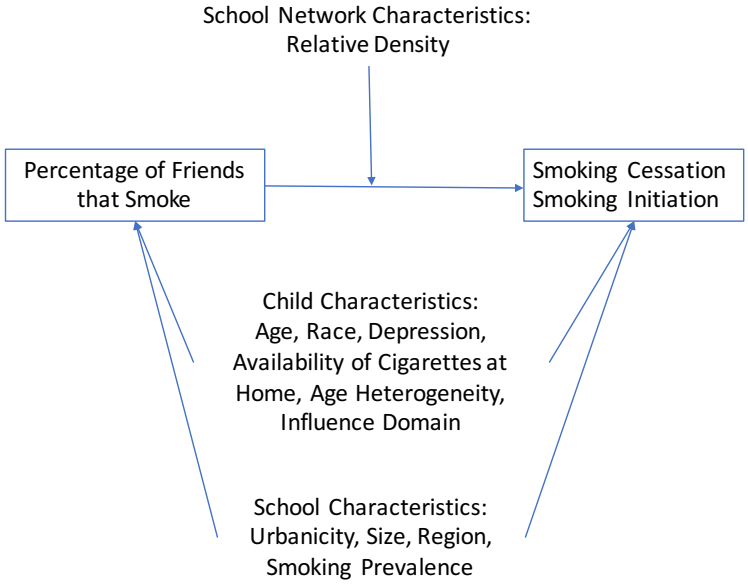


Figure B2 Conceptual Model Diagram of school network (relative) density acting as a modifier on the relationship between percentage of friends that smoke and smoking behavior outcomes.

Table B1 Mean comparisons between retained observations (cleaned in-home data) and observations from population representative data (raw in-home data), Add Health, 1994-1995.

		Raw In-Home Data	Cleaned In-Home Data
Child Age		15.31	15.29
Child Sex (%)			
	Female	50.25	54.28
	Male	49.75	45.72
Child Race (%)			
	Black	15.99	12.90
	Other	17.97	16.73
	White	66.04	70.37
Child Depression (%)			
	Everyday	5.49	5.41
	Often	11.06	10.61
	Occasionally	18.30	17.92
	Rarely	29.13	30.42
	Never	36.02	35.63
Child Age Heterogeneity		0.47	0.49
Child Influence Domain		488.30	535.85
School Urbanicity (%)			
	Rural	21.86	21.86
	Suburban	58.70	19.43
	Urban	19.43	58.70
School Region (%)			
	Midwest	29.49	30.71
	Northeast	15.73	15.46
	South	42.36	42.70
	West	12.42	11.13
School Size (%)			
	Large	33.24	31.92
	Medium	48.79	48.66
	Small	17.97	19.42
Availability of Cigarettes at Home (%)			
	No	67.85	67.89
	Yes	32.15	32.11
Percentage of Friends that Smoke		24.17	25.18
School Relative Density		0.45	0.46
School Smoking Prevalence (%)		35.00	35.00

Table B2 Unadjusted associations between urbanicity and network density as a continuous measure.

Variable		Estimate	Lower 95% CI	Upper 95% CI
Intercept		0.49	0.41	0.59
Urbanicity	Suburban (REF)	1.00	1.00	1.00
	Rural	1.03	0.84	1.26
	Urban	0.78	0.64	0.97

Table B3 Unadjusted risk ratios for smoking initiation and cessation by selected risk factors, stratified by school network density, Add Health, 2001-2002.

		School Relative Density														
		Unadjusted Non-Stratified			1st Quartile			2nd Quartile			3rd Quartile			4th Quartile		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Smoking Initiation																
Region																
	South (REF)															
	Midwest	**1.46†**	**1.11**	**1.93**	**3.39†**	**1.60**	**7.19**	1.66	0.87	3.16	1.05	0.72	1.53	1.23	0.86	1.77
	Northeast	1.26	0.90	1.76	1.84	0.95	3.59	**1.75†**	**1.00**	**3.06**	1.31	0.91	1.88	0.85	0.62	1.15
	West*	0.90	0.59	1.35	**2.66†**	**1.28**	**5.53**	0.75	0.33	1.68	0.86	0.68	1.09			
Urbanicity																
	Suburban (REF)															
	Rural*	0.88	0.63	1.24				0.73	0.28	1.91	0.83	0.59	1.18	0.94	0.61	1.43
	Urban	0.97	0.69	1.36	1.01	0.49	2.07	1.33	0.78	2.25	0.89	0.62	1.27	1.03	0.65	1.63
Percentage of Nominated Friends that Smoke‡		**1.08†**	**1.03**	**1.13**	**1.15†**	**1.02**	**1.29**	1.07	0.98	1.18	1.01	0.95	1.07	**1.14†**	**1.04**	**1.26**
School Smoking Prevalence‡		1.02	0.88	1.19	1.02	0.88	1.19	0.76	0.50	1.15	1.11	0.84	1.48	0.97	0.83	1.13
Influence Domain (i.e., "popularity")	1st Quartile (REF)															
	2nd Quartile	1.05	0.81	1.35	1.21	0.84	1.74	1.06	0.61	1.82	1.08	0.65	1.80	0.84	0.55	1.26
	3rd Quartile	0.89	0.61	1.28	0.55	0.29	1.06	1.32	0.50	3.47	1.10	0.60	1.99	0.71	0.47	1.08
	4th Quartile	0.47	0.22	1.00	0.41	0.13	1.29	0.16	0.01	2.29	0.71	0.46	1.09	**1.55†**	**1.13**	**2.12**
Age Heterogeneity	1st Quartile (REF)															
	2nd Quartile	1.17	0.93	1.46	**1.69†**	**1.10**	**2.59**	0.97	0.71	1.31	1.12	0.75	1.69	1.21	0.64	2.29
	3rd Quartile	1.23	0.96	1.59	0.96	0.56	1.65	1.11	0.83	1.49	1.23	0.87	1.75	1.49	0.67	3.31
	4th Quartile	1.02	0.78	1.33	1.44	0.74	2.80	0.88	0.57	1.36	0.92	0.58	1.45	1.12	0.59	2.14
School Relative Density*	1st Quartile (REF)															
	2nd Quartile	1.32	0.85	2.07	---	---	---	---	---	---	---	---	---	---	---	---
	3rd Quartile	1.34	0.94	1.93	---	---	---	---	---	---	---	---	---	---	---	---
	4th Quartile	1.43	0.94	2.16	---	---	---	---	---	---	---	---	---	---	---	---
Smoking Cessation																
Region																
	South (REF)															
	Midwest	0.94	0.56	1.60	0.99	0.34	2.92	**0.34†**	**0.19**	**0.63**	0.81	0.46	1.44	2.12	0.93	4.87
	Northeast	0.99	0.52	1.89	0.58	0.07	4.49	1.16	0.55	2.45	0.98	0.68	1.42	0.99	0.50	1.95
	West*	1.12	0.76	1.67	0.73	0.26	2.01	1.04	0.52	2.07	1.01	0.71	1.45	---	---	---
Urbanicity																
	Suburban (REF)															
	Rural*	0.69	0.47	1.02	---	---	---	1.18	0.41	3.40	0.87	0.59	1.27	0.53	0.25	1.15
	Urban	1.31	0.82	2.10	0.84	0.30	2.34	**2.22†**	**1.14**	**4.30**	**1.42†**	**1.02**	**1.99**	0.68	0.27	1.73
Percentage of Nominated Friends that Smoke ‡		0.97	0.93	1.02	1.03	0.93	1.15	0.97	0.88	1.07	0.98	0.92	1.04	0.93	0.85	1.01
School Smoking Prevalence‡		**0.74†**	**0.64**	**0.85**	**0.64†**	**0.45**	**0.91**	**0.74†**	**0.56**	**0.97**	0.87	0.65	1.17	0.73	0.48	1.10
Influence Domain (i.e., "popularity")	1st Quartile (REF)															
	2nd Quartile	0.73	0.47	1.15	0.51	0.21	1.28	1.22	0.46	3.27	**1.68†**	**1.02**	**2.77**	0.44	0.17	1.13
	3rd Quartile	1.10	0.74	1.62	0.59	0.24	1.46	1.18	0.58	2.41	**1.73†**	**1.14**	**2.64**	1.15	0.62	2.13
	4th Quartile	1.07	0.67	1.72	1.04	0.55	1.97	1.65	0.71	3.84	1.37	0.59	3.22	**0.52†**	**0.36**	**0.75**
Age Heterogeneity	1st Quartile (REF)															
	2nd Quartile	0.83	0.63	1.11	0.98	0.59	1.62	0.80	0.39	1.63	0.69	0.47	1.00	0.80	0.50	1.26
	3rd Quartile	0.96	0.72	1.27	1.29	0.85	1.95	0.94	0.49	1.83	**0.59†**	**0.36**	**0.96**	1.30	0.83	2.04
	4th Quartile	**0.62†**	**0.48**	**0.81**	0.74	0.31	1.76	**0.54†**	**0.32**	**0.92**	**0.51†**	**0.35**	**0.74**	0.70	0.41	1.22
School Relative Density*	1st Quartile (REF)															
	2nd Quartile	0.93	0.51	1.68	---	---	---	---	---	---	---	---	---	---	---	---
	3rd Quartile	0.77	0.47	1.25	---	---	---	---	---	---	---	---	---	---	---	---
	4th Quartile	0.76	0.38	1.50	---	---	---	---	---	---	---	---	---	---	---	---

* There were no rural schools in the 1st quartile of relative density, therefore, the 2nd quartile is used as the referent in that stratum. There are also no schools located in the western region of the US in the 4th quartile of relative density

† p < 0.05

‡ Estimate is related to a 10% increase in the predictor

Table B4 Full model results of adjusted risk ratios for smoking initiation and cessation by selected risk factors, stratified by school network density, Add Health, 2001-2002.

					School Relative Density														
		Full Model (N=5,404)			No Density Model (N=5,404)			1st Quartile (N=1,330)			2nd Quartile (N=1,373)			3rd Quartile (N=1,323)			4th Quartile (N=1,378)		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Smoking Initiation																			
Intercept		**0.16†**	**0.09**	**0.30**	**0.17†**	**0.09**	**0.31**	0.24†	0.09	0.63	**0.28†**	**0.08**	**1.00**	**0.26†**	**0.09**	**0.80**	**0.05†**	**0.02**	**0.15**
Age		**0.86†**	**0.79**	**0.93**	**0.86†**	**0.79**	**0.93**	1.00	0.80	1.25	0.91	0.81	1.03	0.90	0.80	1.01	**0.72†**	**0.63**	**0.84**
Sex	Male (REF)																		
	Female	**0.71†**	**0.61**	**0.81**	**0.70†**	**0.61**	**0.81**	0.83	0.59	1.17	**0.58†**	**0.47**	**0.71**	**0.73†**	**0.56**	**0.94**	**0.73†**	**0.54**	**0.98**
Race	White (REF)																		
	Black	**0.52†**	**0.35**	**0.78**	**0.53†**	**0.35**	**0.78**	0.57	0.26	1.25	**0.32†**	0.15	0.66	0.72	0.49	1.07	0.47	0.20	1.11
	Other	0.92	0.75	1.15	0.93	0.75	1.15	0.85	0.56	1.30	1.07	0.74	1.55	0.67	0.43	1.06	0.92	0.66	1.27
Depression	Never (REF)																		
	Everyday	1.13	0.78	1.65	1.14	0.79	1.65	1.41	0.59	3.37	0.76	0.29	1.97	1.19	0.64	2.20	1.18	0.61	2.28
	Often	1.16	0.88	1.52	1.16	0.88	1.52	1.62	0.92	2.85	1.05	0.55	2.00	1.13	0.71	1.80	1.05	0.62	1.79
	Occasionally	1.10	0.87	1.39	1.10	0.87	1.39	0.90	0.47	1.72	0.97	0.67	1.39	1.25	0.74	2.14	1.20	0.83	1.74
	Rarely	1.02	0.87	1.19	1.02	0.87	1.19	0.83	0.55	1.24	1.05	0.84	1.33	1.23	0.87	1.74	0.94	0.68	1.30
Availability of Cigarettes at Home	No (REF)																		
	Yes	**1.19†**	1.00	1.43	1.19	0.99	1.42	1.55	0.88	2.71	1.01	0.71	1.43	1.17	0.89	1.53	1.26	0.92	1.71
School Size	Small (REF)																		
	Medium	1.03	0.58	1.83	1.02	0.58	1.79	0.92	0.34	2.50	**10.50†**	**3.36**	**32.81**	1.44	0.80	2.57	**0.48†**	**0.26**	**0.88**
	Large	1.11	0.80	1.55	1.11	0.78	1.58	0.69	0.40	1.20	1.31	0.43	4.03	0.85	0.51	1.42	1.15	0.78	1.70
Region	South (REF)																		
	Midwest	1.24	0.95	1.62	1.19	0.90	1.59	**3.73†**	**1.56**	**8.92**	1.25	0.55	2.83	1.05	0.82	1.35	**2.08†**	**1.31**	**3.31**
	Northeast	1.05	0.79	1.41	1.03	0.75	1.41	1.98	0.78	5.01	0.98	0.53	1.81	0.98	0.66	1.45	1.16	0.85	1.58
	West	0.86	0.61	1.23	0.89	0.65	1.21	1.95	0.89	4.26	0.58	0.27	1.26	1.06	0.74	1.51			
Urbanicity	Suburban (REF)																		
	Rural	0.90	0.71	1.14	0.87	0.70	1.09	---	---	---	0.98	0.60	1.58	0.82	0.57	1.18	**1.99†**	**1.22**	**3.25**
	Urban	**1.01†**	**0.78**	**1.29**	1.04	0.81	1.33	**0.48†**	**0.29**	**0.80**	1.25	0.62	2.51	0.77	0.50	1.16	**2.34†**	**1.53**	**3.59**
Percentage of Nominated Friends that Smoke		**1.07†**	**1.03**	**1.13**	**1.07†**	**1.03**	**1.12**	1.12	0.98	1.27	1.07	0.97	1.18	1.01	0.95	1.08	**1.15†**	**1.07**	**1.24**
School Smoking Prevalence		1.05	0.94	1.18	1.05	0.93	1.17	0.78	0.47	1.27	1.02	0.83	1.25	1.01	0.84	1.23	1.05	0.87	1.27
Influence Domain (i.e., "popularity")	1st Quartile (REF)																		
	2nd Quartile	1.02	0.75	1.39	1.02	0.75	1.38	1.39	0.86	2.24	1.01	0.38	2.64	1.05	0.59	1.88	1.48	0.87	2.53
	3rd Quartile	0.93	0.65	1.32	0.93	0.65	1.33	0.71	0.40	1.23	1.02	0.39	2.63	0.98	0.55	1.73	1.09	0.63	1.88
	4th Quartile	0.51	0.19	1.38	0.53	0.20	1.42	0.38	0.09	1.50	**0.05†**	**0.01**	**0.18**	**0.54†**	**0.35**	**0.86**	**7.01†**	**3.02**	**16.27**
Age Heterogeneity	1st Quartile (REF)																		
	2nd Quartile	1.16	0.93	1.45	1.16	0.94	1.45	**1.66†**	**1.11**	**2.49**	0.92	0.70	1.21	1.11	0.73	1.69	1.18	0.64	2.18
	3rd Quartile	1.23	0.97	1.56	1.23	0.97	1.55	1.00	0.60	1.66	1.15	0.90	1.47	1.23	0.87	1.75	1.31	0.66	2.58
	4th Quartile	1.09	0.83	1.43	1.09	0.83	1.43	1.49	0.75	2.96	0.90	0.59	1.38	1.05	0.66	1.66	1.22	0.66	2.28
School Relative Density*	1st Quartile (REF)																		
	2nd Quartile	1.15	0.80	1.64															
	3rd Quartile	1.03	0.68	1.56															
	4th Quartile	0.91	0.61	1.35															
Smoking Cessation																			
Intercept		1.12	0.54	2.30	1.14	0.58	2.23	3.74	0.91	15.42	0.99	0.26	3.78	0.29	0.03	2.48	2.77	0.38	20.44
Age		0.99	0.90	1.09	0.99	0.90	1.08	1.23	0.91	1.66	0.91	0.71	1.17	0.88	0.75	1.04	1.13	0.96	1.34
Sex	Male (REF)																		
	Female	0.99	0.80	1.23	0.99	0.80	1.23	1.49	0.77	2.87	1.00	0.61	1.64	0.74	0.54	1.02	1.48	0.94	2.32
Race	White (REF)																		
	Black	1.23	0.83	1.81	1.21	0.82	1.78	1.02	0.40	2.58	1.50	0.87	2.59	1.68	0.97	2.90	0.46	0.02	9.08
	Other	0.79	0.57	1.09	0.78	0.57	1.08	0.91	0.46	1.79	0.70	0.38	1.28	0.78	0.41	1.48	0.77	0.31	1.88
Depression	Never (REF)																		
	Everyday	0.64	0.41	1.00	0.64	0.41	1.00	0.44	0.16	1.21	0.87	0.44	1.72	0.54	0.16	1.81	0.58	0.26	1.31
	Often	1.00	0.72	1.38	1.00	0.72	1.38	0.55	0.23	1.28	1.00	0.45	2.25	**1.47†**	**1.00**	**2.16**	0.77	0.49	1.20
	Occasionally	1.00	0.79	1.28	1.00	0.79	1.28	**0.50†**	**0.31**	**0.82**	0.97	0.56	1.67	1.33	0.92	1.92	1.13	0.73	1.76
	Rarely	0.85	0.65	1.10	0.85	0.65	1.10	0.84	0.49	1.46	0.75	0.43	1.31	1.07	0.64	1.78	0.64	0.41	1.01
Availability of Cigarettes at Home	No (REF)																		
	Yes	**0.76†**	**0.63**	**0.93**	**0.76†**	**0.63**	**0.93**	**0.52†**	**0.36**	**0.75**	0.79	0.51	1.24	0.95	0.72	1.24	**0.55†**	**0.37**	**0.82**
School Size	Small (REF)																		
	Medium	0.84	0.46	1.51	0.79	0.42	1.48	0.38	0.11	1.35	0.31	0.04	2.42	0.89	0.17	4.69	1.26	0.27	5.76
	Large	1.06	0.69	1.64	1.05	0.66	1.67	0.84	0.32	2.22	0.96	0.45	2.06	0.70	0.29	1.71	1.14	0.63	2.06
Region	South (REF)																		
	Midwest	0.75	0.53	1.05	0.74	0.53	1.05	0.55	0.17	1.82	**0.44†**	**0.21**	**0.92**	0.73	0.44	1.19	0.93	0.32	2.69
	Northeast	0.70	0.48	1.01	0.71	0.50	1.00	1.11	0.28	4.40	0.64	0.27	1.53	0.75	0.31	1.78	**0.50†**	**0.27**	**0.94**
	West	1.25	0.82	1.90	1.25	0.83	1.88	1.15	0.40	3.28	1.47	0.82	2.66	1.14	0.65	2.00			
Urbanicity	Suburban (REF)																		
	Rural	0.76	0.51	1.12	0.78	0.56	1.10				1.37	0.84	2.25	1.06	0.75	1.51	**0.51†**	**0.29**	**0.89**
	Urban	1.18	0.87	1.61	1.14	0.84	1.54	0.85	0.32	2.27	1.13	0.75	1.69	1.18	0.63	2.21	**0.42†**	**0.21**	**0.84**
Percentage of Nominated Friends that Smoke		0.98	0.94	1.02	0.98	0.94	1.02	1.01	0.90	1.12	0.98	0.90	1.08	1.04	0.95	1.14	**0.90†**	**0.83**	**0.97**
School Smoking Prevalence		**0.80†**	**0.69**	**0.92**	**0.83†**	**0.72**	**0.95**	0.74	0.52	1.04	0.86	0.69	1.07	1.08	0.64	1.82	0.75	0.50	1.14
Influence Domain (i.e., "popularity")	1st Quartile (REF)																		
	2nd Quartile	0.66	0.42	1.02	0.67	0.42	1.05	**0.22†**	**0.09**	**0.51**	0.90	0.48	1.69	1.88	0.88	4.01	**0.34†**	**0.16**	**0.73**
	3rd Quartile	1.28	0.78	2.08	1.28	0.77	2.12	0.78	0.24	2.49	1.30	0.54	3.11	**2.20†**	**1.08**	**4.48**	1.57	0.57	4.31
	4th Quartile	1.16	0.66	2.04	1.18	0.66	2.09	1.03	0.55	1.95	3.69	0.66	20.71	1.37	0.25	7.45	1.29	0.16	10.58
Age Heterogeneity	1st Quartile (REF)																		
	2nd Quartile	0.85	0.63	1.15	0.86	0.63	1.16	0.86	0.54	1.37	0.89	0.44	1.80	0.70	0.46	1.05	0.97	0.54	1.77
	3rd Quartile	0.99	0.74	1.32	1.00	0.75	1.34	1.20	0.85	1.70	1.03	0.53	1.98	0.60	0.36	1.02	**1.44†**	**1.03**	**2.02**
	4th Quartile	**0.66†**	**0.50**	**0.88**	**0.67†**	**0.51**	**0.89**	0.51	0.20	1.31	0.64	0.38	1.07	**0.56†**	**0.35**	**0.89**	0.80	0.49	1.31
School Relative Density*	1st Quartile (REF)																		
	2nd Quartile	1.12	0.72	1.74															
	3rd Quartile	1.24	0.76	2.02															
	4th Quartile	1.20	0.65	2.20															

* There are no rural schools in the 1st quartile of relative density, therefore, the 2nd quartile is used as the referent in that stratum. There are also no schools located in the western region of the US in the 4th quartile of relative density

† p < 0.05

‡ Estimate is related to a 10% increase in the predictor

Table B5 Full model results of adjusted risk ratios for smoking initiation and cessation by selected risk factors, stratified by sex, Add Health, 2001-2002.

		Female (N=3,114)			Male (N=2,290)		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Initiation							
Intercept		**0.10†**	**0.05**	**0.21**	**0.15†**	**0.06**	**0.37**
Age		**0.86†**	**0.77**	**0.95**	**0.85†**	**0.77**	**0.94**
Race	White (REF)						
	Black	**0.40†**	**0.25**	**0.63**	0.64	0.38	1.08
	Other	0.88	0.64	1.22	0.98	0.75	1.28
Depression	Never (REF)						
	Everyday	1.35	0.80	2.27	1.22	0.72	2.08
	Often	1.24	0.85	1.83	0.96	0.60	1.55
	Occasionally	1.30	0.93	1.81	0.90	0.62	1.29
	Rarely	0.90	0.66	1.22	1.11	0.89	1.38
Availability of Cigarettes at Home	No (REF)						
	Yes	1.18	0.97	1.45	1.19	0.90	1.57
School Size	Small (REF)						
	Medium	1.28	0.80	2.05	1.06	0.70	1.62
	Large	1.28	0.56	2.96	0.91	0.49	1.67
Region	South (REF)						
	Midwest	**1.34†**	1.00	1.81	1.13	0.78	1.64
	Northeast	1.20	0.85	1.68	0.95	0.66	1.37
	West	0.69	0.43	1.10	0.86	0.53	1.41
Urbanicity	Suburban (REF)						
	Rural	0.86	0.64	1.15	1.07	0.80	1.45
	Urban	0.84	0.61	1.15	1.23	0.87	1.73
Percentage of Nominated Friends that Smoke‡		**1.10†**	**1.04**	**1.16**	1.05	0.99	1.11
School Smoking Prevalence‡		**1.19†**	**1.03**	**1.37**	0.94	0.82	1.07
Influence Domain (i.e., "popularity")	1st Quartile (REF)						
	2nd Quartile	0.76	0.44	1.30	1.22	0.82	1.82
	3rd Quartile	0.61	0.36	1.03	1.34	0.87	2.06
	4th Quartile	0.40	0.12	1.36	0.72	0.37	1.40
Age Heterogeneity	1st Quartile (REF)						
	2nd Quartile	1.20	0.82	1.74	1.11	0.87	1.41
	3rd Quartile	1.31	0.90	1.89	1.17	0.90	1.52
	4th Quartile	1.04	0.73	1.47	1.06	0.78	1.44
School Relative Density*	1st Quartile (REF)						
	2nd Quartile	0.88	0.54	1.44	1.60	0.99	2.61
	3rd Quartile	0.77	0.44	1.34	1.45	0.84	2.51
	4th Quartile	0.66	0.37	1.19	1.25	0.74	2.10
Cessation							
Effect							
Intercept		0.71	0.24	2.12	1.40	0.42	4.71
Age		0.92	0.80	1.06	1.07	0.91	1.25
Race	White (REF)						
	Black	**1.66†**	**1.06**	**2.60**	0.59	0.27	1.31
	Other	0.83	0.54	1.28	0.66	0.43	1.02
Depression	Never (REF)						
	Everyday	**0.61†**	**0.37**	**0.99**	0.53	0.20	1.43
	Often	1.12	0.73	1.72	0.62	0.29	1.32
	Occasionally	1.03	0.67	1.58	0.98	0.64	1.51
	Rarely	1.03	0.66	1.60	**0.67†**	**0.47**	**0.96**
Availability of Cigarettes at Home	No (REF)						
	Yes	**0.67†**	**0.53**	**0.84**	0.91	0.66	1.25
School Size	Small (REF)						
	Medium	1.15	0.67	1.96	0.95	0.42	2.13
	Large	0.88	0.34	2.30	0.85	0.29	2.48
Region	South (REF)						
	Midwest	0.62	0.37	1.06	0.67	0.38	1.18
	Northeast	0.69	0.44	1.09	**0.32†**	**0.12**	**0.87**
	West	1.11	0.57	2.15	0.98	0.39	2.43
Urbanicity	Suburban (REF)						
	Rural	0.68	0.38	1.21	0.70	0.35	1.40
	Urban	1.12	0.73	1.72	1.20	0.64	2.26
Percentage of Nominated Friends that Smoke‡		0.98	0.93	1.04	0.98	0.92	1.04
School Smoking Prevalence‡		0.92	0.74	1.13	0.71	0.51	0.98
Influence Domain (i.e., "popularity")	1st Quartile (REF)						
	2nd Quartile	0.64	0.34	1.20	0.66	0.25	1.74
	3rd Quartile	1.33	0.76	2.33	1.73	0.71	4.22
	4th Quartile	0.89	0.32	2.51	1.23	0.43	3.49
Age Heterogeneity	1st Quartile (REF)						
	2nd Quartile	0.79	0.50	1.24	0.75	0.44	1.27
	3rd Quartile	0.95	0.67	1.36	1.03	0.65	1.63
	4th Quartile	0.70	0.49	1.00	0.62	0.36	1.07
School Relative Density*	1st Quartile (REF)						
	2nd Quartile	0.95	0.51	1.75	1.45	0.59	3.58
	3rd Quartile	0.96	0.48	1.95	2.02	0.74	5.55
	4th Quartile	0.96	0.38	2.42	1.69	0.61	4.72

† p < 0.05

‡ Estimate is related to a 10% increase in the predictor

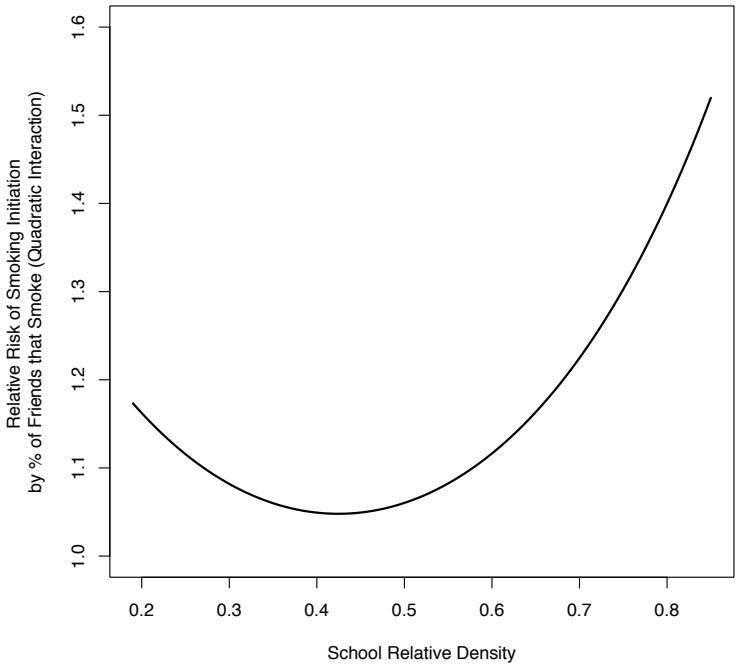


Figure B3 Relative risk of smoking initiation for a 10% increase in percentage of friends that smoke by school relative density. This figure assumes quadratic statistical interaction (i.e., non-linear effect modification) by school network density with the percentage of friends that smoke on the probability of smoking initiation.

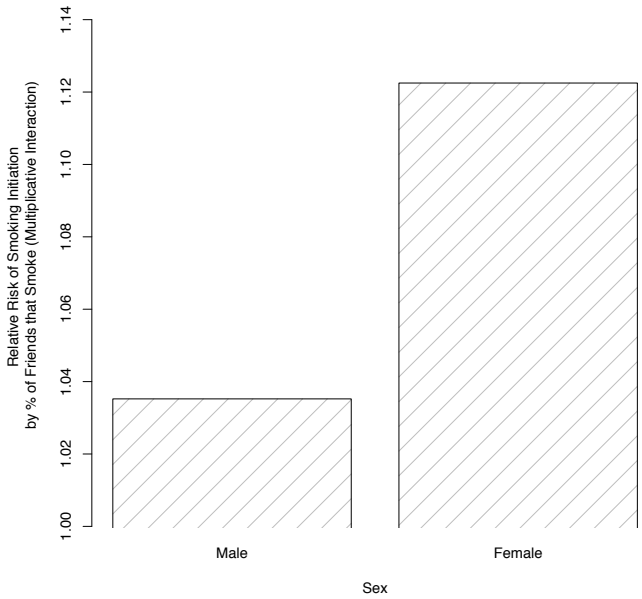


Figure B4 Relative risk of smoking initiation for a 10% increase in percentage of friends that smoke by sex. This figure assumes multiplicative statistical interaction by sex with the percentage of friends that smoke on the probability of smoking initiation.

CHAPTER 4

Connecting Statistical Analyses to Mechanistic Models: Pathway Validation

Background

As demonstrated in the previous chapter, and in prior research, a portion of adolescent smoking behavior could be attributed to peer influence.[1–5] Additionally, differences in peer influence by network density have been demonstrated for other substance use and delinquency more broadly.[6–8] However, the nature of data-driven statistical methods, such as those utilized in Chapter 3 of this dissertation, provide only associations and evidence supporting the causal frameworks used to justify the inclusion of covariates in the statistical model. That is, while statistical regression methods can provide supporting evidence for causal frameworks by revealing associations, these methods are not suited to reveal the underlying mechanisms that produce these associations without additional assumptions and modifications. Previous simulation models of social networks have attempted to disentangle the effects between peer selection and peer influence on substance use,[1] social media sharing,[2] and delinquency,[9] and find that many studies can potentially over- or under- estimate the level of peer influence effects due to difficulties disentangling them from peer selection effects (i.e., potential reverse causality).[10,11] In this chapter, we hypothesize that peer influence diffusion through social adolescent networks (i.e., assuming peer selection) is a key mechanism that generates differences in peer influence of smoking behavior by network density. To explore this hypothesis, we apply the pertinent pathways posited in the previous chapter in a mechanistic computational model to isolate the independent effects of peer influence by network structure. This aim presents an explanatory mechanism integrated into a mechanistic model, that when incorporated with empirical data, can bring insights into the meaning of the statistical associations observed between urbanicity, network density, peer influence, and smoking behavior. Furthermore, these results may provide insights into a range of potential social policy levers that can help further reduce US smoking prevalence in adolescents.

Introduction

Friendship networks have been consistently shown to contribute to the propagation of smoking behavior among adolescents.[4,7,12,13] However, whether smoking behavior outcomes, such as smoking initiation among non-smokers or smoking cessation among current smokers, are due primarily to peer selection, peer influence, or both equally, has also been a topic of considerable contention.[2,10,11,14,15] Moreover, many studies have examined how complex social networks contribute to adolescent smoking initiation,[16] yet most either utilize data-driven statistical methods or mechanistic computational methods that prioritize individual-level network characteristics, such as popularity or other centrality measures, while ignoring broader contextual network characteristics.[1,16] Despite evidence supporting differences in the nature of friendship[17] and smoking behavior across urban contexts,[18,19] we found no published studies explicitly focusing on the differences in associations between friendship influence and urbanicity with smoking behavior outcomes using the National Longitudinal Study of Adolescent to Adult Health (Add Health) networks data. To our knowledge, there are also no mechanistic modeling studies using the Add Health networks data to examine how school network structure can independently change the propagation outcomes of smoking behavior through mechanisms of friendship influence. Thus, although many studies suggest that network density, and perhaps urbanicity by proxy, can change the nature of friendship influence on adolescent behavior,[6,7,13] the mechanisms explaining why we observe these differences remain unclear. The complexities of social networks may result in analytic or conceptual biases that could contribute to erroneous interpretations of social influence. Such results may then encourage ineffective interventions or policies if social influence does not truly cause adolescents to change their health behaviors,[20] or if its influence is lower than inferred from associations using non-mechanistic analysis methods.

While the statistical regression models shown in Chapter 3 demonstrate potential effect modification of peer influence on smoking behavior by network structure, the validity of our model results relies on the validity of the causal framework that motivated the statistical analyses. Mechanistic computational models (such as ABMs) can bring additional value to the causal framework by providing further explanations of the interacting processes that can generate empirical data patterns. These methods are also able to experiment with and test processes in well-defined and closed experimentation environment, free of unknown confounding. Given the literature supporting differences in social network characteristics by built environment (e.g.,

urbanicity and regional location),[21,22] we hypothesize that because friendship influence on smoking behavior changes due to network density, the same differences may be observed through a higher-level proxy measure, such as urbanicity. We test this hypothesis by examining whether the magnitude and direction of friendship influence on smoking initiation and cessation changes by urbanicity designations, an extension of the findings from Chapter 3. Next, we hypothesize that the observed differences from our statistical models can be accounted for solely by differences in network structure, and when assuming competing percolation processes of smoking cessation and smoking initiation. We test this hypothesis by implementing the network structure data from Add Health in an agent-based networks simulation model, assuming that friendship influence on smoking behavior, for both smoking initiation and cessation, spreads across directed friendship ties. This model excludes all other risk factors to isolate the effects of network structure and peer influence on smoking prevalence outcomes. After model simulations are finished, we match individuals in the model to corresponding empirical observations to determine whether the computational model can recover the qualitative characteristics of the outcomes observed from the mixed-effects statistical models presented here and in Chapter 3. This work yields two important results. First, we provide insights into the basic processes contributing to smoking behavior among US adolescents. Second, by showing how statistical and abstract computational methods can coordinate to advance causal frameworks in epidemiology, we demonstrate the role that abstract models can play in the study of causality in epidemiology and public health.

Methods

Data

The statistical methods of chapter uses networks data from the Add Health study, a school-based longitudinal study of a nationally representative sample of adolescents in the United States attending grades 7-12 in 1994.[23] Urbanicity designations were determined in the primary sampling frame of Add Health, using Quality Education Data to ensure that the selected schools in the study are representative of US schools in terms of urbanicity, region, size, type, and race/ethnicity distribution.[23] In wave 1 (1994) of the in-school questionnaire, adolescents were asked to select 5 of their best female friends, in addition to 5 of their best male friends, with many selecting fewer than 10 total friends. Add Health only retained network data for schools with at least 50% participation by the student body on friendship questionnaires to ensure that the social network

samples are representative of the whole school.[23] While the simulation methods of this chapter use only friendship structure data from the Add Health study, adolescent identification values are retained in the model order to identify what adolescent and school characteristics (i.e., sex, race, network density, urbanicity) may be associated with friendship network structure, and whether these associations may give rise to differences in social patterning of smoking behavior.

Statistical Analysis

To determine the potential mechanisms contributing to the patterning of social influence by broader contextual factors in the Add Health survey, we test whether friendship influence on smoking behavior differs by school urbanicity designations. We begin by applying the same data selection method and subsequent dataset from Chapter 3. Then, using univariate analysis, we estimate the relative risks and 95% confidence intervals of risk factors of interest, hypothesized to contribute to smoking initiation and cessation in wave 3 (2001-2002) among non-smokers and smokers, respectively, in wave 1 (1994). These results are stratified by urban, suburban, and rural designations of schools. Next, to determine potential effect modification by urbanicity on the pathway between peer influence and smoking behavior, we use multivariate mixed-effects analyses, stratified by urbanicity, with schools acting as the random intercept while adjusting for fixed effects of all variables analyzed in the univariate models, in addition to a range of potential confounders, such as age, sex, race/ethnicity, self-report of depression, school size, and the availability of cigarettes at home. We additionally test for heterogeneity through interaction analysis of the relationship between urbanicity and peer influence on smoking behavior. These analyses were performed by employing a modified Poisson approach to estimate relative risks and using PROC GLIMMIX in SAS 9.4 to account for the weighted multi-level survey design of the Add Health survey.

Simulation Modeling

Next, we initialize network structures in a separate simulation model, using friendship data from the Add Health study. We then initialize smoking prevalence at random across the entire Add Health population, and then propagate smoking behavior in each of the school networks accounting for peer effects as described below. We additionally matched network identification data to in-school response questionnaires to determine sex (male or female) and race (white, non-

Hispanic black, or other) of participants in the model. "Other" race is an aggregate of individuals identifying as of Hispanic or Latino origin, Asian or Pactific Islander, or American Indian or Native American. Adolescents included in the network data with missing values for sex and race were randomly assigned values for simplicity, and low levels of missing data (less than 1% from each subgroup, and missing at random). While these factors were excluded from the model mechanisms and processes entirely, their associated network characteristics and community structure may differ substantially and thus contribute to differences in the patterning of smoking behavior by urbanicity or network density. Aside from the assumption that the Add Health network structures are representative of US adolescent friendship networks, we make two additional assumptions of smoking behavior among adolescents: 1) that peer influence of smoking behavior manifests in populations like information or infectious disease diffusion; and 2) that peer influence of smoking initiation behaves similarly to peer influence of smoking cessation (i.e., equivalent forces in competing directions).

In operationalizing peer influence and the subsequent diffusion of smoking state changes, we construct a transition probability by time step. This transition probability weights the probability of smoking status changes by the percentage of nominated smoking and non-smoking friends, derived from a formulation of the odds ratio (OR) in logistic regression (see Appendix C for details of this derivation):

$$P(Transition)_{t+1} = \frac{1}{1 + e^{-baseline - OR*pofs_{t-1}}}$$

Where *t* is time and *baseline* is a parameter of the baseline transition rate between states that captures the individual propensity to either start of quit smoking (Baseline_start and Baseline_quit). OR values are also a variable parameter that differs by whether the focal adolescent will quit (OR_quit) or start (OR_init). *pofs* is the percentage of friends that are either non-smokers for a currently smoking focal adolescent, or the percentage of friends that are current smokers, for a non-smoking focal adolescent (from 0 to 100 in units). In other words, the term *OR*pofs* captures the peer influence effects. When *pof*s is 0, the transition probability depends only on the *baseline* parameter.

At every time step in the model, each additional percentage point increase in the percentage of nominated friends that smoke increases the odds of a non-smoker becoming a smoker. Similarly,

each percentage point increase in the percentage of nominated friends that are non-smokers increases the odds of a smoker becoming a non-smoker. The magnitude of the odds of non-smokers becoming smokers and smokers becoming non-smokers are variable parameters in this model, which allow us to examine how smoking behavior outcomes of the model change due to changes in the magnitude of friendship influence on smoking initiation (OR_init) and smoking cessation (OR_quit). The odds of a non-smoker initiating smoking increase the baseline probability of smoking initiation (Baseline_start), while the odds of a current smoker quitting increase the baseline probability of quitting smoking (Baseline_quit). For interpretability purposes, we present a select subset of 9 model outcomes for combinations of two values for Baseline_start and Baseline_quit (0.01, and 0.005), along with two values for OR_init and OR_quit (1.000, 1.025). Heatmaps of how the combinations of our parameters change model smoking prevalence by urbanicity and density are shown in Figures C1-C3.

In this simulation model, the conceptualization of time is not fixed to data, and is therefore abstractly defined in the context of the model dynamics. All model dynamics occur as discrete events in discrete time. That is, each agent has one opportunity at each time step to become a smoker or to quit smoking, depending on their smoking state at that time step. Each model runs for 150 time steps (i.e., until qualitative equilibrium), and all experiments are repeated 15 times for each case to account for stochasticity in each model run. Again, no assumptions in this model are made regarding differences of peer influence by urbanicity, network density, sex, or race. Heterogeneity of agents in this model come exclusively from the friendship network data (i.e., the number of ties for each adolescent is heterogeneous). Importantly, heterogeneity in the percentage of friends that smoke arises from this data. That is, the number of friendship selections affects the percentage of friends that smoke overall. All other variables, such as the baseline probability of smoking initiation and cessation, and the coefficient of the odds of smoking initiation and cessation, are homogenous within the population. This is done to isolate the impact of potential differences in the underlying school networks that contribute to the diffusion of smoking behaviors by friendship influence. Finally, given the potential association between urbanicity and school-level network density, continuous network density values are converted to tertiles for comparison purposes to the three urbanicity levels, although quartiles of network density are retained for the statistical model. Further details discussing the derivation and operationalization of the probability

of smoking state changes by peer influence (C1.1), along with pseudo code of the model (C1.2), can be found in Appendix C.

Results

We begin with an extension of the results from Chapter 3. Table 4.1 shows select statistical multi-level outcomes of smoking initiation in wave 3 among non-smokers in wave 1, and smoking cessation in wave 3 among current smokers in wave 1, stratified by urbanicity. Univariate outcomes for parameters of interest are included in Table C1 and complete model results in Table C2. We observe that among suburban (RR=1.09, 95% CI: 1.02, 1.16) and urban (RR=1.09, 95% CI: 1.02, 1.17) schools, the proportion of friends is important for smoking initiation, though the effect estimate of school-level smoking prevalence for smoking initiation among adolescents in rural schools (RR=1.25, 95% CI: 1.02, 1.51) is larger than that of the percentage of nominated friends that smoke (RR=1.01, 95% CI: 0.93, 1.11), which does not reach significance. For smoking cessation, we observe that school smoking prevalence plays a large role in predicting smoking cessation in urban schools (RR=0.73, 95% CI: 0.61, 0.88), though the effects are smaller in suburban (RR=0.97, 95% CI: 0.74, 1.26) and rural schools (RR=0.87, 95% CI=0.63, 1.20). We additionally find that adolescents attending schools in the 3rd and 4th quartile of network density in rural schools are 0.56 (95% CI: 0.32, 0.98) and 0.46 (95% CI: 0.23, 0.94) times less likely, respectively, to quit smoking than adolescents attending rural schools in the 2nd quartile of network density. Interaction terms between urbanicity and peer influence in predicting smoking initiation and cessation were not found to be statistically significant. Table 3.1 presents the descriptive statistics of the population used for the analyses presented here.

Overall, 69,408 adolescents in 112 schools participated in the in-school friendship networks survey of Add Health. A glossary of network terms and their definitions are available in Table 4.2; network terms previously defined in Table 3.1 are not included in this table. Table 4.3 shows descriptive statistics of the population used in the networks models, and of the networks themselves, stratified by urbanicity and density tertiles. We also show descriptive statistics of networks stratified by sex in Table C3. We observe rural networks have higher number of contacts (degree) and higher excess degree, which is a measure of the efficiency of the network in transmitting communicable traits (see Table 4.2 for network measure definitions). Additionally, we find in the data that a greater percentage of adolescents in rural schools are white and have

higher network centrality measures on average (i.e., degree, excess degree, betweenness, closeness, modularity) than adolescents attending urban or suburban schools. We also observe that rural schools have higher mean and median density than suburban or urban schools, with urban schools reporting the lowest values for all descriptive variables except in cases of the percentage of non-white adolescents. Shifting our focus to density, we observe that most urban schools are concentrated in the 1st tertile of network density, which also contains no rural schools. Adolescents with higher levels of network centrality measures also attend schools in the highest levels of network density. Similarly, white students comprise of 71% of the adolescents attending schools in the 3rd tertile of network density, while schools in the 1st tertile of network density comprise of a lower percentage of white (35%) adolescents, and a higher percentage of non-Hispanic black (25%) and other race (30%) adolescents.

Figure 4.1 shows model results from the agent-based networks model, where panel columns represent static values for OR_init and OR_quit, and panel rows represent static values for Baseline_start and Baseline_quit. Each graph represents a unique combination of these four values. In the cases where OR_quit is low (i.e., 1.000) and OR_init is high (i.e., 1.025), we find that regardless of whether the baseline initiation start or quit values are high (0.01) or low (0.005), we observe increases from smoking prevalence at initialization. In contrast, when OR_quit is high and OR_init is low, we observe that smoking prevalence decreases from initialization values for all cases. When the OR values are equal for both friendship influence of smoking initiation and friendship influence of cessation, we observe decreasing and then slightly increasing prevalence rates when baseline initiation and cessation values are equal. However, decreasing the baseline initiation value to 0.005 while holding the baseline cessation value at 0.01 results in decreases to smoking prevalence. Alternatively, decreasing the baseline cessation rate to 0.005 and maintaining the baseline initiation value at 0.01 results in increases to smoking prevalence when ORs are equal.

Notably, we observe the emergence of differences in smoking prevalence by network density and urbanicity with varying values of ORs. In all cases where the differences emerge in the smoking prevalence, we observe that urban and schools in the first tertile (T1) of density report lowest smoking prevalence. Similarly, rural schools and schools in the third tertile (T3) of density report the highest smoking prevalence. These patterns of smoking prevalence by urbanicity and density are consistent across all cases where we can observe clear differences in smoking prevalence by these strata.

Figure 4.2 shows smoking prevalence stratified by race. We observe similar outcomes to that of urbanicity, with non-Hispanic blacks demonstrating lower smoking prevalence than whites, and other races demonstrating smoking prevalence values slightly higher than blacks in all cases. Figure 4.3 expands on the information from Figure 4.2 by comparing races across urban environments. We observe that all urban groups by race present lower smoking prevalence than rural and suburban whites. Moreover, rural blacks surpass urban whites in smoking prevalence for all cases when observable patterns distinguishing smoking prevalence by these parameters emerge. Finally, Figure 4.4 shows model results by sex. In cases where differences between males and females emerge, females present with higher smoking prevalence than males. Across all cases in Figures 4.1-4.4, we observe differences in the dynamics of smoking prevalence, with the same networks generating dramatically different smoking prevalence outcomes.

Discussion

From the statistical models, we observe that the potential effects of friendship influence on both smoking initiation and cessation vary across urbanicity designations, although statistical tests for heterogeneity do not reach significance. Given that we adjust for network density in the stratified analysis, these results may suggest that the effects of network density (found in Chapter 3) may outweigh any additional differences by urbanicity in our statistical analyses. Nevertheless, among adolescents in rural schools, school smoking prevalence is positively associated with and has a greater influence on smoking initiation than the effect of the percentage of nominated friends that smoke. Differently, among urban and suburban schools, a higher percentage of nominated friends that smoke is a stronger factor for smoking initiation than school-level smoking prevalence, suggesting potentially stronger contextual effects on adolescent smoking initiation in rural areas than in suburban or urban areas. In urban schools, current smokers are much less likely to quit if school-level smoking prevalence is high, and while we observe this association across all urbanicity designations, it does not reach statistical significance in suburban or rural areas. Finally, while we observe that the 3rd and 4th quartile of network density is significantly associated with lower likelihood of smoking cessation in rural schools relative to the 1st quartile of network density, the effect estimates of network density quartiles are potentially non-linear in suburban and urban areas, although these estimates are not statistically significant.

Results from the simulation models demonstrate that racial and urbanicity differences in adolescent smoking behavior can be generated from a simple diffusion assumption through network structure alone. In other words, differences in network configurations by these factors are sufficient to generate the higher smoking prevalence in rural and high density schools. However, these differences vary greatly between different combinations of baseline values and the magnitude of the influence of peers. This suggests a potential feedback between peer selection and peer influence in a way that is strongly influenced by network structure. This may be due to a variety of factors, including what friendship means in different contexts, and how they manifest as explicit pathways through which influence can change behavior. The possibility for differences to emerge in smoking outcomes by urbanicity and race, demonstrate that the effects of network structure cannot be attributed to race alone, although individual network communities and subsequent structure may differ by race across similar urbanicity designations. Similarly, we observe that urbanicity differences in network structure may overpower the peer effects of racial differences in network structure, given the patterning of higher prevalence of smoking among blacks in rural areas than whites in urban areas.

We observe that model results differ by sex, with females propagating smoking prevalence more effectively than males in our simulations. Smoking prevalence is known to be higher in male versus female adolescents (Table 3.1),[24] which could be due to differences in tendencies towards risk behaviors by sex,[25] in addition to differences in social structure by sex.[26] However, while centrality and degree measures may point to popularity within a network, perhaps as a proxy for peer influence, these measures are not largely different between males and females in the network data used in this chapter (Table 4.3 and Table C3). Notably, female and male networks are not independent from each other in the Add Health representations, so the larger smoking prevalence values among females in our simulation develops from a network structure that includes connections between all sex combinations. More complex characteristics of network influence may therefore be responsible for the emergence of these differences, resulting in females experiencing peer influence to a greater, and perhaps different, extent than males. This outcome is potentially supported by studies that find that adolescent females are less likely to quit smoking once they start,[27] and with female adolescent smoking being linked to concerns regarding weight and appearance.[28] Additionally, the disparate results by sex presented here and in Chapter 3 (Table

B8) perhaps lend credence to the hypothesis that adolescent female social networks perpetuate smoking behavior more easily than adolescent male social networks on average.[26,29]

There are several limitations to the results of this study. The first obvious limitation is in its abstract nature. We do not make predictions, nor use individual-level empirical data other than the peer network, to generate our results, and thus, the absolute outcomes produced by this model are not directly interpretable. Additionally, like any other mechanistic simulation model, it is a simplified representation of the world. While the proportion of friends that smoke changes at each time step, and in turn, changes the magnitude and direction of smoking influence over time, we do not implement dynamic networks of changing friends and may observe substantially different results if network properties change dramatically as a result. Finally, the outcomes of this model only suggest a potential explanation for the outcomes that have been empirically observed and analyzed by statistical models, and cannot independently demonstrate that empirical differences in smoking by urbanicity is truly caused by network structure.

However, these limitations are also potential strengths of the model. While the differences in smoking prevalence by urbanicity have been well documented, the overall simplicity of the model isolates one mechanism: peer influence, in the context of social network structure, and independently observes differences in the diffusion of smoking behavior by a designation (i.e., urbanicity, race, sex, and network density) that have no bearing on the diffusion mechanism. This suggests that diffusion, combined with network structure, are sufficient in tandem to generate the differences in smoking prevalence by these factors, independent and absent of all other covariates.

Finally, our results suggest that the same networks can generate dramatically different outcomes due only to the relative baseline initiation and cessation rates, and the relative peer influence on initiation and cessation. Social networks that propagate smoking behavior can also eliminate smoking behavior, or accelerate its decline. With the changing landscape of attitudes toward tobacco and the tobacco policy environment in the US, non-smokers may eventually exert greater influence on quitting than current smokers have on smoking initiation. This phenomenon could result in synergistic effects on decreasing smoking prevalence through pathways of social influence. The results of our model support prior research findings that peer influence can positively impact smoking prevalence over time.[30] Thus, network based interventions to enhance peer effects against tobacco initiation and encouraging cessation through the empowerment of non-

smokers and anti-tobacco use behavior more generally may be effective tobacco control interventions.[31,32]

Conclusions

Computational modeling as an exercise alone encourages researchers to explicitly state their assumptions and justify the inclusion of any hypothesized cause by considering it within a set of contextual processes that result in an outcome.[33] This method of inquiry differs from the process of variable selection in statistics. Here, the stratifying model results by "school network density" or "urbanicity" in a generalized linear model suggests a causal hypothesis that accounts for differences by these variables when determining the effects of peer influence on smoking behavior. Yet the quality of living in a "rural" area, for example, is not itself responsible for the mechanism that perpetuates its relationship with other risk factors and smoking behavior. Rather, it is in the experience of living in a rural area that is hypothesized to contribute to the differences observed by urbanicity in smoking behavior resulting from peer influence. In this chapter, we determined how urbanicity, race, and sex, could function through network density to change the effects of peer influence on smoking behavior. By explicitly describing (via functions) how these peer influence processes can generate the differences that we see observe in the patterning of smoking behavior by urbanicity, race, and sex, we identify a potential sufficient cause process and framework. As a result, the flexibility to define such mechanisms and processes can help us develop better causal frameworks while providing additional insights into smoking etiology.

Table 4.1 Adjusted risk ratios for smoking initiation by selected risk factors, stratified by school urbanicity designations, Add Health, 2001-2002.

		Urbanicity														
		Full Model (N=5,404)			No Urbanicity Model (N=5,404)			Urban (N=1,282)			Suburban (N=2,861)			Rural (N=1,261)		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Smoking Initiation																
Region	South (REF)															
	Midwest	1.24	0.95	1.62	1.27	0.98	1.63	1.09	0.62	1.93	1.65	1.05	2.57	0.76	0.36	1.64
	Northeast	1.05	0.79	1.41	1.07	0.80	1.44	1.08	0.72	1.62	1.12	0.76	1.63	1.00	0.66	1.52
	West	0.86	0.61	1.23	0.84	0.58	1.22	0.85	0.43	1.69	1.29	0.82	2.03	0.58	0.34	0.99
Urbanicity	Suburban (REF)															
	Rural	0.90	0.71	1.14												
	Urban	1.01	0.78	1.29												
Percentage of Nominated Friends that Smoke		**1.07†**	**1.03**	**1.13**	**1.07†**	**1.03**	**1.13**	**1.09†**	**1.02**	**1.17**	**1.09†**	**1.02**	**1.16**	1.01	0.93	1.11
School Smoking Prevalence		1.05	0.94	1.18	1.04	0.93	1.16	0.95	0.76	1.18	1.06	0.86	1.30	**1.24†**	**1.02**	**1.51**
Influence Domain (i.e., "popularity")	1st Quartile (REF)															
	2nd Quartile	1.02	0.75	1.39	1.03	0.76	1.41	1.09	0.65	1.84	0.91	0.60	1.39	**3.01†**	**1.32**	**6.89**
	3rd Quartile	0.93	0.65	1.32	0.92	0.65	1.31	1.12	0.51	2.46	0.79	0.49	1.26	1.84	0.57	5.96
	4th Quartile	0.51	0.19	1.38	0.52	0.19	1.40	0.55	0.20	1.47	0.42	0.13	1.31	1.12	0.71	1.78
Age Heterogeneity	1st Quartile (REF)															
	2nd Quartile	1.16	0.93	1.45	1.16	0.93	1.45	1.17	0.83	1.66	1.21	0.91	1.61	0.84	0.51	1.38
	3rd Quartile	1.23	0.97	1.56	1.23	0.97	1.56	0.92	0.65	1.30	**1.40†**	**1.04**	**1.87**	0.99	0.53	1.86
	4th Quartile	1.09	0.83	1.43	1.09	0.83	1.42	0.95	0.56	1.61	1.09	0.72	1.63	1.06	0.66	1.70
School Relative Density*	1st Quartile (REF)															
	2nd Quartile	1.15	0.80	1.64	1.13	0.80	1.59	1.44	0.85	2.42	1.10	0.66	1.82	---	---	---
	3rd Quartile	1.03	0.68	1.56	1.01	0.68	1.49	1.09	0.51	2.35	1.19	0.67	2.10	0.82	0.55	1.24
	4th Quartile	0.91	0.61	1.35	0.88	0.61	1.26	1.19	0.60	2.35	0.81	0.49	1.32	0.67	0.29	1.54
		Full Model (N=5,404)			No Urbanicity Model (N=5,404)			Urban (N=1,282)			Suburban (N=2,861)			Rural (N=1,261)		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Smoking Cessation																
Region	South (REF)															
	Midwest	0.75	0.53	1.05	0.78	0.54	1.13	**0.51†**	**0.29**	**0.92**	**0.51†**	**0.29**	**0.92**	0.66	0.28	1.52
	Northeast	0.70	0.48	1.01	0.77	0.53	1.12	0.55	0.24	1.25	0.65	0.38	1.09	**0.54†**	**0.31**	**0.95**
	West	1.25	0.82	1.90	1.28	0.83	1.96	0.98	0.45	2.14	1.17	0.66	2.08	1.62	0.70	3.79
Urbanicity	Suburban (REF)															
	Rural	0.76	0.51	1.12												
	Urban	1.18	0.87	1.61												
Percentage of Nominated Friends that Smoke		0.98	0.94	1.02	0.98	0.94	1.02	1.02	0.90	1.15	0.99	0.95	1.04	0.94	0.84	1.06
School Smoking Prevalence		**0.80†**	**0.69**	**0.92**	**0.78†**	**0.67**	**0.90**	**0.73†**	**0.61**	**0.88**	0.97	0.74	1.26	0.87	0.63	1.20
Influence Domain (i.e., "popularity")	1st Quartile (REF)															
	2nd Quartile	0.66	0.42	1.02	0.66	0.43	1.04	0.44	0.18	1.11	**0.53†**	**0.32**	**0.88**	1.55	0.46	5.20
	3rd Quartile	1.28	0.78	2.08	1.26	0.77	2.07	0.77	0.32	1.84	1.08	0.66	1.76	2.19	0.78	6.17
	4th Quartile	1.16	0.66	2.04	1.16	0.66	2.04	0.83	0.46	1.47	1.02	0.48	2.15	0.50	0.14	1.86
Age Heterogeneity	1st Quartile (REF)															
	2nd Quartile	0.85	0.63	1.15	0.86	0.63	1.16	0.93	0.50	1.73	0.84	0.54	1.32	0.75	0.43	1.32
	3rd Quartile	0.99	0.74	1.32	1.00	0.75	1.33	1.25	0.74	2.11	0.99	0.67	1.47	0.78	0.47	1.30
	4th Quartile	**0.66†**	**0.50**	**0.88**	**0.67†**	**0.50**	**0.88**	0.92	0.46	1.82	**0.60†**	**0.39**	**0.91**	**0.63†**	**0.41**	**0.97**
School Relative Density*	1st Quartile (REF)															
	2nd Quartile	1.12	0.72	1.74	1.07	0.69	1.67	0.99	0.52	1.88	0.85	0.42	1.72	---	---	---
	3rd Quartile	1.24	0.76	2.02	1.13	0.70	1.81	2.55	0.94	6.88	0.93	0.47	1.85	**0.56†**	**0.32**	**0.98**
	4th Quartile	1.20	0.65	2.20	1.03	0.58	1.81	0.53	0.27	1.05	1.44	0.52	4.00	**0.46†**	**0.23**	**0.94**

* There are no rural schools in the 1st quartile of relative density, therefore, the 2nd quartile is used as the referent in that stratum. There are also no schools located in the western region of the US in the 4th quartile of relative density

†p < 0.05

‡ Estimate is related to a 10% increase in the predictor

Table 4.2 Summary glossary of network measures described in Tables 4.3

Network Term	Definition
Excess Degree	The number of ties that smoking initiation or cessation influence can spread through, excluding the tie that is responsible for influencing the ego previously. More generally, measures the efficiency of a network to propagate communicable features (e.g., infections, information, health behaviors).
Clustering Coefficient	Quantifies the extent to which an adolescent's friends are to also be friends with each other
Out-Degree Centrality	The fraction of total nodes in the overall network of an adolescent's nominated friends
Shortest Paths	The shortest number of connections that it takes for a focal adolescent to reach a target adolescent
Betweenness Centrality	The number of shortest paths between two other adolescents that pass through the focal adolescent
Closeness Centrality	The sum of the length of the paths that connect an one adolescent to all other adolescents
Modularity	A measure of how connected groups of friends are from one another in the network
Average Race Heterogeneity	A measure of the extent to which the race of nominated friends differ from that of the adolescent
Eigenvector Centrality	The sum of the importance of the alters that nominate the ego as their friend.

Table 4.3 School and individual level descriptive measures, stratified by urbanicity and density tertiles.

Measure	Urban (N=20410)	Standard Deviation	Suburban (N=41018)	Standard Deviation	Rural (N=7980)	Standard Deviation
General Measures						
Number of Schools	31		63		18	
Female (%)	50.833		48.844		49.424	
White (%)	42.959		60.103		69.649	
Other (%)	26.546		18.411		10.025	
Black (%)	19.696		16.454		17.393	
Network Measures						
Average density	0.341	0.098	0.437	0.091	0.512	0.034
Average degree	7.157	5.239	8.829	5.676	10.252	5.688
Average in-degree	3.579	3.275	4.415	3.712	5.126	3.910
Average out-degree	3.579	2.933	4.415	3.052	5.126	3.015
Excess degree	9.419	1.887	10.889	1.839	12.154	0.923
Average clustering coefficient	0.169	0.207	0.180	0.196	0.204	0.188
Average out-degree centrality	0.006	0.009	0.007	0.012	0.011	0.015
Average betweeness centrality	0.004	0.008	0.005	0.007	0.007	0.012
Average closeness centrality	0.098	0.069	0.127	0.072	0.164	0.072
Average race heterogeneity	0.358	0.228	0.263	0.221	0.190	0.183
Average eigenvector centrality	0.011	0.037	0.012	0.037	0.017	0.044
Measure	Density Tertile 1 (N=21730)	Standard Deviation	Density Tertile 2 (N=24521)	Standard Deviation	Density Tertile 3 (N=23157)	Standard Deviation
General Measures						
Number of School	31		38		43	
Female (%)	47.335		51.030		49.899	
White (%)	34.634		61.180		71.041	
Other (%)	29.986		17.842		12.433	
Black (%)	25.081		15.354		12.705	
Urban (person)	11587		7217		1606	
Suburan (person)	10143		15572		15303	
Rural (person)	0		1732		6248	
Urban (school)	18		8		5	
Suburan (school)	13		27		23	
Rural (school)	0		3		15	
Network Measures						
Average degree	6.035	4.634	8.779	5.437	10.521	5.825
Average in-degree	3.017	2.840	4.390	3.565	5.260	4.042
Average out-degree	3.017	2.707	4.390	2.993	5.260	3.016
Excess degree	8.306	1.291	10.864	0.899	12.481	0.812
Average clustering coefficient	0.165	0.223	0.180	0.192	0.193	0.179
Average out-degree centrality	0.004	0.006	0.007	0.009	0.010	0.017
Average betweeness centrality	0.004	0.008	0.005	0.009	0.006	0.008
Average closeness centrality	0.074	0.060	0.129	0.066	0.161	0.068
Average race heterogeneity	0.337	0.232	0.281	0.228	0.234	0.205
Average eigenvector centrality	0.009	0.037	0.012	0.037	0.014	0.041

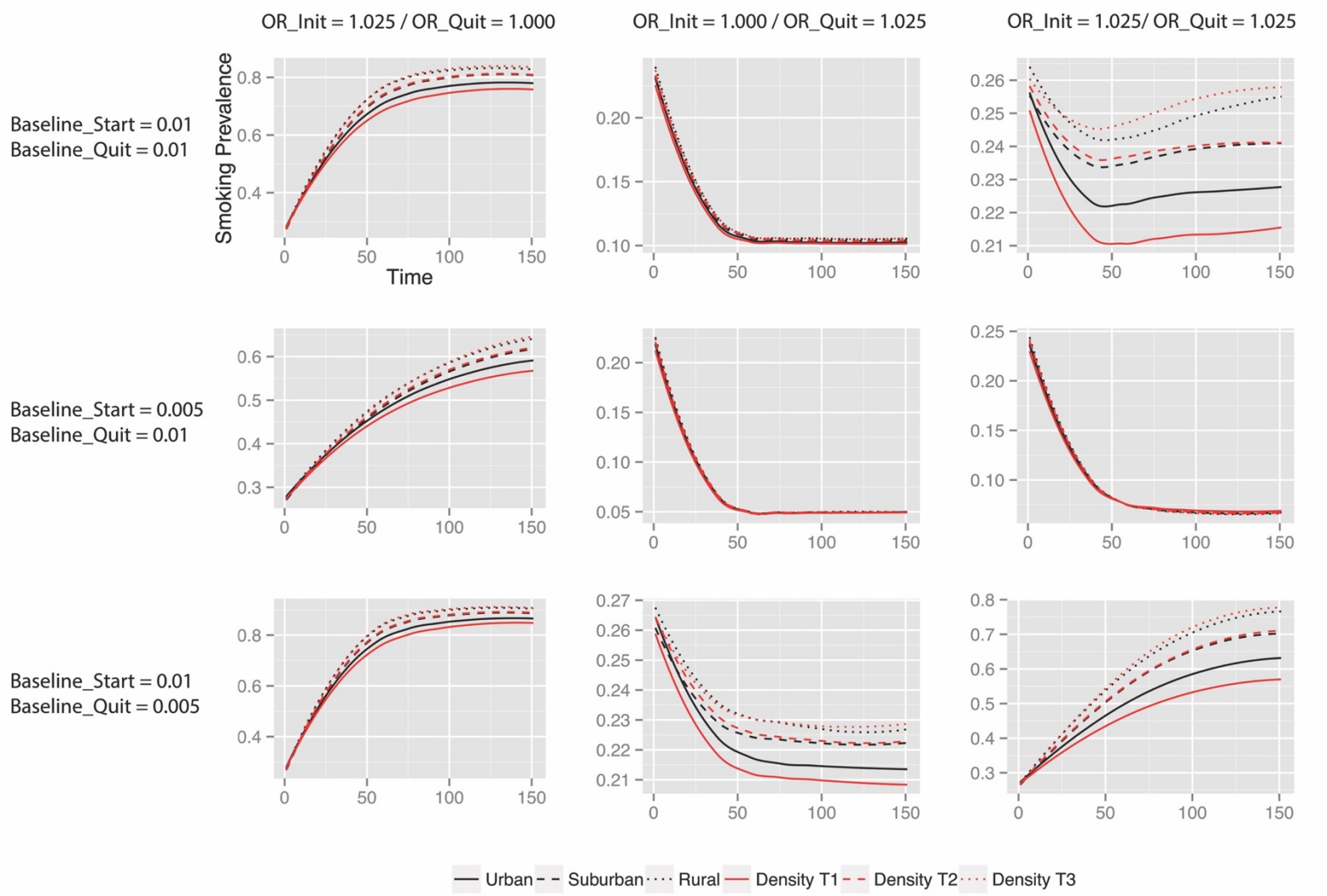


Figure 4.1 Time series plots of smoking prevalence by urbanicity designations and density terciles for all OR_init, OR_quit, Baseline_start, and Baseline_quit combinations. Each column represents a different combination of OR_init and OR_quit values, and each row represents a different Baseline_start and Baseline_quit combination.

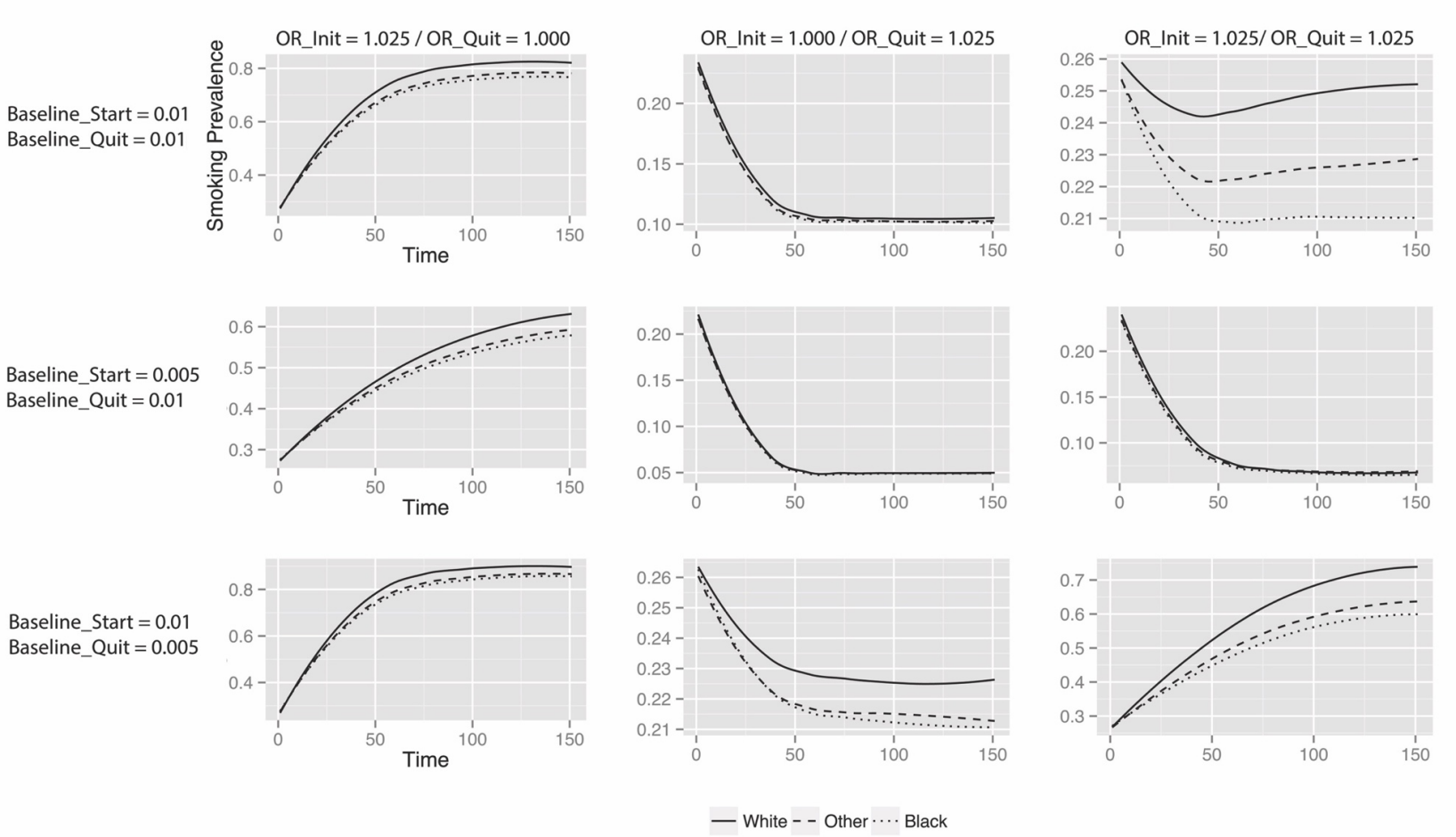


Figure 4.2 Time series plots of smoking prevalence by race for all OR_init, OR_quit, Baseline_start, and Baseline_quit combinations. Each column represents a different combination of OR_init and OR_quit values, and each row represents a different Baseline_start and Baseline_quit combination.

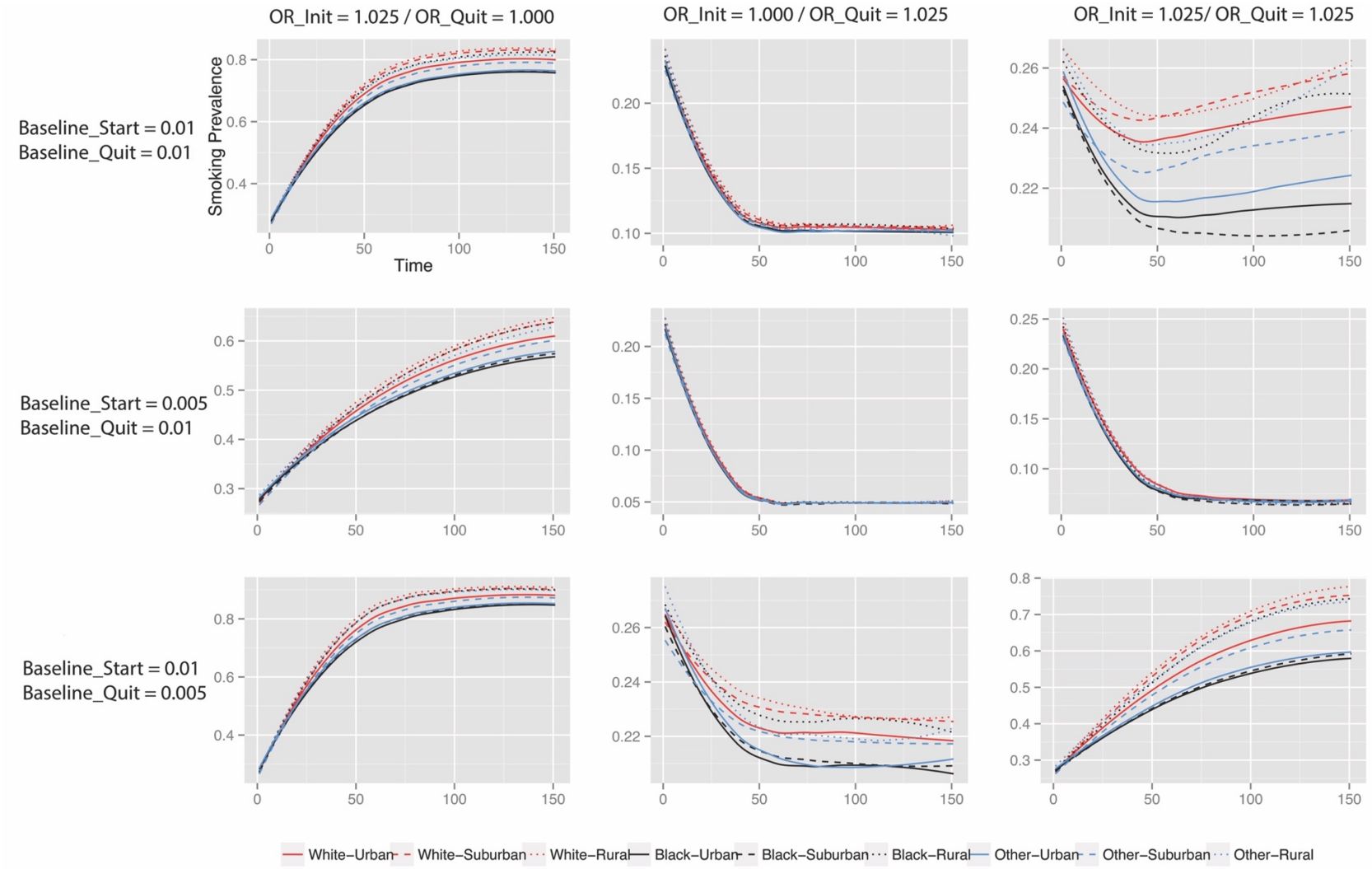


Figure 4.3 Time series plots of smoking prevalence by race, stratified by urbanicity, for all OR_init, OR_quit, Baseline_start, and Baseline_quit combinations. Each column represents a different combination of OR_init and OR_quit values, and each row represents a different Baseline_start and Baseline_quit combination.

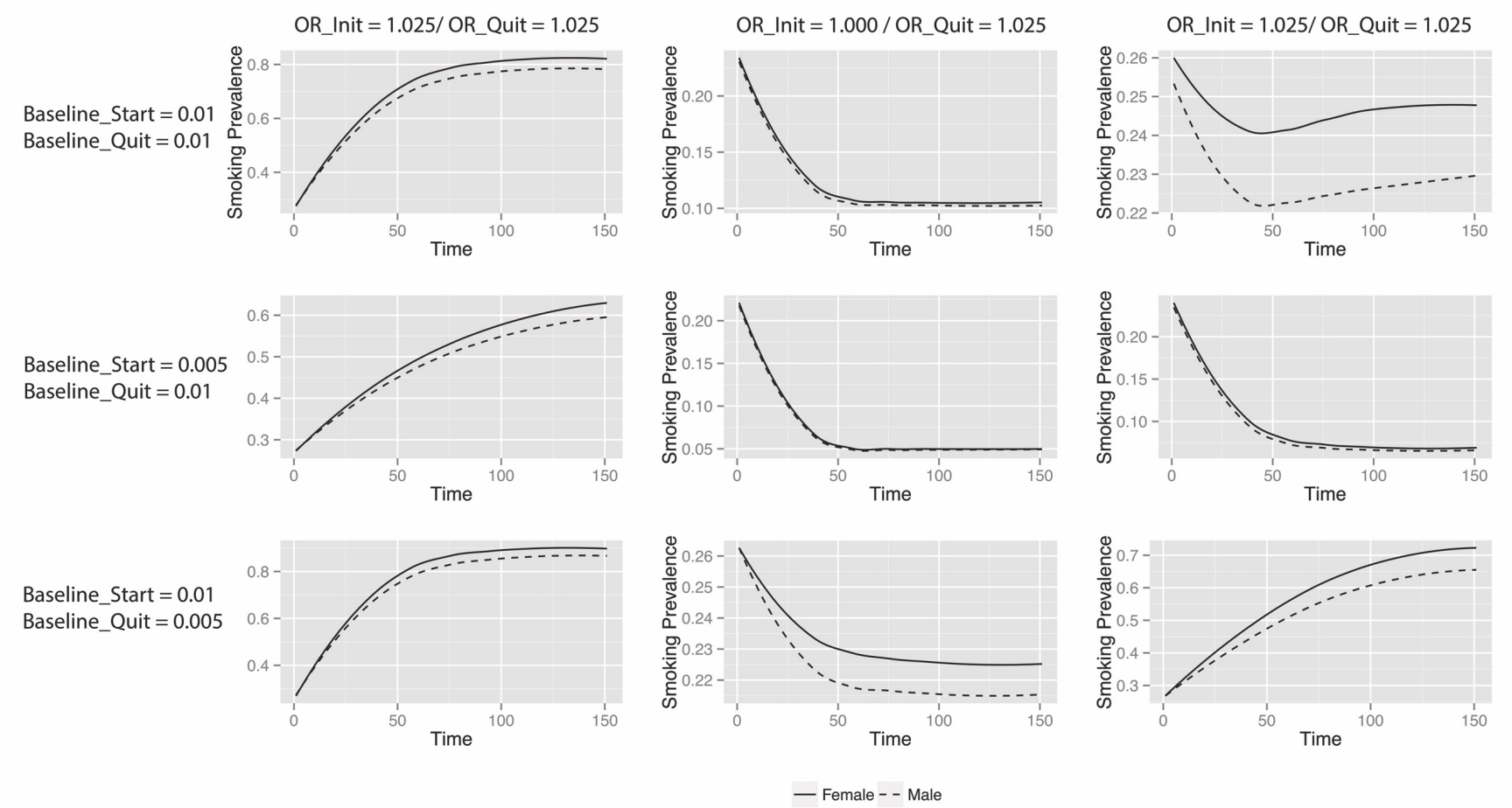


Figure 4.4 Time series plots of smoking prevalence by sex for all OR_init, OR_quit, Baseline_start, and Baseline_quit combinations. Each column represents a different combination of OR_init and OR_quit values, and each row represents a different Baseline_start and Baseline_quit combination.

References

1. Schaefer DR, Adams J, Haas SA. Social Networks and Smoking: Exploring the Effects of Peer Influence and Smoker Popularity through Simulations. *Health Educ Behav*. 2013;40(1 Suppl):24S-32S. doi:10.1177/1090198113493091.
2. Eckles D, Kizilcec RF, Bakshy E. Estimating Peer Effects in Networks with Peer Encouragement Designs. *Proc Natl Acad Sci*. 2016;113(27):7316-7322. doi:10.1073/pnas.1511201113.
3. Haynie DL. Delinquent Peers Revisited: Does Network Structure Matter? *Am J Sociol*. 2001;106(4):1013-1057. doi:10.1086/320298.
4. Kobus K. Peers and Adolescent Smoking. *Addiction*. 2003;98(Suppl 1):37-55.
5. Santor D, Messervey D, Kusumakar V. Measuring Peer Pressure, Popularity, and Conformity in Adolescent Boys and Girls: Predicting School Performance, Sexual Attitudes, and Substance Abuse. *Youth Adolesc*. 2000;29(2):163-182. doi:10.1023/a:1005152515264.
6. McGloin JM, Sullivan CJ, Thomas KJ. Peer Influence and Context : The Interdependence of Friendship Groups, Schoolmates and Network Density in Predicting Substance Use. *J Youth Adolesc*. 2014;43:1436-1452. doi:10.1007/s10964-014-0126-7.
7. Ennett ST, Bauman KE, Hussong A, et al. The Peer Context of Adolescent Substance Use: Findings from Social Network Analysis. *J Res Adolesc*. 2006;16(2):159-186. doi:10.1111/j.1532-7795.2006.00127.x.
8. Mercken L, Snijders TAB, Steglich C, Vartiainen E, de Vries H. Dynamics of Adolescent Friendship Networks and Smoking Behavior. *Soc Networks*. 2010;32(1):72-81. doi:10.1016/j.socnet.2009.02.005.
9. Thornberry TP. Toward an Interactional Theory of Delinquency. *Criminology*. 1987;25(4):863-892. doi:10.1111/j.1745-9125.1987.tb00823.x.
10. Kiuru N, Burk WJ, Laursen B, Salmela-Aro K, Nurmi JE. Pressure to Drink but Not to Smoke: Disentangling Selection and Socialization in Adolescent Peer Networks and Peer Groups. *J Adolesc*. 2010;33(6):801-812. doi:10.1016/j.adolescence.2010.07.006.
11. Ennett ST, Bauman KE. The contribution of influence and selection to adolescent peer group homogeneity: The case of adolescent cigarette smoking. *J Pers Soc Psychol*. 1994;67(4):653-663. doi:10.1037/0022-3514.67.4.653.
12. Lakon CM, Hipp JR, Timberlake DS. The Social Context of Adolescent Smoking : A Systems Perspective. *Am J Public Health*. 2010;100(7). doi:10.2105/AJPH.2009.167973.
13. Alexander C, Piazza M, Mekos D, Valente T. Peers, Schools, and Adolescent Cigarette Smoking. *J Adolesc Heal*. 2001;29(1):22-30. doi:10.1016/S1054-139X(01)00210-5.

14. Manski CF. Identification of Endogenous Social Effects: The Reflection Problem. *Rev Econ Stud*. 1993;60(3):531-542. doi:10.2307/2298123.

15. Bramoullé Y, Djebbari H, Fortin B. Identification of Peer Effects through Social Networks. *J Econom*. 2009;150(1):41-55. doi:10.1016/j.jeconom.2008.12.021.

16. Jeon KC, Goodson P. US Adolescents' Friendship Networks and Health Risk Hehaviors: A Systematic Review of Studies Using Social Network Analysis and Add Health Data. *PeerJ*. 2015;3:e1052. doi:10.7717/peerj.1052.

17. Peguero A a., Portillos EL, Hong JS, González JC, Kahle LL, Shekarkhar Z. Victimization, Urbanicity, and the Relevance of Context: School Routines, Race and Ethnicity, and Adolescent Violence. *J Criminol*. 2013;2013:1-14. doi:10.1155/2013/240637.

18. Eberhardt MS , Ingram DD, Makuc DM et al. *Urban and Rural Health Chartbook, Health United States*. 2001.

19. Doescher MP, Jackson JE, Jerant A, Gary Hart L. Prevalence and Trends in Smoking: A National Rural Study. *J Rural Heal*. 2006;22(2):112-118. doi:10.1111/j.1748-0361.2006.00018.x.

20. Fletcher JM, Ross SL. Estimating the Effects of Friendship Networks on Health Behaviors of Adolescents. *Natl Bur Econ Res Work Pap Ser*. 2012;No. 18253.

21. Herrera-Yagüe C, Schneider CMM, Couronné T, et al. The Anatomy of Urban Social Networks and its Implications in the Searchability Problem. *Sci Rep*. 2015:1-27. doi:10.1038/srep10265.

22. Leyden KM. Social Capital and the Built Environment: The Importance of Walkable Neighborhoods. *Am J Public Health*. 2003;93(9):1546-1551. doi:10.2105/AJPH.93.9.1546.

23. Harris KM. The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009. *Chapel Hill, NC Carolina Popul Center, Univ North Carolina Chapel Hill*. 2009. doi:DOI: 10.3886/ICPSR27021.v9.

24. Singh T, Arrazola RA, Corey CG, et al. Tobacco Use Among Middle and High School Students - United States, 2011-2015. *MMWR Morb Mortal Wkly Rep*. 2016;65(14):361-367. doi:10.15585/mmwr.mm6514a1.

25. Byrnes, JP, Miller, DC, Schafer WD. Gender Differences in Risk Taking: A Meta-Analysis. *Psychol Bull*. 1999;125(3):367.

26. Miething A, Rostila M, Edling C, Rydgren J. The Influence of Social Network Characteristics on Peer Clustering in Smoking: A Two-wave Panel Study of 19-and 23-Year-old Swedes. *PLoS One*. 2016;11(10):1-17. doi:10.1371/journal.pone.0164611.

27. O'Loughlin JL, Sylvestre MP, Dugas EN, Karp I. Predictors of the Occurrence of Smoking Discontinuation in Novice Adolescent Smokers. *Cancer Epidemiol Biomarkers Prev*. 2014;23(6):1090-1101. doi:10.1158/1055-9965.EPI-13-0869.

28. Grogan S, Conner M, Fry G, Gough B, Higgins A. Gender differences in Smoking: A Longitudinal Study of Beliefs Predicting Smoking in 11-15 Year Olds. *Psychol Heal*. 2009;24(3):301-316. doi:10.1080/08870440701746586.

29. Mercken L, Snijders TAB, Steglich C, Vertiainen E, De Vries H. Smoking-based Selection and Influence in Gender-segregated Friendship Networks: A Social Network Analysis of Adolescent Smoking. *Addiction*. 2010;105(7):1280-1289. doi:10.1111/j.1360-0443.2010.02930.x.

30. Hoffman BR, Sussman S, Unger JB, Valente TW. Peer Influences on Adolescent Cigarette Smoking: a Theoretical Review of the Literature. *Subst Use Misuse*. 2006;41(1):103-155. doi:10.1080/10826080500368892.

31. Valente TW, Hoffman BR, Ritt-Olson A, Lichtman K, Johnson CA. Effects of a Social-Network Method for Group Assignment Strategies on Peer-Led Tobacco Prevention Programs in Schools. *Am J Public Health*. 2003;93(11):1837-1843. doi:10.2105/AJPH.93.11.1837.

32. Valente TW. Network Interventions. *Science (80)*. 2012;337(6090):49-53. doi:10.1126/science.1217330.

33. Epstein J. Why model? *J Artif Soc Soc Simul*. 2008;11(4):6. doi:10.1080/01969720490426803.

Appendix C: Supplemental Material for Chapter 4

Model Mechanisms and Pseudo Code

C 1.1 Transition Probabilities

Transition probability calculations are taken from the functional form of logistic regression to better connect an understanding of the underlying theory of statistical methods with mechanistic computational methods. That is, the percentage of friends that smoke should behave as a weight on the odds of transition at baseline, which we assume in this model to control for all other factors. We then convert these odds into a probability to control for the likelihood of transition between states with parameters.

$$b = \frac{1}{1 + e^{-baseline}}$$

Where *b* is the baseline transition probability assuming no friendship influence, and *baseline* is the *baseline* odds of transitioning from one smoking state to another.

$$e^{-baseline} = \frac{(1-b)}{b}$$

$$baseline = -\log \frac{1-b}{b}$$

$$p(transition)_t = \frac{1}{1 + e^{-baseline-\beta * pofs_{t-1}}}$$

Where $pofs_{t-1}$ is the percentage of friends at the previous time step that generate the influence to change states. These are the friends that have different smoking states than the smoking state of the focal adolescent and emerge from the network connections of the focal adolescents in the mode

$$\beta = \log(OR)$$

$$OR = \frac{\left(\frac{p(transition)}{1 - p(transition)}\right)}{\left(\frac{bss}{1 - bss}\right)}$$

Our variable model parameters include β and *baseline*, which are referred to throughout the text as OR_init and OR_quit, along with Baseline_start and Baseline_quit, respectively.

C 1.2 Model Pseudo Code

```
OBJECT agent;
  proportion of friends that smoke;
  proportion of friends that are non-smokers;
  school id;
  personal id;

PROGRAM network_smoking_model;
  initialize agents according to data from 1994 Add Health Network;
  initialize smokers probabilistically (uniform) by weighted population smoking
  prevalence 1994 Add Health Data (29%);
     for every time step from 1 to 150:
         repeat for all agents in the model:
            if non-smoker:
               start at P(start_smoking)
             else:
               quit at P(quit_smoking)
             calculate the percentage of friends that are smokers for next time step;
             calculate the percentage of friends that are non-smokers for next time
             step;
         calculate aggregates;
         write model and agent statistics to output;
         clear model and agent statistics;
```

Table C1 Unadjusted risk ratios for smoking initiation and cessation by selected risk factors, stratified by school urbanicity designations, Add Health, 2001-2002.

		Unadjusted Non-Stratified			Urbanicity: Urban			Urbanicity: Suburban			Urbanicity: Rural		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Smoking Initiation													
Region													
	South (REF)												
	Midwest	**1.46†**	**1.11**	**1.93**	1.01	0.66	1.56	**1.75†**	**1.16**	**2.63**	1.23	0.76	1.98
	Northeast	1.26	0.90	1.76	**1.51†**	**1.03**	**2.21**	1.14	0.73	1.76	1.05	0.85	1.31
	West	0.90	0.59	1.35	0.97	0.52	1.84	1.31	0.87	1.99	**0.50†**	**0.29**	**0.87**
Urbanicity													
	Suburban (REF)												
	Rural*	0.88	0.63	1.24	---	---	---	---	---	---	---	---	---
	Urban	0.97	0.69	1.36	---	---	---	---	---	---	---	---	---
Percentage of Nominated Friends that Smoke		**1.01†**	**1.00**	**1.01**	**1.01†**	**1.00**	**1.02**	**1.01†**	**1.00**	**1.02**	1.00	0.99	1.01
Percentage of Nominated Friends that Smoke‡		1.00	0.99	1.02	0.98	0.97	1.00	1.00	0.98	1.03	**1.02†**	**1.01**	**1.04**
School Smoking Prevalence‡	1st Quartile (REF)												
	2nd Quartile	1.05	0.81	1.35	1.00	0.64	1.56	0.96	0.68	1.34	**1.65†**	**1.02**	**2.68**
	3rd Quartile	0.89	0.61	1.28	0.87	0.36	2.10	0.84	0.55	1.28	0.98	0.61	1.59
	4th Quartile	0.47	0.22	1.00	**0.32†**	**0.16**	**0.65**	0.47	0.18	1.23	0.73	0.52	1.03
Age Heterogeneity	1st Quartile (REF)												
	2nd Quartile	1.17	0.93	1.46	1.17	0.83	1.67	1.23	0.91	1.66	0.86	0.51	1.44
	3rd Quartile	1.23	0.96	1.59	0.95	0.68	1.33	**1.41†**	**1.02**	**1.96**	0.98	0.54	1.79
	4th Quartile	1.02	0.78	1.33	0.90	0.51	1.58	1.04	0.70	1.55	0.96	0.61	1.51
School Relative Density*	1st Quartile (REF)												
	2nd Quartile	1.32	0.85	2.07	**1.70†**	**1.12**	**2.57**	1.26	0.61	2.61	---	---	---
	3rd Quartile	1.34	0.94	1.93	1.28	0.82	1.99	1.41	0.75	2.63	1.19	0.50	2.81
	4th Quartile	1.43	0.94	2.16	1.48	1.00	2.21	1.47	0.73	2.93	1.33	0.59	2.99
Smoking Cessation													
Region													
	South (REF)												
	Midwest	0.94	0.56	1.60	0.99	0.47	2.12	0.81	0.38	1.76	1.01	0.65	1.58
	Northeast	0.99	0.52	1.89	1.27	0.54	2.96	0.80	0.50	1.27	0.57	0.23	1.45
	West	1.12	0.76	1.67	1.31	0.63	2.73	0.87	0.49	1.54	1.32	0.97	1.82
Urbanicity													
	Suburban (REF)												
	Rural*	0.69	0.47	1.02	---	---	---	---	---	---	---	---	---
	Urban	1.31	0.82	2.10	---	---	---	---	---	---	---	---	---
Percentage of Nominated Friends that Smoke‡		1.00	0.99	1.00	1.00	0.99	1.01	1.00	0.99	1.00	0.99	0.98	1.00
School Smoking Prevalence‡		**0.97†**	**0.96**	**0.98**	**0.96†**	**0.94**	**0.98**	0.99	0.96	1.01	**0.97†**	**0.94**	**1.00**
Influence Domain (i.e., "popularity")	1st Quartile (REF)												
	2nd Quartile	0.73	0.47	1.15	0.64	0.21	1.96	0.61	0.35	1.05	1.78	0.59	5.33
	3rd Quartile	1.10	0.74	1.62	0.53	0.22	1.29	1.04	0.64	1.71	**1.68†**	**1.16**	**2.43**
	4th Quartile	1.07	0.67	1.72	**0.65†**	**0.49**	**0.85**	1.01	0.55	1.86	**2.54†**	**1.38**	**4.68**
Age Heterogeneity	1st Quartile (REF)												
	2nd Quartile	0.83	0.63	1.11	0.88	0.50	1.56	0.85	0.56	1.30	0.72	0.48	1.09
	3rd Quartile	0.96	0.72	1.27	1.20	0.69	2.07	0.95	0.66	1.38	0.72	0.37	1.41
	4th Quartile	**0.62†**	**0.48**	**0.81**	0.90	0.44	1.81	**0.56†**	**0.39**	**0.80**	**0.62†**	**0.40**	**0.97**
School Relative Density*	1st Quartile (REF)												
	2nd Quartile	0.93	0.51	1.68	1.72	0.80	3.69	0.67	0.30	1.48	---	---	---
	3rd Quartile	0.77	0.47	1.25	1.21	0.63	2.33	0.71	0.35	1.44	0.70	0.27	1.86
	4th Quartile	0.76	0.38	1.50	0.75	0.31	1.82	1.02	0.39	2.64	0.57	0.22	1.53

* There were no rural schools in the 1st quartile of relative density, therefore, the 2nd quartile is used as the referent in that stratum. There are also no schools located in the western region of the US in the 4th quartile of relative density

† p < 0.05

‡ Estimate is related to a 10% increase in the predictor

Table C2 Full model results of adjusted risk ratios for smoking initiation and cessation by selected risk factors, stratified by school network density, Add Health, 2001-2002.

		Urbanicity														
		Full Model (N=5,404)			No Urbanicity Model (N=5,404)			Urban (N=1,282)			Suburban (N=2,861)			Rural (N=1,261)		
		Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Smoking Initiation																
Intercept		0.16	0.09	0.30	0.16	0.09	0.29	0.24	0.12	0.48	0.11	0.05	0.27	0.17	0.08	0.36
Age		0.86	0.79	0.93	0.86	0.79	0.93	0.91	0.80	1.04	0.83	0.74	0.94	0.85	0.73	0.99
Sex	Male (REF)															
	Female	0.71	0.61	0.81	0.71	0.61	0.81	0.58	0.44	0.75	0.73	0.61	0.88	0.73	0.55	0.97
Race	White (REF)															
	Black	0.52	0.35	0.78	0.52	0.35	0.78	0.35	0.16	0.75	0.71	0.48	1.03	0.37	0.14	0.98
	Other	0.92	0.75	1.15	0.92	0.75	1.15	0.95	0.72	1.25	0.91	0.65	1.27	0.83	0.45	1.53
Depression	Never (REF)															
	Everyday	1.13	0.78	1.65	1.13	0.78	1.64	2.32	1.38	3.89	0.77	0.43	1.37	1.11	0.53	2.32
	Often	1.16	0.88	1.52	1.16	0.88	1.52	1.88	1.06	3.34	1.16	0.82	1.63	0.57	0.26	1.25
	Occasionally	1.10	0.87	1.39	1.10	0.87	1.39	1.08	0.76	1.55	1.09	0.79	1.50	1.26	0.79	2.02
	Rarely	1.02	0.87	1.19	1.02	0.87	1.19	1.16	0.90	1.50	0.97	0.81	1.18	1.07	0.67	1.72
Availability of Cigarettes at Home	No (REF)															
	Yes	1.19	1.00	1.43	1.19	1.00	1.43	1.32	0.96	1.80	1.19	0.93	1.52	1.01	0.72	1.42
School Size	Small (REF)															
	Medium	1.11	0.80	1.55	1.12	0.81	1.55	0.86	0.53	1.41	1.28	0.82	2.00	0.66	0.30	1.45
	Large	1.03	0.58	1.83	1.05	0.59	1.86	0.62	0.28	1.37	1.67	0.77	3.63	0.53	0.23	1.22
Region	South (REF)															
	Midwest	1.24	0.95	1.62	1.27	0.98	1.63	1.09	0.62	1.93	1.65	1.05	2.57	0.76	0.36	1.64
	Northeast	1.05	0.79	1.41	1.07	0.80	1.44	1.08	0.72	1.62	1.12	0.76	1.63	1.00	0.66	1.52
	West	0.86	0.61	1.23	0.84	0.58	1.22	0.85	0.43	1.69	1.29	0.82	2.03	0.58	0.34	0.99
Urbanicity	Suburban (REF)															
	Rural	0.90	0.71	1.14												
	Urban	1.01	0.78	1.29												
Percentage of Nominated Friends that Smoke		1.07	1.03	1.13	1.07	1.03	1.13	1.09	1.02	1.17	1.09	1.02	1.16	1.01	0.93	1.11
School Smoking Prevalence		1.05	0.94	1.18	1.04	0.93	1.16	0.95	0.76	1.18	1.06	0.86	1.30	1.24	1.02	1.51
Influence Domain (i.e., "popularity")	1st Quartile (REF)															
	2nd Quartile	1.02	0.75	1.39	1.03	0.76	1.41	1.09	0.65	1.84	0.91	0.60	1.39	3.01	1.32	6.89
	3rd Quartile	0.93	0.65	1.32	0.92	0.65	1.31	1.12	0.51	2.46	0.79	0.49	1.26	1.84	0.57	5.96
	4th Quartile	0.51	0.19	1.38	0.52	0.19	1.40	0.55	0.20	1.47	0.42	0.13	1.31	1.12	0.71	1.78
Age Heterogeneity	1st Quartile (REF)															
	2nd Quartile	1.16	0.93	1.45	1.16	0.93	1.45	1.17	0.83	1.66	1.21	0.91	1.61	0.84	0.51	1.38
	3rd Quartile	1.23	0.97	1.56	1.23	0.97	1.56	0.92	0.65	1.30	1.40	1.04	1.87	0.99	0.53	1.86
	4th Quartile	1.09	0.83	1.43	1.09	0.83	1.42	0.95	0.56	1.61	1.09	0.72	1.63	1.06	0.66	1.70
School Relative Density*	1st Quartile (REF)															
	2nd Quartile	1.15	0.80	1.64	1.13	0.80	1.59	1.44	0.85	2.42	1.10	0.66	1.82	---	---	---
	3rd Quartile	1.03	0.68	1.56	1.01	0.68	1.49	1.09	0.51	2.35	1.19	0.67	2.10	0.82	0.55	1.24
	4th Quartile	0.91	0.61	1.35	0.88	0.61	1.26	1.19	0.60	2.35	0.81	0.49	1.32	0.67	0.29	1.54
Smoking Cessation																
Intercept		1.12	0.54	2.30	1.24	0.63	2.41	2.29	0.66	7.91	0.88	0.34	2.27	1.42	0.35	5.72
Age		0.99	0.90	1.09	0.99	0.90	1.08	0.95	0.80	1.14	0.94	0.81	1.09	1.06	0.88	1.28
Sex	Male (REF)															
	Female	0.99	0.80	1.23	0.99	0.80	1.23	0.87	0.63	1.18	0.98	0.72	1.33	1.32	0.80	2.19
Race	White (REF)															
	Black	1.23	0.83	1.81	1.21	0.82	1.79	0.92	0.54	1.57	1.40	0.92	2.13	1.19	0.41	3.49
	Other	0.79	0.57	1.09	0.78	0.57	1.08	0.79	0.44	1.43	0.81	0.52	1.29	0.71	0.30	1.67
Depression	Never (REF)															
	Everyday	0.64	0.41	1.00	0.64	0.41	1.00	1.05	0.40	2.72	0.54	0.29	0.99	0.78	0.40	1.55
	Often	1.00	0.72	1.38	0.99	0.72	1.37	1.32	0.75	2.33	0.83	0.52	1.34	1.22	0.73	2.03
	Occasionally	1.00	0.79	1.28	1.00	0.78	1.27	0.66	0.34	1.27	0.86	0.61	1.21	1.74	1.22	2.48
	Rarely	0.85	0.65	1.10	0.84	0.65	1.10	1.28	0.73	2.27	0.80	0.58	1.11	0.68	0.35	1.31
Availability of Cigarettes at Home	Cigs No (REF)															
	Yes	0.76	0.63	0.93	0.76	0.63	0.93	0.74	0.50	1.10	0.84	0.66	1.06	0.63	0.40	1.01
School Size	Small (REF)															
	Medium	1.06	0.69	1.64	1.05	0.68	1.64	1.09	0.64	1.83	1.06	0.53	2.12	0.74	0.24	2.30
	Large	0.84	0.46	1.51	0.88	0.48	1.61	0.67	0.31	1.42	0.94	0.38	2.37	1.00	0.33	3.01
Region	South (REF)															
	Midwest	0.75	0.53	1.05	0.78	0.54	1.13	0.51	0.29	0.92	0.51	0.29	0.92	0.66	0.28	1.52
	Northeast	0.70	0.48	1.01	0.77	0.53	1.12	0.55	0.24	1.25	0.65	0.38	1.09	0.54	0.31	0.95
	West	1.25	0.82	1.90	1.28	0.83	1.96	0.98	0.45	2.14	1.17	0.66	2.08	1.62	0.70	3.79
Urbanicity	Suburban (REF)															
	Rural	0.76	0.51	1.12												
	Urban	1.18	0.87	1.61												
Percentage of Nominated Friends that Smoke		0.98	0.94	1.02	0.98	0.94	1.02	1.02	0.90	1.15	0.99	0.95	1.04	0.94	0.84	1.06
School Smoking Prevalence		0.80	0.69	0.92	0.78	0.67	0.90	0.73	0.61	0.88	0.97	0.74	1.26	0.87	0.63	1.20
Influence Domain (i.e., "popularity")	1st Quartile (REF)															
	2nd Quartile	0.66	0.42	1.02	0.66	0.43	1.04	0.44	0.18	1.11	0.53	0.32	0.88	1.55	0.46	5.20
	3rd Quartile	1.28	0.78	2.08	1.26	0.77	2.07	0.77	0.32	1.84	1.08	0.66	1.76	2.19	0.78	6.17
	4th Quartile	1.16	0.66	2.04	1.16	0.66	2.04	0.83	0.46	1.47	1.02	0.48	2.15	0.50	0.14	1.86
Age Heterogeneity	1st Quartile (REF)															
	2nd Quartile	0.85	0.63	1.15	0.86	0.63	1.16	0.93	0.50	1.73	0.84	0.54	1.32	0.75	0.43	1.32
	3rd Quartile	0.99	0.74	1.32	1.00	0.75	1.33	1.25	0.74	2.11	0.99	0.67	1.47	0.78	0.47	1.30
	4th Quartile	0.66	0.50	0.88	0.67	0.50	0.88	0.92	0.46	1.82	0.60	0.39	0.91	0.63	0.41	0.97
School Relative Density*	1st Quartile (REF)															
	2nd Quartile	1.12	0.72	1.74	1.07	0.69	1.67	0.99	0.52	1.88	0.85	0.42	1.72	---	---	---
	3rd Quartile	1.24	0.76	2.02	1.13	0.70	1.81	2.55	0.94	6.88	0.93	0.47	1.85	0.56	0.32	0.98
	4th Quartile	1.20	0.65	2.20	1.03	0.58	1.81	0.53	0.27	1.05	1.44	0.52	4.00	0.46	0.23	0.94

* There are no rural schools in the 1st quartile of relative density, therefore, the 2nd quartile is used as the referent in that stratum. There are also no schools located in the western region of the US in the 4th quartile of relative density

† p < 0.05

‡ Estimate is related to a 10% increase in the predictor

Table C3 Individual level descriptive statistics of network measures, stratified by sex.

Measure	**Female (N=34354)**	**Standard Deviation**	**Male (N=34554)**	**Standard Deviation**
Network Measures				
Average degree	9.035	5.512	8.004	5.713
Average in-degree	4.461	3.570	4.057	3.713
Average out-degree	4.574	2.917	3.948	3.147
Excess degree	10.677	1.878	10.532	2.054
Average clustering coefficient	0.193	0.197	0.168	0.199
Average out-degree centrality	0.008	0.012	0.007	0.012
Average betweeness centrality	0.005	0.009	0.004	0.008
Average closeness centrality	0.131	0.068	0.115	0.078
Average race heterogeneity	0.283	0.223	0.280	0.228

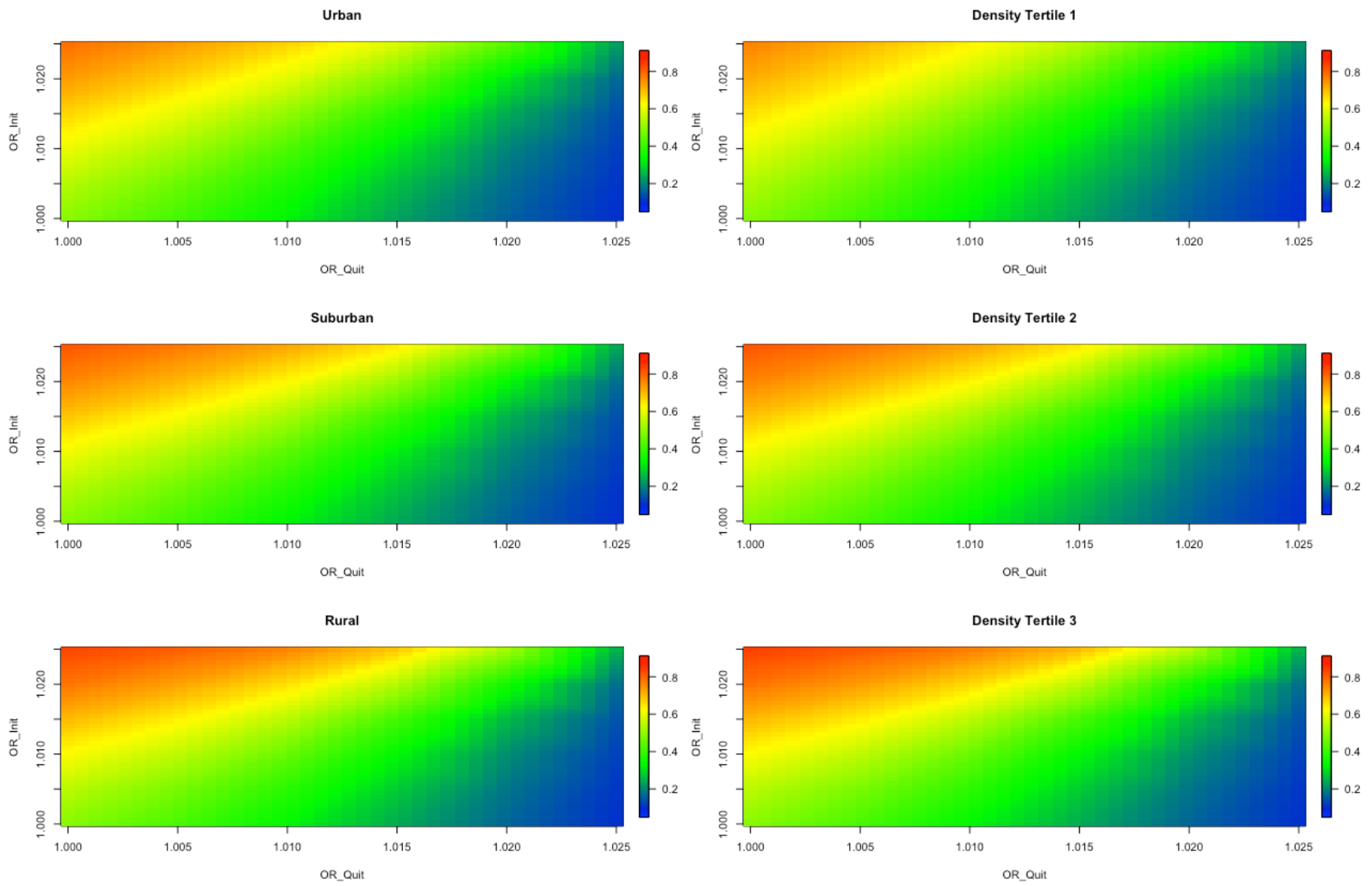


Figure C1 Model results comparing smoking prevalence values stratified by urbanicity and density tertiles, averaged over last 5 model time steps, where Baseline_start = 0.01 and Baseline_quit = 0.01.

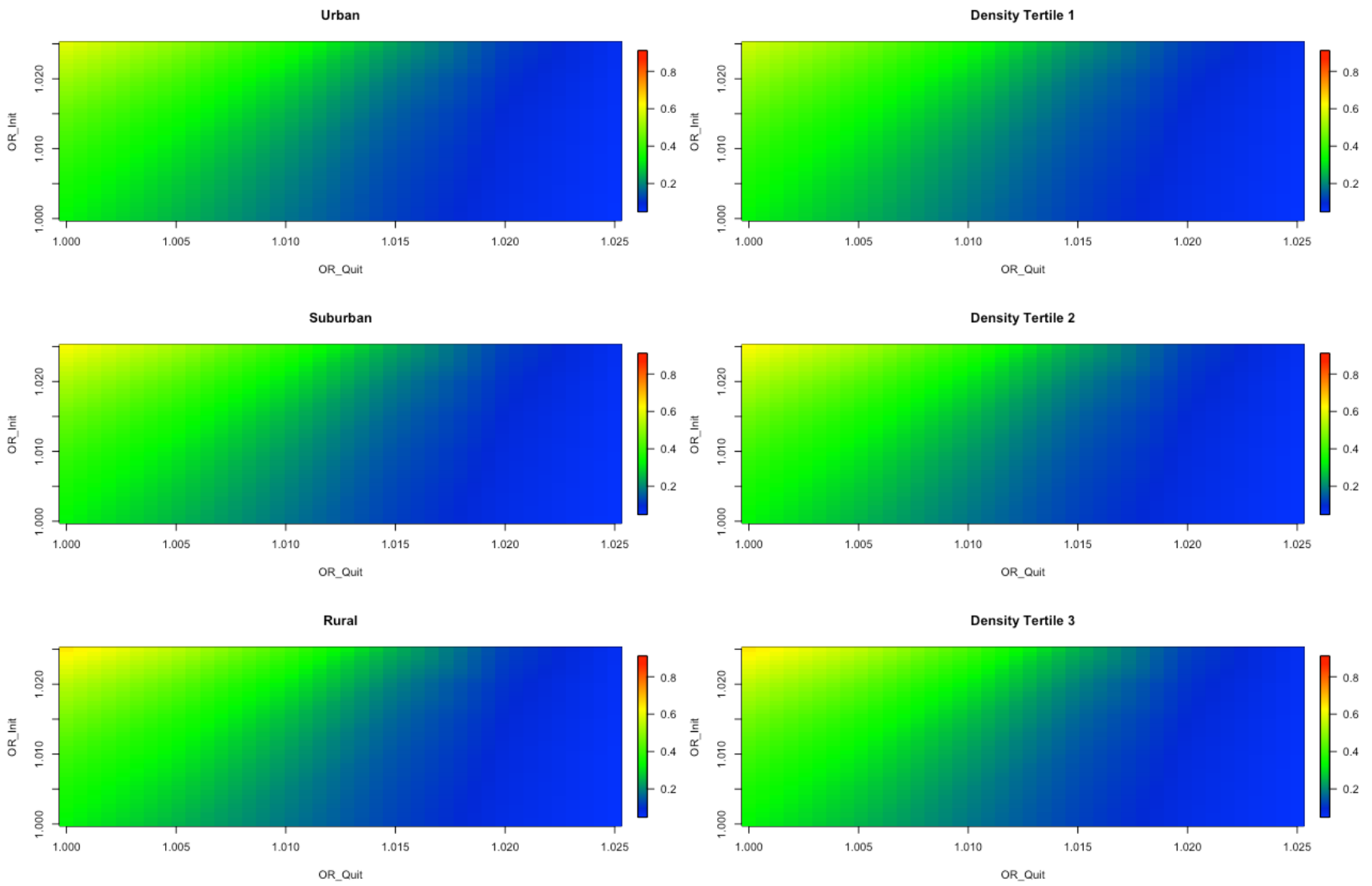


Figure C2 Model results comparing smoking prevalence values stratified by urbanicity and density tertiles, averaged over last 5 model time steps, where Baseline_start = 0.005 and Baseline_quit = 0.01.

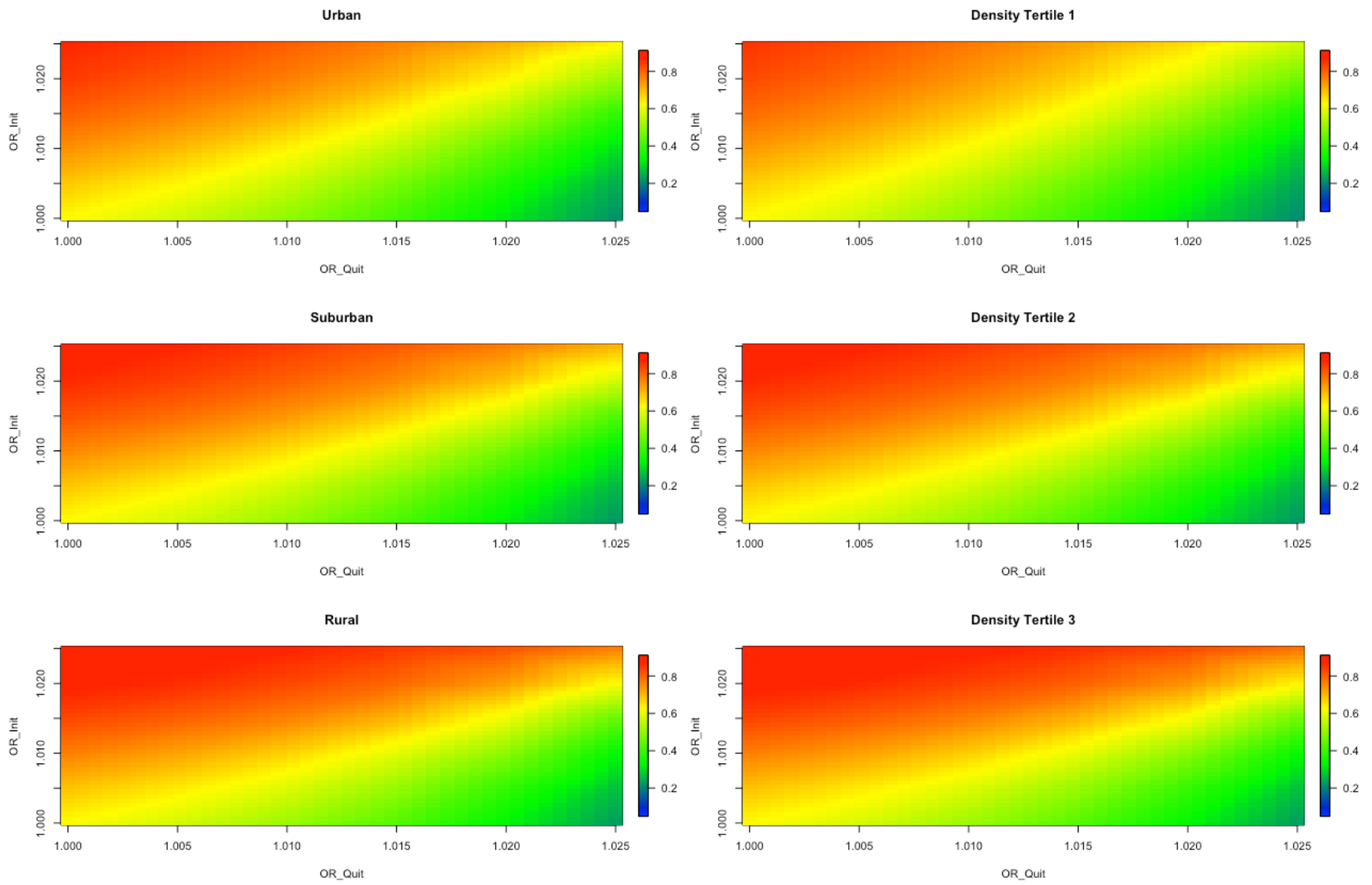


Figure C3 Model results comparing smoking prevalence values stratified by urbanicity and density tertiles, averaged over last 5 model time steps, where Baseline_start = 0.01 and Baseline_quit = 0.005.

CHAPTER 5

Conclusions: Advancing Causal Frameworks in Epidemiology

Background

What constitutes the explanation of a social phenomenon? Epstein suggests that the answer to the question of "can you explain it?" can be better interpreted as "can you grow it?"[1] That is, growing a phenomenon necessarily means that we can provide a plausible explanation of its potential causes. This suggestion appropriately integrates with Rothman's sufficient-component cause model which proposes that there can exist some set of component causes, that when occurring together, are sufficient to result in disease.[2] While the latter of these two examinations provides a theoretical basis for causality, the former proposes a practical application of sufficient-component cause theory, with important implications for causal thinking in epidemiology. In the following sections, I conclude this dissertation by discussing 1) some existing conceptualizations of causal pathways in traditional epidemiology; 2) how combining statistical and computational models can provide better insights into causality frameworks of smoking behavior than either method alone; 3) how a better understanding of these insights, even in abstract forms such as those presented in previous chapters, can provide insights into potential causal mechanisms; and 4) directions for future work on the topics of this dissertation.

Discussion of Causality

Causality in History, Science, and Epidemiology

The field of epidemiology as a scientific discipline with roots in biology, logic, and the philosophy of science, for which applied statistics is only a tool.[3] Concepts such as causality and counterfactuals, which are pillars of epidemiologic theory, logic, and complex systems, also exist as fundamental theories of knowledge throughout the history of science and philosophy. A commonly cited philosopher in the epidemiologic (and epistemological) methods literature is

David Hume. In his endeavor to define knowledge, Hume provides the first explicit conceptualization of counterfactual theory when discussing causality in the Enquiry Concerning Human Understanding, stating that "if the first object had not been, the second never had existed."[4] Hill references this concept of temporality in his set of criteria in determining causality,[5] with Popper later refining this understanding of causality in epidemiology by stressing that science progresses through rejection and modifications of causal hypotheses, not through proof of causality itself.[6] These ideas are all later referenced by Rothman as "basic concepts" of causal inference in epidemiology.[2] Broad conceptualizations of disease etiology are discussed through causal frameworks in both historical and modern epidemiology, beginning with germ theory representations of monocausal frameworks (i.e., necessary causes) such the relationship between microorganisms and disease, to the etiological connection between cigarette smoking and a variety of negative health outcomes, and more recently through directed acyclic graphs (DAGs), a graphical summarization of multiple causal links between risk factors to an outcome of interest.

All these notions of causality in epidemiology are constrained by temporal directionality: cause(s), then effect; exposure(s), then disease. In practice, this relationship is complicated by heterogeneity of causal types, which are representations of how the same exposure can result in different outcomes across individuals due to between-individual characteristic differences (i.e., comparability). Modern epidemiologic practice designates four causal types: doomed, immune, preventive, and causal.[2] Causal types are often presented through counterfactual conceptualizations of the relationships between exposures and disease.[2,7] As an example, if the presence of an exposure is sufficient to explain the occurrence of disease among an individual of the "causal" causal type, the counterfactual explanation of the causal effect requires that the disease would not have occurred in the absence of that same exposure. This explanation accounts for two different, though intersecting, causal theories. In their explanation of sufficient-component cause models, Greenland and Brumbak portray causal mechanisms as the fundamental unit of analysis, instead of the person-level unit of analysis used in the causal types theory discussed previously.[2] Yet the two must interrelate to provide a full understanding of the causal mechanisms at play between exposures and disease. That is, if someone is immune to an exposure for a disease, yet that same exposure is causal for another individual for the same disease, the same sufficient cause conceptualization of the exposure cannot apply to both individuals.

The framework of causal types and sufficient-component cause models imply three broad and interfacing concepts of populations that researchers must consider when performing epidemiologic studies: 1) measured and unmeasured heterogeneity of individual traits, 2) the relationships between these traits to exposures, and 3) causal processes contributing to the occurrence of disease. These considerations provide natural support for reductionist (i.e., statistical) approaches to estimate individual-level effects of exposures on diseases from population data. That is, if a defined exposure is causal for a disease, we should be able to observe a progression towards disease among those that are exposed, and from a population level perspective, these exposed individuals should be similar in some traits, and will aggregate to represent the portion of the population that presents with disease. The same must be true in the opposite direction, we should be able to make broad causal inferences about the relationship between disease and exposure at the person level by isolating a specific risk factor from an aggregate population level. After all, disease must occur in individuals, and if we assume that the population is simply an aggregate of individuals, then we should be able to make both bottom-up and top-down inferences.

Establishing Causality in Epidemiology

Despite the rich history of causal theory and scientific interrogation of natural and social processes contributing to the epidemiologic causal frameworks described previously, these explanations fall short of independently distinguishing causal from non-causal associations, and definitive tests of causality in epidemiology remain elusive. To address the prior shortcomings of causal theory in epidemiology, Hill suggested a set of considerations to be accounted for when attempting to make causal inferences: strength, consistency, specificity, temporarily, biological gradient, plausibility, coherence, experimental evidence, and analogy.[2] In fact, Hill used his work with Doll that established the relationship between smoking and lung cancer as example for the use of the criteria.[5] This list is an expansion of the criteria offered in the landmark U.S. Surgeon General's report on Smoking and Health in 1964, which included the review of over 7,000 scientific articles, ultimately stating that cigarettes are *causally* related to lung cancer and other health conditions.[8]

An overarching goal of the modern epidemiologic approach is to establish a sufficient body of evidence to support a causal relationship between an exposure of interest and an outcome of

interest. Such approaches are traditionally based in regression analyses and, from a philosophical perspective, require two key components: 1) comprehensive and unbiased data for exposures of interest, relevant covariates, and the outcome of interest, and 2) a unidirectional causal framework providing support for the relationships tested (i.e., directed acyclic graphs). When examining the relationship between social influence and cigarette smoking, the body of scientific evidence supporting this causal relationship draws from a mixture of epidemiological, sociological, and economical research – many of which use regression analyses to examine the association.[9–14] Expanding simple regression models to account for correlated data and time dependencies using generalized estimating equations,[15] time series analysis,[16] and marginal structural equation models,[17] provide additional methods to account for the inherent complexity of the mechanisms of social influence and subsequent behavioral outcomes. Nevertheless, despite the strength and formalization of regression analyses for examining patterns and magnitudes of relationships between indicator variables and outcomes, drawing causal conclusions (i.e., if-then) while using these methods remains challenging.

Beyond limitations in the statistical approaches used in epidemiology to isolate the independent effects of one risk factor against an outcome, analyzing outcomes from observational studies that attempt to mimic randomized experiments[18] also limits the scientific pursuit of causality. For example, the intermixing of risk factors and disease outcomes in datasets creates difficulties in determining what associations are due to the absence of accounting for confounding effects of various sources (e.g., missing data, study design, failing to account for unknown confounders) relative to what is truly causal of the outcomes of interest. The possible contributors to the causal effects occurring between the exposure and the disease in the observed world are theoretically infinite. Thus, as done with smoking and various negative health outcomes in the 1964 Surgeon General's Report,[8] the practical requirements for establishing causation, or even strong enough associations to warrant policy intervention, requires enormous human capital, years of focused inquiry, multiple robust datasets, consistent outcomes in both the form of magnitude and outcome of the effect, and substantial contributions from multiple disciplines of science.[2] Epidemiologists are tasked with satisfying a large subset of requirements when making causal inference, while also bearing brunt of hundreds of years of science applied to problems of the highest dimension: humans, their life history, their relationships and interactions with each other, and their environments.

A potential tool that may help us better understand the problem of causality in applied epidemiologic practice is the use of computational modeling and complex systems theory. In contrast to statistical methods, computational models provide flexibility when the pertinent data is insufficient or unavailable.[19–22] Rather than attempt to understand the data through effect isolation and reductionist methods, computational models can, following Epstein's logic, potentially grow and reproduce the same observed patterns of data through a set of causal hypotheses. These models can then provide insights into the mechanisms that generate observed patterns in the real world if their hypothesized processes are feasible explanations of the outcome. The finite environment of a model is limited to the processes and traits that the researchers choose to include. As a result, the combination of the processes and assumptions of various mechanisms in the computational model are, by definition, sufficient causes for every outcome that the model can produce. Thus, the application of computational modeling provides a method to validate the causal frameworks of statistical models.

Added Value from Computational Models in Causality and Epidemiology

The benefits of mechanistic models, applied to the field of epidemiology, are numerous and public health science may benefit from expanding their application towards causal frameworks. Here, I focus primarily on ABMs given their typical application in growing population-level patterns from individual-level processes. The most commonly cited ABMs in the complex systems literature are substantiated on the notion that hypothesized mechanisms contribute to the emergence of a phenomenon, effectively through component cause theory. For example, Schelling's segregation model was simple enough to implement on a checkerboard with nickels and dimes, yet sufficient to explain a simple causal mechanism for segregation: slight preference for homophily.[23] Similarly, Doringo demonstrated that solution sharing, otherwise known as swarm intelligence, is sufficient to define path optimization between two points, such as that between an ant hive and a food colony.[24] Both models do not use empirical data collected from observational or controlled studies in the models themselves. Instead, these models use existing theories derived from a large body of empirical studies (i.e., those utilizing statistics) to develop data-free algorithms of interaction that produce patterns observed in the real world. The outcomes of the models are thus able to demonstrate the potential causal underpinnings of said patterns in a closed-solution environment, free from unobserved confounding that hounds existing

statistical methods or the excessive complexity that may obfuscate potentially important mechanisms or processes. In epidemiologic research, robust abstract models have presented numerous opportunities for a better understanding of population health, although they remain isolated examples. In cancer, multistage carcinogenesis models, such as the Armitage-Doll model[25] or the Two-Stage clonal expansion model[26–28] can reproduce the population patterns of cancer incidence based on minimal assumptions, and have been used since the 1950s to generate hypotheses about cancer mechanisms and epidemiology. Similarly, in infectious disease epidemiology, the SIR (Susceptible-Infectious-Recovered) model by Kermack-McKentrick[29] is based on simple principles of disease transmission, and can produce simulated population patterns of disease consistent with observed epidemics.[30–32] These models have been able to shed light on the mechanisms of herd immunity, and has led to the evolution of a mathematical (infectious) disease modeling subfield in epidemiology. Coupled with empirical data and statistical methods, the conceptual frameworks of these models provide powerful insights into the fundamental causes of observed phenomenon. Yet the dearth of such models in social epidemiology leave much to be explored.

The utility of systems science methods in social epidemiology has been previously discussed in the literature with Auchincloss,[33] Diez Roux[34] and Galea[35] who all suggest that agent-based models (ABMs) are likely to contribute to an understanding of social dynamics and towards the development of theories of disease causality. A discussion of the distinct traits of ABMs compared to other computational modeling methods can be found in Appendix A2.1. Despite this potential, ABMs in the epidemiologic literature are often implemented as a data-driven method rather than as a theoretical one. Examples of this include recent attempts by Marshall and Galea to formalize the role of ABMs in epidemiologic theory of causality by contextualizing the method in the counterfactual frameworks[36], El-Sayed et al.'s exploration of obesity interventions,[37] Hennessy et al.'s step-by-step guide to build an ABM to explore obesity interventions,[38] and Yang et al.'s analysis of the contributions of land use and socioeconomic differences in walking behavior.[39] One study by Cerdá et al. includes full regression models and corresponding beta estimates for risk factors in the ABM to determine the probability of violence perpetration.[40] While these studies all utilize the ability of ABMs to account for interference, interdependence, and interaction, their primary focus is not on building causal frameworks. Rather, the previously discussed studies use ABMs within existing epidemiologic frameworks to quantify counterfactual effect estimates.

When examining the applications of ABMs in the epidemiologic literature, Auchincloss et al.[41] is the only publication, to our knowledge, to approach modeling from the sufficient cause perspective in the epidemiologic sciences literature. This dearth of ABMs exploring sufficient cause mechanisms is perhaps due to modeling philosophy in epidemiology, which Galea has described as a "pragmatic" field.[42] Moreover, even simple mechanistic models take substantially more time and effort to understand, build, and tune, relative to statistical methods, for which existing frameworks of analyses and black box software,[43] are readily available. Ethical, time, and resource constraints limit epidemiologists to analyzing outcomes from observational studies that attempt to mimic randomized experiments. Yet as Hernán observes, in attempting to obtain and scrutinize high-quality data from many individuals using statistical methods of association, the role of theory in epidemiology has been minimized.[18]

Throughout this dissertation, we sought to present a different understanding of ABMs through highlighting the importance of epidemiologic theory from a causal framework perspective (i.e., providing generalizable and broader explanations for why we observe certain statistical associations from the empirical data). Our goal is to demonstrate an instance of a causality in an epidemiologic framework that is relatively free of linear assumptions and data driven results (i.e., in comparison to statistical methods, Chapter 4). We also highlight the importance of applied statistical methods in revealing hypothesized causal mechanisms that can be further explored by sufficient cause computational models (Chapter 3). Additionally, statistical models provide quantitative boundaries representative of the real world, such that when abstract models function within these boundaries, their results can have real world implications (Chapter 2).

Summary of Findings

In the second chapter, we observe that based on current knowledge of the patterns of adult e-cigarette use by smoking status and the heavy concentration of e-cigarette use among current smokers, the simulated effects of e-cigarettes on smoking cessation generate substantially larger changes to smoking prevalence compared to their effects on smoking initiation.[44] Additionally, our findings do not change, even when the model accounts for large initiation effects whereby youth and adult smoking initiation rates escalate due to e-cigarettes.[45] Overall, we observe that under current conditions, the cessation effects of e-cigarettes on adult smoking offset these

undesirable "gateway" effect even if adolescents initiate e-cigarette use at higher rates than currently observed, and e-cigarettes increase smoking initiation among these adolescents by substantially. This model also considers the influence of population levels of e-cigarette use on individual likelihood of e-cigarette initiation. Additionally, we consider feedbacks of e-cigarette and cigarette use influence at the individual level to initiate e-cigarettes, smoking, or to quit smoking. Despite lacking predictive power, the results of this model quantify relative estimates, and highlight the how the modeling process itself can contribute to a better understanding of a complex public health issue from the perspective of populations and individuals. While robust longitudinal data on e-cigarette initiation and prevalence patterns are still in development, simulation models such as ours can help to determine the best public policies that tip the balance toward desirable public health outcomes.

The third and fourth chapter suggests that overall differences observed in smoking prevalence by urbanicity, network density, race, and sex can be explained by the interfacing of differences in network structure with peer influence. Computational modeling as an exercise alone encourages researchers to explicitly state their assumptions and justify the inclusion of any hypothesized cause by considering it within a set of contextual processes that result in an outcome.[20] This method of inquiry differs from the process of variable selection in statistics. In these chapters, stratifying model results by "school network density," "urbanicity," or "race" in a generalized linear model suggests a causal hypothesis that accounts for differences by these variables when determining the effects of peer influence on smoking behavior. Yet the quality of living in a "rural" area, for example, is not itself responsible for the mechanism that perpetuates its relationship with other risk factors and smoking behavior. Rather, it is in the experience of living in a rural area that is can contribute to the differences observed by urbanicity in smoking behavior resulting from peer influence. In mechanistic framework, we determine how urbanicity changes individual-level experiences by explicitly describing (via functions) how these processes can generate the differences that we see observe the magnitude and direction of the effect that urbanicity may have on the relationship between peer influence and smoking behavior. This flexibility to define mechanisms and processes can help us develop better frameworks of causality while providing additional insights into smoking etiology.

Hernán suggests that "many practicing epidemiologists, attached to their cherished data, may not be prepared to jump head first into the world of agent-based modeling."[18] The continued

dearth of abstract ABMs in social epidemiology suggest that many are perhaps still unwilling to embrace the theoretical benefits of the method over its ability to generate quantitative outcomes. However, this patterning of published ABMs in the epidemiologic literature may be a symptom of epidemiologists embracing the added value of modeling, yet using ABMs in situations that are suboptimal, when other computational or advanced statistical methodologies are better suited to the stated problems. Ultimately, we find many published ABMs justifying the approach by stating that traditional epidemiologic methods do not account for nonlinearity, feedbacks of exposures, or correlated data.[46–49] These claims are somewhat deceiving in the face of advanced statistical methods, for example the maximum likelihood and generalized estimating equation techniques by Vanderweele et. al.[50,51]

In consideration of the criticisms of the still-developing literature that applies and discusses the utility of ABMs and mechanistic models,[51] the outcomes of this this dissertation attempt to provide a stepping stone towards the development of a better understanding of the potential theoretical and scientific value added by abstract computational models to epidemiologic research. The outcomes described in this chapter are not meant to simply demonstrate potential and abstract causal mechanisms of the differences in smoking behavior across a variety of contexts with varying assumptions. Rather, these outcomes are an extension of an enormous body of statistical models that analyze population representative data, which provided insights and practical implications for the mechanistic models built in this dissertation. As a result, we can infer potential real world implications of what would otherwise only be considered theoretical results. In other words, while these results are not predictive, they provide greater insight into the inner workings of tobacco use among adults and adolescents, beyond quantitative estimates, and suggesting sufficient cause mechanisms of real world patterning. These mechanisms can contribute to a better understanding of why we observe the results of statistical outcomes, thereby providing justifications and support for public health intervention. Our results suggest evidence to consider when determining the appropriate e-cigarette regulation policies, and in the development of network based interventions to encourage adolescents to abstain from or quit smoking. In environments where vast amounts of time and money are invested into better understanding causal mechanisms of health behavior, computational models in combination with statistical models can provide better guidance for intervention and randomized experiments than either method independently.

This dissertation provides a large platform for future research in the application of computational modeling methods towards tobacco control epidemiology. As longitudinal data of e-cigarettes become more available, the model presented in Chapter 2 can be extended to account for additional factors, including excess mortality due to e-cigarette use, as well as incorporating additional tobacco products such as smokeless tobacco or heat not burn cigarettes. Additionally, from a mechanistic standpoint, we could tie the e-cigarette model from Chapter 2 into the network models of Chapter 3 and 4 to provide insights into potential tipping points of peer influence, and worst- and base- case scenarios of cigarette smoking behavior resulting from propagation of e-cigarette use through social networks. Finally, we could extend the findings in Chapter 3 and 4 to provide a standardized framework of validating causal pathways considered in statistical models through the application of mechanistic simulation models more broadly. The practice of science has begun an interdisciplinary integration process that has been able to expand our understanding of nature and social systems. Mathematics and biology integrate to better understand cancer from a cellular level,[52] while physics and sociology have united to contribute an entirely new understanding of social science through the study of networks.[53,54] This dissertation attempts to tie understandings from two already vastly interdisciplinary fields: complex systems and epidemiology. It is with feelings of great anticipation that these two disciplines can standardize their understanding of one another, such that the added value of computational models can extend far beyond fitting models to data and prediction of future disease dynamics, to providing causal frameworks for revealing the underlying mechanisms of disease and health behavior.

References

1. Epstein JM. Generative Social Science: Studies in Agent-Based Computational Modeling. *Gener Soc Sci*. 2006:356. doi:10.1017/CBO9781107415324.004.
2. Rothman KJ, Greenland S, Lash T. *Modern Epidemiology, 3rd Edition*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. doi:10.1002/hast.292.
3. Rothman KJ. *Epidemiology: An Introduction*. 2nd ed. New York, NY: Oxford University Press; 2012.
4. Hume D. *An Equiry Concerning Human Understanding*. (Millican P, ed.). New York, NY: Oxford University Press; 2007.
5. Bradford-Hill A. The Enviroment and Disease: Association or Causation? *Proc R Soc Med*. 1965;58:295-300. doi:DOI: 10.1016/j.tourman.2009.12.005.
6. Lucas RM, McMichael AJ. Association or Causation: Evaluating Links Between "Environment and Disease". *Bull World Health Organ*. 2005;83(10):792-795.
7. Hernan MA, Robins JM. *Causal Inference*. Forthcoming: Taylor and Francis Group; 2017.
8. United States Public Health Service. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*.; 1964.
9. Kiuru N, Burk WJ, Laursen B, Salmela-Aro K, Nurmi JE. Pressure to Drink but Not to Smoke: Disentangling Selection and Scialization in Adolescent Peer Networks and Peer Groups. *J Adolesc*. 2010;33(6):801-812. doi:10.1016/j.adolescence.2010.07.006.
10. McGloin JM, Sullivan CJ, Thomas KJ. Peer Influence and Context : The Interdependence of Friendship Groups, Schoolmates and Network Density in Predicting Substance Use. *J Youth Adolesc*. 2014;43:1436-1452. doi:10.1007/s10964-014-0126-7.
11. Kobus K. Peers and Adolescent Smoking. *Addiction*. 2003;98(Suppl 1):37-55.
12. Duncan GJ, Boisjoly J, Kremer M, Levy DM, Eccles J. Peer Effects in Drug use and Sex Among College Students. *J Abnorm Child Psychol*. 2005;33(3):375-385. doi:10.1007/s10802-005-3576-2.
13. Ennett ST, Bauman KE, Hussong A, et al. The Peer Context of Adolescent Substance Use: Findings from Social Network Analysis. *J Res Adolesc*. 2006;16(2):159-186. doi:10.1111/j.1532-7795.2006.00127.x.
14. Ali MM, Dwyer DS. Estimating Peer Effects in Adolescent Smoking Behavior: A Longitudinal Analysis. *J Adolesc Heal*. 2009;45(4):402-408. doi:10.1016/j.jadohealth.2009.02.004.

15. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations between Neighborhood Risk Factors and Health. *Epidemiology*. 2010;21(4):467-474. doi:10.1097/EDE.0b013e3181caeb90.

16. Ionides EL, Bretó C, King AA. Inference for Nonlinear Dynamical Systems. *Proc Natl Acad Sci*. 2006;103(49):18438-18443. doi:10.1073/pnas.0603181103.

17. VanderWeele TJ. Marginal Structural Models for the Estimation of Direct and Indirect Effects. *Epidemiology*. 2009;20(1):18-26. doi:DOI 10.1097/EDE.0b013e31818f69ce.

18. Hernán MA. Invited commentary: Agent-based models for Causal inference - Reweighting Data and Theory in Epidemiology. *Am J Epidemiol*. 2015;181(2):103-105. doi:10.1093/aje/kwu272.

19. United States Department of Health and Human Services. Appendix 15.1 Tobacco Control Simulation Models. In: *The Health Consequences of Smoking—50 Years of Progress A Report of the Surgeon General*. 2014.

20. Epstein J. Why model? *J Artif Soc Soc.* 2008;11(4):6. doi:10.1080/01969720490426803.

21. Feirman SP, Donaldson E, Glasser AM, et al. Mathematical Modeling in Tobacco Control Research: Initial Results From a Systematic Review. *Nicotine Tob Res*. 2015;05/16(3):1-14. doi:10.1093/ntr/ntv104.

22. National Cancer Insitute. Greater Than the Sum. Systems Thinking in Tobacco Control. *Tob Control Monogr*. 2007;18:1-67.

23. Schelling TC. Dynamic Models of Segregation. *J Math Sociol*. 1971;1(2):143-186. doi:10.1080/0022250X.1971.9989794.

24. Bonabeau E, Dorigo M, Theraulaz G. *Swarm Intelligence: From Natural to Artificial Systems*. Vol 4.; 1999. doi:10.1080/09540090210144948.

25. Armitage P, Doll R. A Two-stage Theory of Carcinogenesis in Relation to the Age Distribution of Human Cancer. *Br J Cancer*. 1957;11(2):161-169.

26. Meza R, Jeon J, Moolgavkar SH, Luebeck EG. Age-specific Incidence of Cancer: Phases, Transitions, and Biological Implications. *Proc Natl Acad Sci* . 2008;105(42):16284-16289. doi:10.1073/pnas.0801151105 .

27. Moolgavkar SH, Knudson AG. Mutation and Cancer: A Model for Human Carcinogenesis 1,2. *J Natl Cancer Inst*. 1981;66(6):1037-1052.

28. Meza R. Some Extensions and Applications of Multistage Carcinogenesis Models. 2006.

29. Kermack WO, McKendrick AG. Contributions to the Mathematical Theory of Epidemics. *Proc R Soc London*. 1927;115(772):700-721. doi:10.1098/rspa.1927.0118.

30. Mayer BT, Eisenberg JNS, Henry CJ, Gomes MGM, Ionides EL, Koopman JS. Successes and Shortcomings of Polio Eradication: A Transmission Modeling Analysis. *Am J Epidemiol*. 2013;177(11):1236-1245. doi:10.1093/aje/kws378.

31. Eisenberg JNS, Li S, Spicknall IH, Koopman JS. Dynamics and control of infections transmitted from person to person through the environment. *Am J Epidemiol*. 2009;170(2):257-265. doi:10.1093/aje/kwp116.

32. Koopman J. Modeling Infection Transmission. *Annu Rev Public Health*. 2004;25(1):303-326. doi:10.1146/annurev.publhealth.25.102802.124353.

33. Auchincloss AH, Diez Roux AV. A New Tool for Epidemiology: The Usefulness of Dynamic-agent Models in Understanding Place Effects on Health. *Am J Epidemiol*. 2008;168(1):1-8. doi:10.1093/aje/kwn118.

34. Diez Roux AVD. Complex Systems Thinking and Current Impasses in Health Disparities Research. *Am J Public Health*. 2011;101(9):1627-1634. doi:10.2105/AJPH.2011.300149.

35. Galea S, Riddle M, Kaplan GA. Causal Thinking and Complex System Approaches in Epidemiology. *Int J Epidemiol*. 2010;39(1):97-106. doi:10.1093/ije/dyp296.

36. Marshall BDL, Galea S. Formalizing the Role of Agent-Based Modeling in Causal Inference and Epidemiology. *Am J Epidemiol*. 2015;181(2):92-99. doi:10.1093/aje/kwu274.

37. El-Sayed AM, Seemann L, Scarborough P, Galea S. Are Network-based Interventions a Useful Antiobesity Strategy? An Application of Simulation Models for Causal Inference in Epidemiology. *Am J Epidemiol*. 2013;178(2):287-295. doi:10.1093/aje/kws455.

38. Hennessy E, Ornstein JT, Economos CD, et al. Designing an Agent-Based Model for Childhood Obesity Interventions: A Case Study of ChildObesity180. *Prev Chronic Dis*. 2016;13:E04. doi:10.5888/pcd13.150414.

39. Yang Y, Diez Roux A V., Auchincloss AH, Rodriguez DA, Brown DG. A Spatial Agent-Based Model for the Simulation of Adults' Daily Walking within a City. *Am J Prev Med*. 2011;40(3):353-361. doi:10.1016/j.amepre.2010.11.017.

40. Cerdá M, Tracy M, Ahern J, Galea S. Addressing Population Health and Health Inequalities: The Role of Fundamental Causes. *Am J Public Health*. 2014;104(SUPPL. 4). doi:10.2105/AJPH.2014.302055.

41. Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential segregation. *Am J Prev Med*. 2011;40(3):303-311. doi:10.1016/j.amepre.2010.10.033.

42. Galea S. An Argument for a Consequentialist Epidemiology. *Am J Epidemiol*. 2013;178(8):1185-1191. doi:10.1093/aje/kwt172.

43. Joppa LN, McInerny G, Harper R, et al. Troubling Trends in Scientific Software Use. *Science (80)*. 2013;340(6134):814 LP-815. http://science.sciencemag.org/content/340/6134/814.abstract.

44. Cherng ST, Tam J, Christine PJ, Meza R. Modeling the Effects of E-Cigarettes on Smoking Behavior: Implications for Future Adult Smoking Prevalence. *Epidemiology*. 2016. doi:10.1097/EDE.0000000000000497.

45. Cherng ST, Tam J, Christine PJ, Meza R. The Authors Respond. *Epidemiology*. 2017;28(1).

46. El-Sayed AM, Scarborough P, Seemann L, Galea S. Social network analysis and agent-based modeling in social epidemiology. *Epidemiol Perspect Innov*. 2012;9(1):1. doi:10.1186/1742-5573-9-1.

47. Luke DA Da, Stamatakis KA. System Science Methods in Public Health: Dynamics, Networks, and Agents. *Annu Rev Public Health*. 2012;33(1):357-376. doi:10.1146/annurev-publhealth-031210-101222.Systems.

48. Bonabeau E. Agent-based Modeling: Methods and Techniques for Simulating Juman Systems. *Proc Natl Acad Sci*. 2002;99(suppl. 3):7280-7287. doi:10.1073/pnas.082080899.

49. Homer JB, Hirsch GB. System Dynamics Modeling for Public Health: Background and Opportunities. *Am J Public Health*. 2006;96(3):452-458. doi:10.2105/AJPH.2005.062059.

50. VanderWeele TJ, Tchetgen EJT, Halloran ME. Interference and Sensitivity Analysis. *Stat Sci*. 2014;29(4):687-706. doi:10.1214/14-STS479.

51. Naimi AI. Integrating Complex Systems Thinking into Epidemiologic Research. *Epidemiology*. 2016;27(6):843-847. doi:10.1097/EDE.0000000000000538.

52. Anderson ARA, Quaranta V. Integrative Mathematical Oncology. *Nat Rev Cancer*. 2008;8(3):227-234. http://dx.doi.org/10.1038/nrc2329.

53. Ball B, Newman MEJ. Friendship Networks and Social Status. *Netw Sci*. 2013;1(1):16-30. doi:10.1017/nws.2012.4.

54. Budak C, Watts D. Dissecting the Spirit of Gezi: Influence vs. Selection in the Occupy Gezi Movement. *Sociol Sci*. 2015;2:370-397. doi:10.15195/v2.a18.