Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

______________________________ _______________

Yuefan Shao Date

From Epidemiology to Decision Making:
A Systems Science Approach to Evaluate Effectiveness of Complex Behavioral Interventions

By

Yuefan Shao
Doctor of Philosophy
Epidemiology

Alvaro Alonso, MD, PhD
Chair

Viola Vaccarino, MD, PhD
Co-Chair

Shakira F. Suglia, ScD, MS
Committee Member

David Mendez, PhD, MS
Committee Member

Weihua An, PhD, MA
Committee Member

Accepted:

Kimberly Jacob Arriola, PhD
Dean of the James T. Laney School of Graduate Studies

Date

From Epidemiology to Decision Making:
A Systems Science Approach to Evaluate Effectiveness of Complex Behavioral Interventions

By

Yuefan Shao
MPH, University of Michigan, Ann Arbor, 2017
BS, University of Michigan, Ann Arbor, 2016

Advisors: Alvaro Alonso, MD, PhD and Viola Vaccarino, MD, PhD

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2022

Abstract

From Epidemiology to Decision Making:
A Systems Science Approach to Evaluate Effectiveness of Complex Behavioral Interventions

By Yuefan Shao

Modifiable health behaviors are key to cardiometabolic disease prevention. A significant number of behavioral interventions have been proposed for healthy behavior promotion. However, identifying the most effective type of behavioral intervention for a given population remains challenging due to two main reasons. First, there is a lack of data for behavioral intervention effectiveness evaluation. Second, effectiveness of complex behavior intervention is dependent on multi-level factors, which poses challenges in intervention outcome evaluation. To address these challenges, this dissertation first empirically identified contributing factors associated with patterns of cigarette smoking and physical activity throughout the life-course. In addition, a complex systems modeling approach was used to evaluate effectiveness of different types of behavioral intervention, given target population network characteristics and individual behavior incentive.

Aim 1 characterized trajectories of physical activity and cigarette smoking from early adolescence to adulthood. Using latent class growth mixture model, results showed that there are three sub-groups of individuals sharing similar patterns of physical activity and past 30-day cigarette smoking behavior from early adolescence to adulthood. Age, socio-demographic and early-life psychological factors are important predictors of trajectories for both behaviors.

Aim 2 used social network analysis and regression methods to evaluate the association between social network characteristics and physical activity/cigarette smoking behaviors during adolescence and the adolescence to young adulthood transition. Results suggest that individuals' health behaviors at younger age are the strongest predictors of health behaviors during young adulthood. In addition, an individual's social position during adolescence is a predictor for physical activity level during young adulthood but not for cigarette smoking.

Using computational models, Aim 3 showed that when taking into consideration diffusion of interventions within a network, a highly clustered network does not imply the necessity of network-based intervention. Paradoxically, for networks with longer average path length or unknown network structure, incentivizing individuals might be more effective than interventions on popular opinion leader.

Collectively, findings highlighted that both social network and individual-level heterogeneity are key to shaping population level distributions of health behaviors. In addition, researchers need to embrace a systems science lens when evaluating complex behavioral intervention outcomes.

From Epidemiology to Decision Making:
A Systems Science Approach to Evaluate Effectiveness of Complex Behavioral Interventions

By

Yuefan Shao
MPH, University of Michigan, Ann Arbor, 2017
BS, University of Michigan, Ann Arbor, 2016

Advisors: Alvaro Alonso, MD, PhD and Viola Vaccarino, MD, PhD

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2022

ACKNOWLEDGEMENTS

This dissertation would not have been possible with the support and love for all the important people/companions in my life.

First and foremost, I want to thank my advisor, Alvaro Alonso for all his support and guidance throughout all these years. Alvaro, thank you so much for being the best advisor I could ever ask for. I have learned so much from you, not just about epidemiology and research, but also about life. Thank you for sharing the wisdom and all the guidance throughout all these years. Even though, I probably will never be able to have the super power like yours to manage all the projects so well, I am really happy that I now procrastinate slightly less. Dr. Shakira Suglia, Dr. Viola Vaccarino, and Dr. Weihua An, thank you all for being the Rockstar committee members for all the great support and flexibility throughout the entire process. Dr. David Mendez and Dr. Marisa Eisenberg from University of Michigan, thank you both so much for introducing me to the fascinating field of complex systems and encouraging me to keep pursuing the questions that I am genuinely passionate about. And Dr. Mendez, thank you for being a life mentor throughout all these years as well.

I want to thank my epidemiology cohort at Emory for their companionship throughout the journey. Ryan Saelee, Minxuan Huang, Qi Zhang, Haomin Li, Vinita Subramanya, Lihan Zhao, Randi J. Holtz, and Brittany Flores, I owe you all tons. Thank you for being the best friends and support I could ask for. The laughter and tears shared with you all have been the best memories of the past five years.

Last, I want to thank my parents and partner Hayden Hedman for always being my rock. I also want to thank the best kitties in the world: Jake and Sydney for being my motivation and companion whenever life gets hard. You guys are the best goofballs in the world. And thank you Loki and Tony for being most supportive doggies and for bringing so much joy into my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Background and Literature Review

Recent Advancement in Cardiovascular Disease Prevention and Modifiable Behavioral Risk Factors

The year 2020 marked the end of the 7-year global action plan on non-communicable disease prevention from the World Health Assembly(1,2) and the final year of the Healthy People 2020 initiative to promote nationwide health in the United States.(3) Over the past decade, significant progress has been achieved in cardiovascular disease (CVD) prevention and burden reduction. Based on the latest statistics, higher proportions of US population across all age groups are meeting the ideal levels for cardiovascular health indicators such as total cholesterol level, blood pressure and smoking.(4) However, heart disease remains the leading cause of death in the US, and disparities across different sociodemographic groups persist.(4) Moreover, CVD remains the single most costly disease. In 2016, CVD-related expenditure was around $550 billion in the US. If no significant improvements are made in CVD prevention, such expenditure is estimated to be around $1 trillion by the year 2035.(5)

Key modifiable behavioral risk factors including smoking and physical inactivity have long been identified as critical preventable components for cardiovascular health.(6) Over the last decade, with several policy-level interventions on tobacco use in place and increasing number of personalized commercial products for physical activity promotion, the prevalence of tobacco use and physical inactivity has declined drastically over the past decade. The prevalence of physical inactivity in the US dropped to 26%, which was below the targeted 32% for Healthy People 2020 initiative.(4) Nonetheless, substantial disparity persists in tobacco product use and physical inactivity, disproportionally affecting individuals identified as racial/ethnic minority, lesbian, gay, bisexual, transgender or from lower socioeconomic background.(4) In addition, with the emergence of electronic and flavored tobacco product, the prevalence of such new forms of tobacco product is on the rise, especially among adolescents and young adults.(4) Collectively, evidence indicates a dire need for a continuous effort in CVD prevention and disparity reduction, especially through addressing burdens in tobacco use behavior and physical inactivity.

Overview of Interventions for Physical Activity Promotion

During the early 2000s, as public health services expanded more in the realm of chronic disease management and behavioral intervention, physical activity promotion started to gain increasing amount of attention not only from medical professionals, but also from organizations in both the public and private sectors. Over the past two decades, two major transformations occurred in the field of physical activity promotion.

Firstly, prior to 2013, increasing amount of social and behavioral theories were incorporated into behavioral interventions. Major theories included the Transtheoretical Model(7), the Social Cognitive Theory(7), the Social Ecological Model(8), and the Social Marketing Theory(9). Based on the most prevalent theory to-date for behavioral intervention - the Social Ecological Model (SEM), actions can be taken on five different levels to initiate long-term behavioral change: intrapersonal, interpersonal, community, institution and policy. These theories were well incorporated into a large amount of governmental or privately-funded community and group-based interventions that emerged during the early 2000s. By early 2000s, three major types of physical activity interventions were present: informational approaches, socio-behavioral approaches, and environmental/policy approaches.(10) Informational approaches aimed to promote physical activity through ensuring access to information among populations. Key informational approaches included: 1) point-of-decision prompts, which were signs next to places such as escalators to encourage individuals to be more active; 2) Community-wide or mass media campaigns through advertisement display, message delivery via newspapers or radios. Such approaches mainly targeted the intra-personal and community levels, based on the SEM. The second type of intervention -- socio-behavioral approaches, such as school/classroom-based physical education, family-based support interventions, and social support interventions -- aimed to facilitate individual physical activity promotion and behavioral change through peer pressure, peer support and social support. These approaches targeted mainly the inter-personal and community level of the SEM. Examples include the Stanford Five-City Project and the Pawtucket Heart Health Program.(11–15) One special type of social-behavioral interventions different from previously mentioned ones was individually-adapted behavioral change programs. In such programs, tailored behavioral change activities such as set daily goals of physical activity level, behavioral reinforcement through self-reward or positive rewards were delivered either in a group setting or directly to individuals through phone/mail. Individually adapted behavioral change programs allowed for the integration of both intrapersonal and interpersonal perspectives of the SEM.(10) The last major behavioral intervention approach that emerged during the early 2000s was the environmental/policy approach. Such interventions included federal and local effort to promote physical activity and examples included enhancing neighborhood characteristics such as biking lane establishment, safety lighting in the neighborhood, etc.(10)

Starting 2005, the digital revolution reached a new stage and facilitated the second major transformation in physical activity promotion interventions.(16) During this period, increasing amount of individuals had access to smartphones and social media. Novel digital physical activity promotion methods including smartphone applications, wearable devices, video games and social media physical activity engagement started to replace the ones that used traditional routes such as physical signs, radio or advertisement display on billboards.(17) Two distinct new features that made physical activity promotion in the digital era drastically different from the traditional ones prior to early 2000s were: 1) increased incorporation of digital technology; 2) more tailored to the individual.(16) These digitalized physical activity interventions have been implemented in various settings ranging from individual homes to school, and to the entire social network platform. Majority of these interventions utilized the

Social Cognitive Theory to facilitate individual physical activity promotion through establishing rewards and behavioral motivation via gamification and positive reinforcement. However, access to these interventions varied largely by individuals' socio-economic background and adoption. Additionally, adherence to these programs relied more heavily upon individual's incentives. Examples of these individual-based interventions included Fitbit, smartphone application linked to social network account for socially engaged physical activity (e.g. Keep), Ring Fit Adventure on Nintendo Switch and Physical Activity on Prescription.(18),(19)

<u>Overview of Interventions Targeting Tobacco Product Use</u>

<u>Tobacco Control Efforts Targeting Conventional Tobacco Products</u>

The potential hazard associated with tobacco product use was first identified in the early 1950s. In 1954, Doll & Hill's paper "*The mortality of doctors and their smoking habits*" marked the beginning of the battle for tobacco control in the United States.(20) In 1965, the first federal act requiring a warning label display on all cigarette packages was passed. Meanwhile, the federal excise tax on smokeless tobacco product was repealed. Fast forward, in the next three decades, major breakthroughs in federal and community-level tobacco control effort were achieved. By 1998, California had become the first state that passed a statewide smoke-free indoor air law and cigarette taxation had been enacted across all states in the United States. Starting early 2000s, increasing number of mass media educational campaigns were launched by the Center for Disease and Control and the Food and Drug Administration (FDA).(21) Till today, four main types of federal and state level tobacco control efforts targeting traditional tobacco products are in place: cigarette excise tax, statewide smoke-free indoor air law, required warning label display and mass media campaigns. These four types of federal and state level tobacco control actions can be further categorized into two types: 1) One type with institutionalized power that target individuals' conventional tobacco product use through law enforcement (e.g. smoke-free indoor air law); 2) One type aiming to alter individuals' incentive through increasing perceived negative externality (e.g. increasing cigarette price through cigarette excise tax, establishing perceived negative consequences associated with tobacco product use through media campaign messages). Collective evidence suggests that excise taxes, mass reach anti-tobacco campaigns and smoke-free indoor air laws have been the most important contributors to the declining prevalence of tobacco product use in the past decade.(22,23)

In addition to the large-scale public health interventions targeting traditional tobacco product use listed above, individual and group-based interventions have also been critical for tobacco control, especially in recent years. Similar to interventions targeting other types of modifiable health behaviors, these interventions utilize socio-behavioral theories such as the Transtheoretical Model and the Social Cognitive Theory. Examples of these programs include: group-based smoking cessation programs, individual-based counseling, text-based messaging, and social media based social groups for smoking cessation.(23) These interventions mainly aim to alter individual’s behavior incentives, either through social support, peer pressure, or through altering perception regarding utility of tobacco product use. Sometimes these interventions are combined with pharmacological intervention such as medication prescription, given the addictive nature of nicotinic products.(22–24)

Tobacco Control Effort Targeting Electronic Cigarettes

In 2003, the first electronic cigarette (e-cigarette) was produced, and was later introduced to the U.S marketplace in 2007. Emergence of e-cigarette as a disruptive technology changed the landscape of tobacco use behavior as well as tobacco control. Initially, e-cigarettes were marketed as “less-harmful substitute” for traditional cigarettes.(25) However, increasing amount of evidence started to suggest that e-cigarettes, sometimes also known as vapes, or electronic nicotine delivery system (ENDS), were harmful, especially for youth, young adults and pregnant women.(25,26),(27) To-date, e-cigarettes have become increasingly popular and slowly started to replace the traditional tobacco products, especially among younger adults, despite the tremendous amount of health hazards associated with them. Studies have shown that different from conventional cigarettes, e-cigarettes have branded themselves successfully as “modern”, “cool” for younger audiences through marketing strategies, such as celebrity marketing, associating the product with social status.(27),(28) Evidence suggested that exposure to e-cigarette product marketing content was associated with higher rate of e-cigarette initiation among younger adults and e-cigarette was associated with more positive perceived norm as compared to conventional tobacco products.(29), (30) Moreover, many e-cigarette products came with kid-friendly flavors and allowed for use with other substances including cannabis to make them more appealing to young adults. As a result, by 2018, about one in every five high school students and one in twenty middle school students were e-cigarette users. (25,26)

The tremendous success of marketing and sales of e-cigarettes especially among younger adults suggest that designs of interventions targeting e-cigarettes need to be able to address the unique features of e-cigarettes that make them particularly attractive to younger populations. Therefore, to address the epidemic of e-cigarette use two major types of interventions are currently in place: federal tobacco product marketplace regulations(31) and individual/group-based interventions.(22,23) Through the Family Smoking Prevention and Tobacco Control Act signed in 2009(31) (the Tobacco Control Act), FDA has been authorized to regulate marketplace e-cigarettes through regulating e-cigarettes’ manufacturing, premarket review, marketing (e.g. packaging and advertising), and distribution. These federal actions can facilitate modifications of individual’s e-cigarette use behavior through limiting individuals’ access to e-cigarettes or altering individuals’ perceived utility to initiate e-cigarettes through information displayed on product packaging or advertisements. In addition to federal level efforts that restrict e-cigarette access, limited number of individual and group-based interventions are also available to assist with e-cigarette use control. However, different from those targeting conventional tobacco products, the target population of these interventions are mostly adolescents and young adults. Examples of these programs include the smartphone application for youth vaping cessation *2Morrow Health*, the *Not-On-Tobacco* initiative and the *INDEPTH* program by the American Lung Association, as well as *Get Your Head Out of the Cloud campaign*.(32,33) Similar to conventional individual/group-based tobacco control programs, these programs offer individual/group-based counselling sessions that facilitate behavioral change through established peer support, psychological counseling to alter individual perceived utility associated with e-cigarette use and positive rewards-based learning.

<u>Intervention Effectiveness for Physical Activity Promotion and Tobacco Cessation</u>

To sum up, there are four major domains of interventions targeting physical activity and tobacco use behavior: policy initiatives, community-based programs, informational approaches, and individual-based interventions.(34) Policy initiatives are enforced upon a designated population, hence the entire targeted population have access to the "intervention" and will adhere to it subsequently. Examples include smoke-free indoor air laws, cigarette excise taxes, and complete tobacco cessation policies such as health insurance coverage of cessation program.(35),(36) Such interventions have been estimated to be most effective due to their institutional power.(37), (38) However, feasibility and implementation of policies remain a challenge, especially when given time and resource constraints. Unlike policy initiatives, the other three domains lack the institutional power and therefore, cannot guarantee intervention uptake. Community/group-based programs such as school/classroom based physical activity promotion and group-based tobacco cessation programs utilize capacity building within a targeted population and foster individual behavioral change through reinforcement learning, peer support, and social norm establishment.(39),(40) However, the effect of the interventions did not manifest until years after the initial introduction and adherence remained the key challenge for behavioral change in the long term.(11),(41), (42) Informational approaches including mass media campaigns and text messaging interventions have been shown effective in promoting more physical activity and smoking cessation in some cases, but findings regarding the effectiveness of these campaigns were quite mixed across heterogeneous campaign design and study populations.(43–45) Lastly, with the rapid expansion of the global mobile health (mHealth) market and increasing awareness of healthy lifestyles, individual-based programs including physical activity on prescription and mHealth applications to promote health-benefiting behaviors are becoming increasingly prevalent in the population. Nonetheless, findings on the effectiveness of these interventions are still inconsistent.(46–48) Moreover, as previously mentioned, selective access to these interventions and adherence remain a main obstacle to reduce burden of health-damaging behavior at the population level, especially among the socially disadvantaged.(49–51)

Existing evidence has several important implications. First, four major types of behavioral interventions incorporated different aspects of the Social Ecological Model and some target at two or more levels simultaneously. Second, there is a lack of consensus on existing intervention effectiveness targeting modifiable behavioral risk factors. Third, given the complexity of existing intervention designs and population characteristics, a comprehensive intervention effectiveness evaluation framework is needed.

Using a Systems Science Approach to Evaluate Behavioral Intervention Effectiveness

An intervention targeting modifiable behavioral risk factors can be considered effective if it is able to address the following aspects: whether it will reach the targeted population as desired; and, once accessed, whether initiated behavioral change will be sustained to ensure long-term effectiveness. Two important, but often overlooked factors that affect both aspects of intervention effectiveness are social network and individual behavioral choices. Extensive studies and theories suggest that individual health behaviors are correlated within a social network.(52–54) Such association might be explained through mechanisms such as homophily, perceived social norms and peer pressure.(55),(56),(57) Social network structural characteristics including network density, clustering, average path length and cohesion play a significant role in shaping the mechanisms mentioned above as well as facilitating intervention diffusion.(58)

However, utilizing social network alone might not be sufficient for a behavioral intervention to be effective. Individual behavioral choices need to be taken into consideration simultaneously. Unlike transmission patterns of communicable diseases on a social network, individual health behaviors might not be changed immediately or ever after exposure to an intervention or a network connection's health behavior.(59),(60) In addition, game theoretical models have shown that highly connected network structures can impede the diffusion of behaviors.(61),(62) These aspects are especially crucial to interventions given the presence of CVD disparities in the population. Residential segregation and clustering is associated with higher prevalence of CVD.(63),(64) As a result, target populations of community-based behavioral interventions to address CVD disparities often reside in areas of high prevalence of CVD with highly clustered network structure. Under such circumstances, certain network-based intervention designs might not be as effective for this particular target population due to hindered intervention diffusion and high payoff for new behavior adoption, given high levels of social inertia in a segregated community.(65) To conclude, social network and individual behavioral choices need to be taken into consideration simultaneously when selecting the intervention target population and determining whether an intervention would be effective in the long term.(66),(67)

Nevertheless, a comprehensive evaluation of intervention effectiveness incorporating both aspects can be challenging using traditional statistical methods. In addition, the dynamic interplay between information diffusion and individual behavioral choices within a network can lead to emergence of unforeseen patterns of collective behavior,(68) which is important to consider when assessing the optimal population-level impact of an intervention under limited resources. Therefore, a systems science approach is needed to evaluate intervention effectiveness on behavioral risk factors, taking into consideration the nonlinear dynamic interactions of inter-dependent factors and emerging collective behavioral patterns on a social network.

A systems science approach allows for an exploration of dynamic interactions among multilevel interdependent micro-level factors that collectively affect the intervention effectiveness on the population scale.(69),(70),(71) One systems modeling approach—agent-based modeling (ABM) —allows for an investigation of the population level impact of an intervention in the presence of nonlinear dynamic interactions of multi-level interdependent factors and heterogeneous individuals' behaviors on a social network. It has been used extensively to understand policy implications, social/economic theories and collective behaviors.(70),(72),(73) Different from other equation-based dynamic models such as compartmental model and Markov model, the bottom-up model building process of ABM allows for relaxation in unrealistic assumptions such as population and decision-making process homogeneity.(74) Moreover, it can function as a bridge between theoretical concepts and empirical research.(75) It is almost impossible to have complete information of interest in empirical research to address the research question of interest most of the time. In the context of this dissertation, to identify the optimal behavioral intervention design given target population characteristics, it would be essential to incorporate data or hypothesized mechanisms of individuals' behavioral incentives when evaluating how effective an intervention is in term of sustaining individual behavioral change. However, collection of such information is near impossible for all population of interest due to time, political and budget constraints. Therefore, ABM would allow us to incorporate behavioral economics theories and individual-level mechanistic hypotheses to help address the research question of interest, given limited empirical information. Furthermore, results from such model will help identify which information is key to intervention effectiveness evaluation, thus motivating future empirical research design and data collection.

Dissertation Aims

Aim 1. To characterize the trajectories of cigarette smoking and physical activity from early adolescence to adulthood, using the National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave I through V survey data.

Aim 2. To assess the association between social network characteristics and cigarette smoking/physical activity during adolescence and the transition period from adolescence to young adulthood, using the National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave I and III survey data.

Aim 3. To develop an agent-based model evaluating the effectiveness of network-based and individual-targeting behavioral interventions.

Public Health Significance

Evidence on the effectiveness of interventions targeting modifiable health behaviors for CVD prevention are mixed and, to date, no existing study comprehensively evaluated the potential effectiveness of different types of behavioral interventions. Different from existing studies, this dissertation aims to use a systems science approach to comprehensively evaluate the effectiveness of two major types of behavioral interventions – network-based intervention (e.g. popular opinion leader intervention) and individual-targeting intervention (e.g. incentivizing healthy behaviors), incorporating theories of individual behavioral choices and intervention diffusion on a social network. Through a combination of observational studies and computational experiments, this dissertation will address the following questions: 1). What are the network-level and individual-level contributing factors to individual's health behaviors throughout life-course; 2) Given the knowledge of community network structure, target population characteristics and established community behavioral norms, what type of intervention and target population selection strategy would be most effective in maximizing the prevalence of health-benefiting behaviors in the long run; 3) Given specific target population's characteristics, what type of data would be needed in the future behavioral intervention design and effectiveness evaluation? This dissertation will be the first to summarize major behavioral intervention strategies into computational model designs with pre-specified target population selection and model set-up to alter individual incentives. Also, different from prior research(76–78) that centered around the influence of *direct* network connections (peer influence) on individual's health behavior, this dissertation will evaluate the population-level impact of different intervention scenarios, conditioning upon both network-based diffusion processes and operationalized individual incentive as individual behavioral choice utility functions. Most importantly, this dissertation provides an opportunity to compare and contrast the empirical and systems science modeling approach for identifying causal mechanisms linking individual and network level factors to population level outcomes of complex behavioral interventions. Findings from this dissertation will motivate future behavioral intervention trial design and data collection to address CVD burden and disparities, in terms of selection of target population for intervention and of intervention design (population-based vs individual-based), to reduce the prevalence of health damaging behaviors in the long term.

Chapter 2: Characterization of Trajectories of Physical Activity and Cigarette Smoking from Early Adolescence to Adulthood

Introduction

Cigarette smoking and physical inactivity (PA) have long been identified as two critical behavioral risk factors for cardiometabolic diseases and all-cause mortality.(79) As two modifiable health behaviors, despite their distinct nature, PA and cigarette smoking have been shown to be correlated.(80–83) From a theoretic perspective, Social Cognitive Theory(84),(85),(86) and the Social Ecological Model(87–89) have been the two most widely applied theories in explaining patterns associated with both behaviors. Based on these two theories, PA and cigarette smoking could be correlated due to overlapping cognitive factors (e.g. belief), behavioral factors (e.g. self-efficacy) and environmental factors (e.g. social support) associated with behavior initiation, as well as maintenance. Additionally, empirical studies have shown that physical activity might help thwart craving of cigarette smoking(90) and cigarette smoking may decrease individual's physical activity level due to impaired respiratory function.(91),(92) Nevertheless, results from intervention trials targeting PA and cigarette smoking simultaneously have been mixed.(88),(93),(94),(95) Therefore, it is important to compare and contrast patterns of these two health behaviors simultaneously in order to design interventions that might be effective in modifying both as well as evaluating why certain interventions fail at targeting at one behavior versus the other.

To better intervene upon these two modifiable health behaviors, exploratory studies have been put forth in recent years to identify long-term patterns of cigarette smoking and physical activity behaviors respectively in heterogenous population.(96–105) However, amongst all, the majority of such trajectory analyses were based on estimating trajectories of population mean over time assuming population homogeneity. A small number of studies based on a group-based trajectory modeling approach have used latent class growth analysis (LCGA), which pre-assumes a number of distinct sub-groups of population sharing similar trajectories of behaviors over time given no individual-level heterogeneity.(106) These approaches often fail to capture distinct behavior patterns, taking into consideration individual heterogeneity, which is vital to the design and evaluation of person-centered behavioral interventions. Latent Class (Growth) Mixture Models (LCMM) allow for exploration of number and characteristics of unobserved sub-group population that share similar behavior patterns, incorporating individual-level random effects.(106),(107) Using the National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave I through V survey data, this study aims to explore distinct sub-groups of population sharing similar patterns of PA and cigarette smoking from adolescence to adulthood, as well as predictors of sub-group membership. Results from this study will be able to provide further insights into features of PA and cigarette smoking as two modifiable health behaviors to better inform future person-centered behavioral interventions targeting each or both.

Methods

Study Design

The Add Health study is a longitudinal cohort study that enrolled a nationally representative sample of adolescents in the United States between grades 7 and 12 at baseline.(108) It was originally designed to facilitate a multidisciplinary approach to better understand causes of adolescent health behavior and outcomes throughout multiple developmental phases. At baseline (Wave I, 1993-1994), 20745 participants completed an in-school interview or at-home interview with a mean age of 15 years old. In addition, participants' parents were invited to complete interviews regarding parental socio-demographic background and household-level socio-economic information. Four additional waves of data were collected subsequently: Wave II (1995-1996) N=14,738, mean age = 16 yrs old; Wave III (2001-2002) N=

15,197, mean age = 22 yrs old; Wave IV (2008) N =15,701, mean age = 28 yrs old; Wave V (2016-2018) N = 19,828, mean age = 36 yrs old. Across all five waves, the following information was collected: participants' socio-demographic information, school performance, peer relationship, biomarker information, health outcomes, health behaviors, romantic relationship, familial and neighborhood-level socio-environmental contextual information, and geospatial information. The present analysis utilizes the in-school questionnaire, parental interview questionnaire and in-person interview questionnaire of the Add Health study. The use of the data was reviewed and approved by the Institutional Review Board at Emory University and the Add Health study review boards.

Cigarette Smoking. Survey respondents were asked to self-report cigarette smoking behaviors during in-school and in-home interviews. Questions regarding life-time history of cigarette smoking and past 30-day (p30-day) cigarette smoking behavior were asked. In Wave I and II, the following questions were asked to determine respondents' current smoking status and p30-day cigarette smoking intensity: 1) Have you ever tried cigarette smoking, even just 1 or 2 puffs? 2) Have you ever smoked cigarettes regularly, that is, at least 1 cigarette every day for 30 days? 3) During the past 30 days, on how many days did you smoke cigarettes? 4) During the past 30 days, on the days you smoked, on average, how many cigarettes per day did you smoke? In Wave III through V, the following questions were asked: 1) Have you ever tried cigarette smoking, even just one or two puffs? 2) Have you ever smoked an entire cigarette? 3) Have you smoked at all in the past 30 days? 4) During the past 30 days, on how many days did you smoke cigarettes? 5). During the past 30 days, on the days you smoked, on average, how many cigarettes per day did you smoke? Based on these sets of questions, respondents were categorized as current smoker and current non-smoker. Current smokers were defined as those that have tried cigarettes and smoked cigarettes in the past 30 days. P30-day cigarette smoking intensity was defined as total number of cigarettes smoked in the past 30 days. If respondent was categorized as current non-smoker, p30-day cigarette smoking intensity was zero. Otherwise, p30-day cigarette smoking intensity was calculated as the product of number of days smoked in the past 30 days and number of cigarettes smoked on average on the days respondents smoked. In addition, whether smokers were present in the household during baseline visit was reported as a binary response. During Wave V, respondents' electronic cigarette use was also indicated as a dichotomous variable.

Physical Activity. Study respondents were asked to self-report their weekly frequency (times per week) of a series of standard physical activities including: jogging, walking, karate, jumping rope, gymnastics, dancing, roller-blading, roller-skating, skate-boarding, bicycling, or active sports. Previous studies(109),(110) have frequently used the definition of moderate-vigorous leisure-time physical activity through approximating number of metabolic equivalents. In this study, instead of using number of metabolic equivalents approximated, we generated a physical activity score corresponding to self-reported physical activity frequency of each questionnaire item to account for change in reported activity categorization in Wave V. If frequency was zero in the past seven days, then the score was assigned as zero. If frequency was either once or twice in the past seven days, then the score was assigned as 1.5. Otherwise a score of 3.5 was assigned. A summary physical activity score was generated by summing up physical activity scores across all questionnaire items at each wave. Additionally, a standardized physical activity score across all five waves was generated by dividing the summary score by number of activities included in each wave's questionnaire to account for an increased number of activities included in questionnaires starting Wave III. Detailed questionnaires corresponding to tobacco use behavior and physical activity are available on https://www.cpc.unc.edu/projects/addhealth/documentation/publicdata.

Other Variables of Interest

Socio-demographic Characteristics. Socio-demographic variables of interest included biological sex, race/ethnicity, parental education and household income reported at baseline visit. Survey respondents self-identified as White, African American, Hispanic, Asian/Pacific Islander/Native American/Alaska Indians, or Others. 83% (N = 17238) respondents' parents participated in the baseline parental interview questionnaire in 1994. Highest level of parental education obtained by 1994 was reported. Respondents' parents were further dichotomized as having received a degree no more than high school or having received a degree beyond high school. In addition, total household pre-tax income including welfare benefits, dividends, and others was reported. A three-level ordinal variable was generated based on tertiles of reported household income.

*Baseline Neighborhood Characteristics***.** Respondents' closeness with people in the neighborhood was captured by a survey question asking whether they knew most people in the neighborhood. In addition, all respondents to the in-home interview were asked about whether they were happy with the present neighborhood, whether they felt safe in the current neighborhood and whether they had access to a fitness or recreational center in the neighborhood.

Socio-psychological Factors. Perceived parental, peer and teachers' support was captured during baseline in-home interviews through questions on whether respondents felt cared for by adults, teachers, and friends. Whether respondents perceived as part of the school or close to others at school were also recorded as a binary response in in-school questionnaires at baseline. Detailed baseline in-home and in-school questionnaires are available on https://www.cpc.unc.edu/projects/addhealth/documentation/publicdata.

Statistical Analysis

Participants who participated in Wave I through Wave V in-school and in-home interview as well as baseline parental interview questionnaire of the Add Health study with non-missing information on age, PA and cigarette smoking behaviors were included in the analyses. To ensure participants of five waves of the Add Health study were comparable, key socio-demographic characteristics of all waves of study participants were assessed. (**Table 2-1**) To identify sub-groups of physical activity and cigarette smoking trajectories from young adolescence to adulthood, latent class mixture models (LCMM) were used. LCMM allows for exploration of population-level outcome heterogeneity by identifying the underlying N number of latent classes and accounting for individual-level measure heterogeneity.(111),(112)

To determine the optimal number of latent classes and class-specific trajectories of physical activity scores from young adolescence to adulthood for Add Health participants, we fit LCMMs with standardized physical activity score as the outcome measure and age as a continuous time variable. Maximum likelihood measures of a single latent class model were used as the initial values for model estimation. For each model with hypothesized number of latent classes, model fitting and estimation process were iterated over random vectors of initial values through an automatic grid search algorithm until model achieved the best log-likelihood measure. Moreover, quadratic trajectories of physical activity scores were explored in addition to linear trajectories. Based on prior literature(110,113,114), we hypothesized that there were three classes of distinct trajectories. Hence, all model fitting and estimation procedures were iterated over two to four hypothesized number of latent classes in LCMM. Posterior probabilities of participants belonging to a class, given the hypothesized number of classes were obtained. Optimal number of classes for physical activity score trajectories was determined based on the following six factors: model entropy, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), trajectory shape fitting between predicted and observed data, average posterior probability of individuals belonging to the assigned class (ideally greater or equal to 0.7), and proportion of individuals belonging to each class. With respect to trajectories of p30-day cigarette smoking intensity, a similar approach as described above was used. We hypothesized that three classes of trajectories would be observed.(105) Log of p30-day cigarette smoking intensity was used as the outcome measure for model fitting purposes. To further identify predictors of latent class membership, multinomial logistic regression was used to assess the association of individual-level predictors (e.g. socio-demographic, psychological factors) and community-level predictors (e.g. household and neighborhood factors) with trajectory class membership for both PA and cigarette smoking. All statistical analyses were performed in R 3.5.2.

Results

Of 20745 baseline study participants, 14736 (71%) participated in Wave II, 15197 (73%) participated in Wave III, 15701 (76%) participated in Wave IV, and 12283 (59%) participated in Wave V. Across all five waves, prevalence of female participants and participants of different race/ethnicity were comparable. Similarly, participants across all waves reported similar levels of parental education levels. (**Table 2-1**) Of all baseline study participants, 20734 had completed at least one set of physical activity-related questions across five waves and 20689 had completed at least one set of cigarette smoking-related questions across fives waves. Amongst respondents included in PA trajectory analyses, 10474 were female and 10292 were non-white participants. With respect to cigarette smoking, 10455 respondents included in the final analyses were female and 10308 were non-white participants. (**Table 2-2**, **Table 2-3**)

Trajectory Classes and Class Member Profile for Standardized Physical Activity Score

We identified three distinct sub-groups of PA trajectories in the study population: moderately active group (Class 1, N= 1067, 5%), persistently inactive group (Class 2, N= 14257, 69%) and progressing inactive group (Class 3, N= 5410, 26%) since three classes resulted in model with the highest entropy, smallest AIC/BIC as well as mean posterior probability of individual actually belonging to each class greater than 0.70. (**Table 2-2**, **Supplemental Table 2-1**, and **Figure 2-1**). Moderately active group maintained a moderate PA level till 30 years old, when PA level dropped. Persistently inactive group had the lowest PA level across all groups over time. Meanwhile, the magnitude of change in PA level overtime has been the smallest in this group. The progressing inactive group had the highest PA level prior to 15 years old. Nonetheless, the PA level dropped drastically starting at 15 years old and leveled off starting at 25 years old. The magnitude of change in physical activity level was the biggest amongst this group. Overall, prior to 18 years old, progressing inactive group had the highest mean PA level whereas moderately active group became the most active group amongst all groups starting at 18 years old (**Figure 2-1**). Socio-demographic characteristics were comparable across all three groups. The persistently inactive group had the highest proportion of females (N = 8324, 58%), parents that did not receive a high school degree or above (N = 6167, 43%), and households with an income in the lowest tertile (N = 4104, 29%) whereas the moderately active group had the lowest proportion of females (N= 303, 28%), lowest number of parents that did not receive a high school degree or above (N = 388, 36%) and lowest number of households with an income in the lowest tertile (N = 238, 22%). With respect to baseline neighborhood characteristics, the persistently inactive group had the lowest proportion of respondents that reported being happy with their present neighborhood (N=12754, 89%), had an access to recreational center in the neighborhood (N = 2360, 17%), felt safe in the neighborhood (N = 12382, 87%), or knew almost everyone in the community (N=9577, 67%). Perceived closeness with peers at school and perceiving as a part of the school appeared to be two differentiating factors between the persistently inactive group and the other two groups. Proportions of individuals that perceived as close with peers at school and perceived as part of the school were lowest in the persistently inactive group. (**Table 2-2**)

Trajectory Classes and Class Member Profile for Past 30-day Cigarette Smoking Intensity

With respect to past 30-day cigarette smoking intensity, we observed three distinct groups in the study population: persistent non-smoker (Class 1, N=14939, 72%), progressing non-smoker (Class 2, N=2357, 11%), and progressing smoker (Class 3, N=3393, 16%) since three classes resulted in model with the highest entropy, smallest AIC/BIC as well as mean posterior probability of individual actually belonging to each class greater than 0.70. (**Table 2-3**, **Supplemental Table 2-1**). The progressing non-smoker group increased p30-day cigarette smoking intensity prior to 18 years old and started reducing p30-day cigarette smoking intensity throughout adulthood. The persistent non-smoker group remained as non-smoker throughout the entire study follow-up period. In the meantime, the progressing smoker group increased p30-day cigarette smoking intensity from adolescence to adulthood consistently. However, the magnitude of increase was the highest from adolescence to young adulthood. Rate of increase in p30-day cigarette smoking intensity lowered starting at 22 years old and plateaued around 26 years old till the end of the study follow-up. Overall, the progressing non-smoker group had the highest mean log (p30-day cigarette smoking intensity) prior to 23 years old. After 23 years old, p30-day cigarette smoking intensity amongst the progressing smoker group became the highest. (**Figure 2-2**) Among the three groups, the progressing smoker group had the lowest proportion of females (N = 1474, 43%), meanwhile the persistent non-smoker group had the highest. (N = 7800, 52%) The progressing non-smoker group had the lowest proportion of racially disadvantaged population. (N = 623, 26%) Also, the progressing non-smoker group had the most individuals indicating the presence of smokers in the household (N = 1435, 61%) whereas persistent non-smoker had the fewest. (N = 4822, 32%). Interestingly, the persistent non-smoker group had the fewest respondents indicating knowing most people in the neighborhood (N = 10187, 68%), whereas the progressing smoker group had the most. (N = 2611, 77%) With respect to baseline socio-psychological factors, perceived closeness with peers at school, perceived care from teachers, and perceived as part of the school appeared to be three differentiating factors. The persistent non-smoker group had the highest proportions of individuals that responded positively to all questions above whereas progressing non-smoker group had the lowest proportions. (**Table 2-3**)

Predictors of Physical Activity and P30-day Cigarette Smoking Intensity Class Membership

Out of the entire sample population, about 50% individuals were concurrently persistently inactive and persistent non-smokers. Across all three classes of PA trajectory, highest proportion of individuals were consistently persistent non-smokers whereas fewest individuals were progressing non-smoker. (**Table 2-4**) For cigarette smoking, highest proportion of individuals belonged to the persistently inactive group and fewest number of individuals belonged to the moderately active group. (**Table 2-4**) Based on results from multinomial logistic regression analyses, sex, baseline parental education, baseline parental income, access to fitness center, knowing almost everyone in the neighborhood, feeling close to others at school and feeling as part of the school were significant predictors of being in persistently inactive group as compared to being in moderately active group. Sex, race/ethnicity, and baseline parental income were significant predictors of being in the progressing inactive group as compared to being in the moderately active group. (**Table 2-5**) Overall, females were more likely to be in the persistently inactive and progressing inactive group. (RRR = 3.59, 95% CI: 2.97-4.33; RRR = 1.47, 95% CI: 1.20-1.78) Non-Hispanic white individuals were more likely to be in the moderately active group. Baseline neighborhood characteristics only differentiated individuals in persistently inactive group versus moderately active group. In addition, not feeling as part of school was a significant predictor of an individual being persistently inactive. (RRR= 0.72, 95% CI: 0.53-0.96)

With respect to cigarette smoking, females were less likely to be a progressing smoker as compared to being a progressing smoker. (RRR = 0.74, 95% CI: 0.64, 0.86) Race was a significant predictor of individuals being a progressing non-smoker as compared to the other two smoking trajectory groups. Non-Hispanic black individuals were more likely to be persistent non-smokers (RRR = 7.32, 95% CI: 5.54, 9.66) and progressing smokers (RRR = 4.86, 95% CI: 3.61, 6.53). Non-Hispanic white individuals were more likely to be xxx as compared to all other racial groups. (**Table 2-6**) Individuals that reported feeling happy about present neighborhood, perceived care from teachers at baseline visit, and feeling as part of the school were significantly more likely to be persistent non-smokers. Meanwhile, presence of smokers in the household was associated with higher risk of being progressing non-smoker or progressing smoker as compared to persistent non-smoker. Interestingly, perceiving care from teachers, feeling close to others and feeling as part of the school were associated with higher risk of being a progressing smoker as compared to be a progressing non-smoker. Nonetheless, the magnitude of association is bigger between persistent non-smokers and progressing non-smoker as compared to progressing smoker and progressing non-smoker. (**Table 2-6**)

Discussion

This study explored sub-groups of individuals sharing similar trajectories of physical activity and past 30-day cigarette smoking behavior from adolescence to adulthood, as well as predictors of specific sub-group membership. Our study revealed three distinct groups of individuals following similar patterns of physical activity (moderately active, persistently inactive, progressing inactive) and three distinct groups of individuals following similar patterns of cigarette smoking behavior in the past 30 days (persistent non-smoker, progressing non-smoker, progressing smoker). In general, physical activity level decreases from adolescence to adulthood. However, cigarette smoking behavioral patterns differ significantly across the three groups from adolescence to adulthood. Interestingly, for both physical activity and cigarette smoking behavior, there is one group of individuals that had a consistent behavioral pattern throughout the entire study follow-up and the population size of these groups were the highest. Additionally, transition from adolescence to young adulthood and late adulthood both appeared to be critical to altering individuals' physical activity patterns whereas transition from young adulthood to late adulthood might be a critical time window for change in cigarette smoking behavior. These findings are consistent with several earlier studies that also found 3 to 4 sub-groups of trajectories for both PA and cigarette smoking.(105,110,115),(116),(117) Similar to our study, one existing study on life-course trajectories of PA based on a sample of Finnish population have shown that overall PA level declines overtime and persistently low-activity group makes up the large proportion of the study population. Major changes in PA level also started during the transition period between adolescence to young adulthood around 21 years old.(117) Contrastingly, one large study in a sample of US population described 10 sub-groups of trajectories, among which there were additional groups of individuals that remained persistently at high levels of PA, as well as increasingly levels of PA from adolescence to late adulthood.(118) For cigarette smoking, very few studies explored the long-term trajectories of cigarette smoking behavior spanning across adolescence to adulthood.(105) One recent study based on a Northern Finnish study population with a 46-year follow-up characterized six different groups of individuals sharing similar patterns of behavior over time. However, the overall patterns of these six trajectories are similar to findings from our study, indicating the presence of progressing smoker, never smoker and progressing non-smoker (labeled as quitters in the study).(119)

Our study showed that sex and race/ethnicity are significant socio-demographic predictors of long-term trajectories of PA level and cigarette smoking. Meanwhile, baseline parental education and parental income have more impact over trajectory class membership of PA and not so much on cigarette smoking intensity trajectory class membership. The observed sex difference in PA patterns overtime has long been in discussion. Theories and studies have suggested that females are more likely to be more inactive as compared to their male counterparts, potentially due to long-established gender norms, in addition to physiological differences.(120),(121) Consistent with previous studies, our study has indicated that females of disadvantaged racial profile as well as lower familiar socio-economic status are more likely to be inactive over their lifetime.(113) Combined with our additional analyses on association between household/neighborhood level factors and PA trajectory, adjusting for socio-demographic predictors mentioned above, it appeared that socio-demographic factors are key predictors of PA trajectory and such observation could be partially explained through downstream household and neighborhood level factors that impact one's access to physical activity, for example, no access to recreational center and safety issue in a low socio-economic status neighborhood. In terms of socio-psychological predictors of PA trajectory, our study found that perceived closeness with peers as well as perceiving as part of the school at baseline was associated with a decreased risk of being persistently inactive throughout life course, even after adjusting for critical socio-demographic factors. Significant numbers of empirical research have shown that peer influence plays a role in moderating physical activity behaviors.(122–124) In addition to existing findings, our study has shown the possibility of lasting impact of peer influence and perceived normative behavior at younger age over lifelong trajectories of physical activity.

For cigarette smoking, our study showed that females are more likely to be progressing non-smoker as compared to progressing smoker, which is contrasting to existing study indicating that it might be more difficult for females to quit smoking once started.(125),(126) Consistent with existing studies, results from our study showed that racially disadvantaged individuals are more likely to be progressing smokers versus progressing non-smokers, which suggests that racial minorities, once initiated cigarette smoking, are less likely to quit in the long term.(127) Interestingly, in our study, racially disadvantaged individuals are more likely to be persistent non-smokers than their White counterparts. Previous studies have also echoed such findings. Various studies(128–130) have shown that African American, Hispanic and American Indian individuals, despite their age of cigarette smoking initiation, are lighter and intermittent smokers as compared to Whites. Different from PA trajectories, we found that household and neighborhood level predictors are important to differentiate persistent non-smokers as compared to progressing smokers/non-smokers even after adjusting for individual socio-demographic factors, suggesting the importance of contextual exposure to cigarette smoking behavior. However, the presence of a smoker in the household is associated with higher risk of being a progressing smoker and progressing non-smoker as compared to never smoker. This finding is similar to a previous study that household smoking is not linearly and positively correlated with cigarette smoking or quitting in the long-term.(131) With regards to socio-psychological predictors, different from PA trajectory membership, we did not find an association between baseline socio-psychological factors and long-term trajectories of cigarette smoking intensity.

Results from this study have several important indications for behavioral intervention design targeting physical activity and cigarette smoking. First, interventions are still much needed for both behaviors given the large proportions of individuals that are persistently inactive and progressing non-smokers in our study population. Second, when designing interventions targeting PA, individual socio-demographic and socio-psychological factors might be important to take into consideration, especially when considering the sex difference in motivation to engage in PA. One study has shown that for females, individual-based interventions such as positive messaging and campaigns have been more effective in the long run than changes in built environment.(120) For cigarette smoking, however, individual socio-demographic and contextual factors such as household and neighborhood level factors are key. Third, with regards to potential target population characteristics, our analyses indicate that individuals with more significant change of PA behavior are more likely to male, coming from a better socio-economic background, and having more perceived care/support from peers at a younger age whereas females, and individuals from disadvantaged familial background with less perceived social/psychological support are more reluctant to change behaviors over time. Meanwhile, for cigarette smoking, males, non-Hispanic White individuals, as well as those perceiving support from neighborhood/peers are more likely to change their cigarette smoking behaviors.

Collectively, these findings showed that different types of behavioral interventions might be needed when targeting PA versus cigarette smoking. In addition, despite a dire need to address disparity in these two behaviors due to inequality and inequity, a careful examination of intervention design prior to implementation is needed due to greater inertia to behavioral change amongst disadvantaged population.

Our study has several strengths. First, to our knowledge, it is one of the first studies that characterized trajectories of physical activity and cigarette smoking behavior from adolescence to adulthood using a comprehensive large nationally presentative longitudinal study. Second, utilizing latent class growth mixture model, our study was able to identify specific trajectories and sub-group population, taking into consideration both group and individual level heterogeneity. Third, a comprehensive exploration of predictors associated with trajectories allowed for investigation of important factors associated with sub-group membership, which is crucial to identify behavioral intervention target population characteristics as well as intervention design strategies. In the meantime, several limitations need to be acknowledged. First, the number of Add Health study participants decreased by Wave V. Approximately 50% of the study participants were part of Wave V of the study. Even though the socio-demographic characteristics of study participants were comparable across all five waves (**Table 2-1**), individuals that were not part of the study during later waves of study might lead to missingness in outcome data that are not at random, which might lead to bias our study finding. Second, both physical activity and past 30-day cigarette smoking data were obtained through self-reported survey. However, the design of the questionnaire items regarding to those two behaviors were not consistent across all waves, which might lead to measurement error of outcomes. Third, Wave I through III did not have questions on average cigarette smoking intensity on a typical smoking day instead of the past 30-day daily cigarette smoking intensity. Therefore, this study used past 30-day cigarette smoking intensity as an outcome, which might not be representative of all cigarette smokers' typical smoking behavior annually. Lastly, latent class (growth) mixture model is a post-hoc analytical approach that is constrained by parameters imposed on model specification, such as hypothesized number of groups, as well as whether the trajectory would follow a linear, quadratic or cubic pattern. Even though our study explored different model specification, similar research question needs to be explored in other studies to further confirm the research findings.

Conclusion

To conclude, our study indicates that there are three sub-groups of individuals sharing similar patterns of physical activity and past 30-day cigarette smoking behavior from early adolescence to adulthood. In addition, age, socio-demographic and psychological factors are all important predictors of these two behaviors. Future behavioral interventions targeting physical activity and cigarette smoking behaviors need to take into consideration both timing and target population characteristics to be effective.

Table 2-1. Characteristics of the Add Health Study Participants across Five Waves of Study Follow-up, 1994 – 2018

	Wave I	**Wave II**	**Wave III**	**Wave IV**	**Wave V**
N					
Total	20745	14736	15197	15701	12283
Parental Education					
College or above	6941 (33)	5129 (35)	5278 (35)	5372 (34)	4457 (36)
Graduated from high school but not college	4192 (20)	2938 (20)	3088 (20)	3269 (21)	2565 (21)
Not graduated from high school	8572 (41)	6070 (41)	6153 (41)	6386 (41)	4764 (39)
Race/Ethnicity					
Non-Hispanic White	10455 (50)	7573 (51)	7864 (52)	8294 (53)	6842 (56)
Non-Hispanic Black	4669 (23)	3244 (22)	3316 (22)	3498 (22)	2473 (20)
Hispanic	3525 (17)	2487 (17)	2447 (16)	2498 (16)	1825 (15)
Non-Hispanic Asian/Pacific Islander/American Indian/Alaska Native	1467 (7)	1004 (7)	1108 (7)	947 (6)	793 (6)
Other	629 (3)	428 (3)	435 (3)	437 (3)	328 (3)
Female	10263 (49)	7182 (49)	7167 (47)	7349 (47)	5324 (43)
Current Smoker	5326 (26)	4648 (32)	4786 (32)	5508 (35)	2984 (24)
Mean (SD)					
Age (years)	15.7 (1.7)	16.2 (1.6)	22.0 (1.8)	28.5 (1.8)	37.5 (1.9)
Standardized physical activity score	0.4 (0.2)	0.4 (0.2)	0.2 (0.2)	0.2 (0.2)	0.1 (0.1)
**SD: standard deviation*					

Table 2-2. Baseline Class Member Profile of Physical Activity Trajectory

	Physical Activity Trajectory Class Profile		
	Class 1 (Moderately active)	**Class 2 (Persistently inactive)**	**Class 3 (Progressing inactive)**
N (%)	1067 (5)	14257 (69)	5410 (26)
Socio-demographic			
Female	303 (28)	8324 (58)	1847 (34)
Non-Hispanic White	503 (47)	7059 (50)	2891 (53)
Non-Hispanic Black	247 (23)	3348 (23)	1072 (20)
Hispanic	181 (17)	2447 (17)	893 (17)
Other	136 (12)	1403 (10)	554 (10)
Parental Education (Less than High school)	388 (36)	6167 (43)	2016 (37)
Household Income (lowest tertile)	238 (22)	4104 (29)	1324 (24)
Neighborhood			
Knowing most people in the neighborhood	814 (76)	9577 (67)	4089 (76)
Happy with present neighborhood	986 (92)	12754 (89)	4968 (92)
Access to recreational center in neighborhood	302 (28)	2360 (17)	1548 (29)
Feel safe in the neighborhood	968 (91)	12382 (87)	4831 (89)
Socio-psychological			
Perceived adults' care	1022 (96)	13657 (96)	5218 (96)
Perceived teachers' care	924 (87)	12111 (85)	4695 (87)
Perceived friends' care	1040 (97)	13733 (96)	5236 (97)
Perceived closeness with peers at school	467 (44)	5168 (26)	2404 (44)
Perceived as part of school	457 (43)	5101 (36)	2357 (44)

Table 2-3. Baseline Class Member Profile of Past 30-day Cigarette Smoking Intensity Trajectory

	Past 30-day Cigarette Smoking Trajectory Class Profile		
	Persistent Non-smoker	**Progressing Non-smoker**	**Progressing Smoker**
N (%)	14939 (72)	2357 (11)	3393 (16)
Socio-demographic			
Female	7800 (52)	1181 (50)	1474 (43)
Non-Hispanic White	6691 (45)	1734 (74)	2012 (59)
Non-Hispanic Black	3781 (25)	128 (5)	742 (22)
Hispanic	2900 (19)	270 (11)	342 (10)
Other	1567(11)	225 (9)	297 (9)
Parental Education (Less than High school)	5899 (39)	1095 (46)	1556 (46)
Household Income (lowest tertile)	3951 (26)	641 (27)	1058 (31)
Presence of smoker in household	4822 (32)	1435 (61)	1732 (51)
Neighborhood			
Knowing most people in the neighborhood	10187 (68)	1660 (70)	2611 (77)
Happy with present neighborhood	13537 (91)	2093 (89)	3046 (90)
Socio-psychological			
Perceived adults' care	14353 (96)	2253 (96)	3254 (96)
Perceived teachers' care	13162 (88)	1765 (75)	2774 (82)
Perceived friends' care	14401 (96)	2302 (98)	3270 (96)
Perceived closeness with peers at school	6026 (40)	769 (33)	1232 (36)
Perceived as part of school	6049 (40)	671 (28)	1184 (35)

Table 2-4. Class membership of physical activity trajectories conditioned on cigarette smoking trajectory class membership

	Cigarette Smoking Trajectory Class Membership			
Physical Activity Class	**Class 1 (Persistent non-smoker)**	**Class 2 (Progressing non-smoker)**	**Class 3 (Progressing Smoker)**	**Total**
Class 1 (Moderately active)	808 (4%)	99 (0.5%)	160 (1%)	**1067 (5%)**
Class 2 (Persistently inactive)	10082 (49%)	1777 (9%)	2358 (11%)	**14217 (69%)**
Class 3 (Progressing inactive)	4046 (20%)	481 (2%)	875 (4%)	**5402 (26%)**
Total	**14939 (72%)**	**2357 (11%)**	**3393 (16%)**	**20745**
*Percentage represent percentage out of the entire study population				

Table 2-5. Predictors of Physical Activity Trajectory Class Membership

	Class (Reference: Class 1: Moderately active)	Relative Risk Ratio (95% CI) *	Relative Risk Ratio (95% CI) **
Female	2 (Persistently inactive)	3.53 (3.08, 4.05)	3.59 (2.97, 4.33)
	3 (Progressing inactive)	1.31 (1.13, 1.51)	1.47 (1.20, 1.78)
Race (reference = non-Hispanic White)			
Non-Hispanic Black	2 (Persistently inactive)	0.95 (0.81, 1.12)	0.84 (0.67, 1.04)
	3 (Progressing inactive)	0.75 (0.64, 0.89)	0.61 (0.48, 0.78)
Hispanic	2 (Persistently inactive)	0.98 (0.82, 1.17)	0.78 (0.60, 1.02)
	3 (Progressing inactive)	0.86 (0.72, 1.04)	0.75 (0.57, 0.99)
Others	2 (Persistently inactive)	0.76 (0.62, 0.93)	0.65 (0.49, 0.86)
	3 (Progressing inactive)	0.71 (0.58, 0.88)	0.63 (0.47, 0.85)
Parent graduated from high school or above			
	2 (Persistently inactive)	0.71 (0.62, 0.82)	0.70 (0.57, 0.85)
	3 (Progressing inactive)	0.95 (0.83, 1.09)	0.88 (0.72, 1.07)
Parental baseline income (reference = first tertile)			
25%+	2 (Persistently inactive)	0.70 (0.58, 0.83)	0.73 (0.58. 0.92)
	3 (Progressing inactive)	0.79 (0.66, 0.96)	0.78 (0.61, 0.99)
75%+	2 (Persistently inactive)	0.65 (0.54, 0.78)	0.78 (0.61, 0.99)
	3 (Progressing inactive)	0.87 (0.73, 1.05)	0.93 (0.73, 1.20)
Have access to fitness center in the community			
	2 (Persistently inactive)	0.54 (0.47, 0.63)	0.57 (0.47. 0.69)
	3 (Progressing inactive)	1.04 (0.90, 1.21)	1.13 (0.93, 1.38)
Feel safe in the neighborhood			
	2 (Persistently inactive)	0.71 (0.57, 0.88)	0.88 (0.63, 1.24)
	3 (Progressing inactive)	0.81 (0.64, 1.01)	0.95 (0.67, 1.35)
Knows everyone in the neighborhood			
	2 (Persistently inactive)	0.63 (0.55, 0.74)	0.67 (0.55, 0.82)
	3 (Progressing inactive)	0.94 (0.81, 1.10)	0.96 (0.77, 1.19)
Feel happy about present neighborhood			

	2 (Persistently inactive)	0.70 (0.55, 0.89)	0.83 (0.57, 1.20)
	3 (Progressing inactive)	0.86 (0.67, 1.11)	0.87 (0.58, 1.28)
Perceived care from adults			
	2 (Persistently inactive)	0.93 (0.66, 1.31)	1.09 (0.66, 1.86)
	3 (Progressing inactive)	1.10 (0.77, 1.58)	1.12 (0.64, 1.97)
Perceived care from teachers			
	2 (Persistently inactive)	0.83 (0.69, 1.01)	1.17 (0.89, 1.54)
	3 (Progressing inactive)	0.97 (0.80, 1.19)	1.17 (0.88, 1.57)
Perceived care from friends			
	2 (Persistently inactive)	0.67 (0.43, 1.04)	0.65 (0.31, 1.36)
	3 (Progressing inactive)	0.72 (0.46, 1.14)	0.72 (0.33, 1.54)
Feel close to others at school			
	2 (Persistently inactive)	0.55 (0.44, 0.70)	0.73 (0.53, 0.99)
	3 (Progressing inactive)	1.07 (0.84, 1.38)	1.15 (0.83, 1.59)
Feel as part of the school			
	2 (Persistently inactive)	0.62 (0.50, 0.77)	0.72 (0.53, 0.96)
	3 (Progressing inactive)	1.03 (0.82, 1.29)	0.85 (0.63, 1.16)

*: RRR corresponds to models adjusting for sex and race/ethnicity only: e.g. model for parental baseline income was obtained from model adjusting for parental baseline income, sex and race/ethnicity
**: RRR corresponds to model adjusting from all covariates of interest

Table 2-6. Predictors of P30-day Cigarette Smoking Intensity Trajectory Class Membership

	Class (Reference: Class 2: Progressing non-smoker)	Relative Risk Ratio (95% CI) *	Relative Risk Ratio (95% CI) **
Female	1 (Persistent Non-smoker)	1.08 (0.99, 1.18)	1.07 (0.94, 1.21)
	3 (Progressing Smoker)	0.75 (0.68, 0.84)	0.74 (0.64, 0.86)
Race (reference = White)			
Non-Hispanic Black	1 (Persistent Non-smoker)	7.64 (6.36, 9.19)	7.32 (5.54, 9.66)
	3 (Progressing Smoker)	5.02 (4.12, 6.12)	4.86 (3.61, 6.53)
Hispanic	1 (Persistent Non-smoker)	2.79 (2.43, 3.19)	2.21 (1.80, 2.71)
	3 (Progressing Smoker)	1.09 (0.91, 1.29)	0.97 (0.75, 1.25)
Others	1 (Persistent Non-smoker)	1.81 (1.56, 2.10)	1.70 (1.33, 2.17)
	3 (Progressing Smoker)	1.13 (0.94, 1.36)	1.43 (1.00, 1.32)
Parent graduated from high school or above			
	1 (Persistent Non-smoker)	1.58 (1.44, 1.73)	1.15 (1.00, 1.32)
	3 (Progressing Smoker)	1.04 (0.94, 1.17)	0.93 (0.79, 1.09)
Parental baseline income (reference = first tertile)			
25%+	1 (Persistent Non-smoker)	1.29 (1.14, 1.45)	0.98 (0.83, 1.16)
	3 (Progressing Smoker)	0.99 (0.86, 1.14)	0.89 (0.74, 1.08)
75%+	1 (Persistent Non-smoker)	1.52 (1.35, 1.72)	0.99 (0.83, 1.18)
	3 (Progressing Smoker)	0.93 (0.80, 1.07)	0.75 (0.61, 0.91)
Presence of smoker in the household			
	1 (Persistent Non-smoker)	0.25 (0.22, 0.28)	0.30 (0.27, 0.35)
	3 (Progressing Smoker)	0.58 (0.51, 0.65)	0.61 (0.52, 0.72)
Knows everyone in the neighborhood			
	1 (Persistent Non-smoker)	0.90 (0.82, 0.99)	0.83 (0.71, 0.96)
	3 (Progressing Smoker)	1.33 (1.18, 1.50)	1.22 (1.02, 1.46)
Feel happy about present neighborhood			
	1 (Persistent Non-smoker)	1.56 (1.35, 1.80)	1.40 (1.12, 1.76)
	3 (Progressing Smoker)	1.25 (1.05, 1.49)	1.28 (0.98, 1.67)
Perceived care from adults			
	1 (Persistent Non-smoker)	1.54 (1.22, 1.94)	1.11 (0.76, 1.62)
	3 (Progressing Smoker)	1.34 (1.02, 1.77)	1.14 (0.73, 1.77)
Perceived care from teachers			

	1 (Persistent Non-smoker)	2.84 (2.54, 3.18)	2.29 (1.92, 2.73)
	3 (Progressing Smoker)	1.61 (1.41, 1.85)	1.60 (1.30,1.96)
Perceived care from friends			
	1 (Persistent Non-smoker)	0.99 (0.72, 1.37)	0.75 (0.45, 1.26)
	3 (Progressing Smoker)	0.82 (0.57, 1.17)	0.69 (0.39, 1.23)
Feel close to others at school			
	1 (Persistent Non-smoker)	1.78 (1.54, 2.06)	1.03 (0.84, 1.27)
	3 (Progressing Smoker)	1.35 (1.14, 1.61)	1.01 (0.79, 1.28)
Feel as part of the school			
	1 (Persistent Non-smoker)	2.35 (2.05, 2.70)	1.82 (1.50, 2.22)
	3 (Progressing Smoker)	1.57 (1.34, 1.86)	1.32 (1.05, 1.65)

*: RRR corresponds to models adjusting for sex and race/ethnicity only: e.g. model for parental baseline income was obtained from model adjusting for parental baseline income, sex and race/ethnicity
**: RRR corresponds to model adjusting from all covariates of interest

Supplemental Table 2-1. Model Selection Criteria for Trajectories of Physical Activity Score and Cigarette Smoking Intensity from Early Adolescence to Adulthood

Physical Activity			
	AIC	**BIC**	**Entropy**
1-class model	-49912.7	-49833.3	1.00
2-class model	-49904.7	-49793.5	0.31
3-class model	-52101.1	-51958.2	0.60
4-class model	-51717.3	-51542.7	0.18
Optimal Model Class Membership			
	Class 1	**Class 2**	**Class 3**
N (%)	1067 (5)	14257 (69)	5410 (26)
Distribution of posterior probability of class membership	**1st Quartile**	**Mean**	
Class 1	0.62	0.78	
Class 2	0.77	0.85	
Class 3	0.60	0.72	
Log(Past 30-day cigarette smoking intensity)			
	AIC	**BIC**	**Entropy**
1-class model	304046.8	304126.1	1.00
2-class model	304054.8	304165.9	0.46
3-class model	284475.4	284618.3	0.90
4-class model	284483.4	284658.1	0.59
3-class Model Class Membership (Optimal Model)			
	Class 1	**Class 2**	**Class 3**
N (%)	14939 (72)	2357 (11)	3393 (16)
3-class model probability of class membership	**1st Quartile**	**Mean**	
Class 1	0.99	0.97	
Class 2	0.89	0.92	
Class 3	0.90	0.92	

Figure 2-1. Subject-specific Trajectories of Standardized Physical Activity Score from Early Adolescence to Adulthood

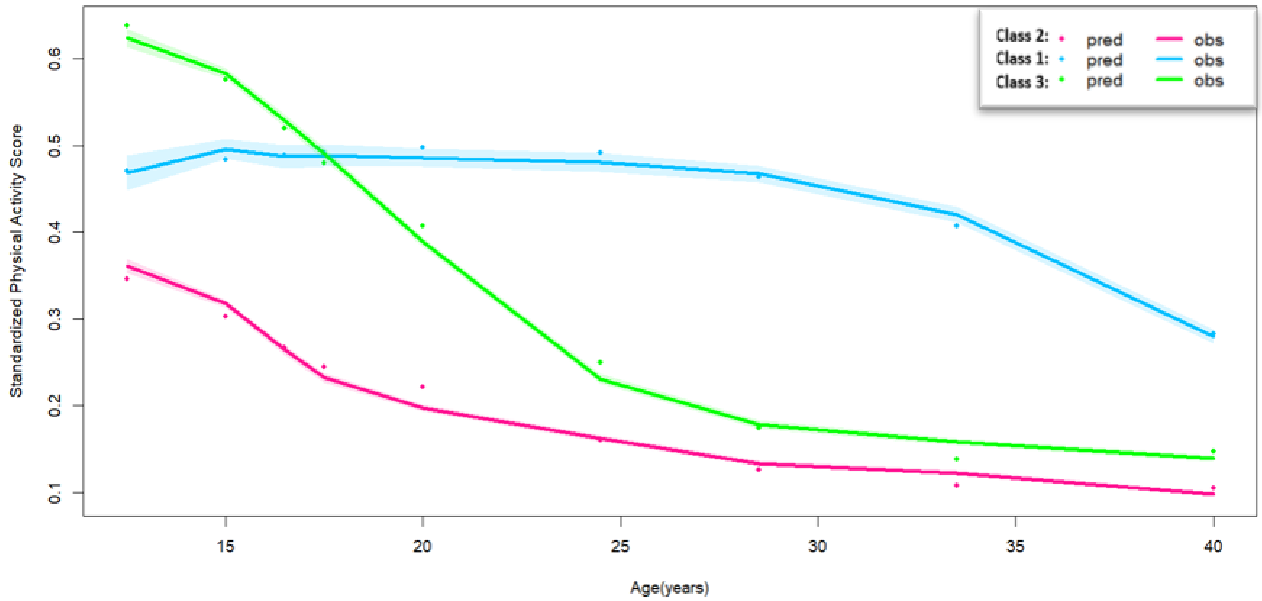
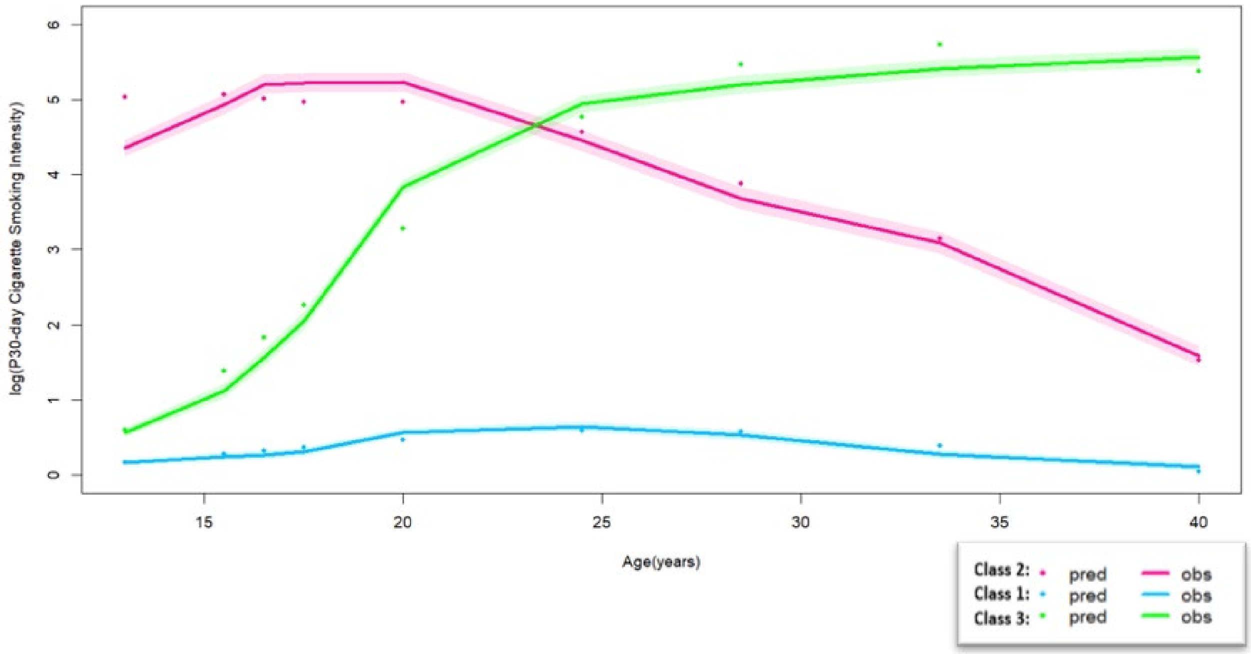


Figure 2-2. Subject-specific Trajectories of Log (Past 30-day Cigarette Smoking Intensity) from Early Adolescence to Adulthood



Chapter 3: The impact of Social Network Characteristics on Cigarette Smoking and Physical Activity during Adolescence and Transition from Adolescence to Young Adulthood

Introduction

The concept of social network as a structural representation of relations linking social actors was first formalized in the 1950s by Barnes and Bott to analyze social ties in the sociological context of traditional kinship.(132) In 1979, Berkman and Syme's paper(133) shed light on the importance of social networks in health outcomes, which led to an increasing interest in research investigating the role of social networks as a determinant of modifiable health behaviors, including cigarette smoking and physical activity (PA). As a result, a large number of observational studies have emerged in recent years addressing the effect of social networks on cigarette smoking and PA as evidence to inform behavioral intervention design.(134–138)

Extensive studies and theories suggest that individual health behaviors are correlated within a social network, especially among adolescents.(52–54) Such findings have commonly been explained through mechanisms such as perceived social norm and perceived peer pressure.(55),(56),(57) Moreover, different from other stages throughout life-course, adolescence is a time period where individuals are most susceptible to social influence, highlighting the need to further explore the impact of social network on health behaviors in adolescents.(139)

Despite the importance of the role of social networks to influence health behaviors, the term "social network" has been used loosely across previous studies. A majority of empirical studies in sociology have focused on the definition of social network as the set of actors and relational ties amongst them, and have been often conducted by social network analysis (SNA), where individuals' social relations are captured through a graph. Measures of interest from SNA include network structural characteristics such as network size, density and clustering. Existing studies using SNA to study modifiable health behaviors mainly addressed questions about how individual-level characteristics and behaviors affect the network formation process and ultimately the network structure.(140,141) Contrastingly, observational studies in epidemiology often address the question of how the social networks affect individual-level characteristics and behaviors, where perceived social support measured with the Social Network Index (SNI) was often used to address such question without obtaining individual-level relational data.(142,143) Nonetheless, a limited number of studies to date examined the relationship between network structural characteristic and individual health behavior, utilizing measures from SNA.

Different from SNI metrics that reflect perceptions of network from an individual point of view, measures derived from SNA can function as quantitative proxy for several sociological theoretical constructs that are key to mechanisms linking social network to population level distributions of health behaviors. For example, local clustering coefficient, an individual-level measure of how clustered the network is around each individual, has been shown to be indicative of behavior diffusion on a social network.(58) Therefore, measures from SNA provide an excellent opportunity to elucidate mechanisms linking social network characteristics to population level distributions of health behavior in observational studies.

Using the National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave I and Wave III data, this study aims to explore individual and network-level predictors of cigarette smoking and PA. Utilizing both the relational data obtained from friendship nomination survey and perceived social support from in-person interviews, this study will first obtain individual-level network structural characteristics from social network analysis and assess the associations between network measures and both PA and cigarette smoking behavior at Wave I. In addition, associations between network characteristics at Wave I and PA/cigarette smoking behaviors at Wave III will be assessed to investigate whether social network characteristics during adolescence are associated with individual health behaviors later in life. We hypothesize that social network characteristics are associated with both individuals' present health behaviors as well as their behaviors later in life. Results from this study will be able to provide further insights into whether social network characteristics play a role in shaping population level patterns of PA and cigarette smoking in order to inform future behavioral intervention design.

Methods

Study Population

The Add Health study is a longitudinal cohort study that enrolled a nationally representative sample of adolescents in the United States between grades 7 and 12 at baseline.(108) It was originally designed to facilitate a multidisciplinary approach to better understand causes of adolescent health behavior and outcomes throughout multiple developmental phases. At baseline (Wave I, 1993-1994), 20745 respondents completed an in-school interview and an at-home interview to collect information including health behaviors, and household and neighborhood-related characteristics. Respondents' parents were invited to complete interviews regarding parental sociodemographic background and household-level socioeconomic information. At Wave III (2001-2002), similar information on individual health behaviors was collected through in-home interviews.

All 20745 baseline participants completed a friendship nomination survey. Out of these 20745 baseline participants, 15760 individuals also participated in the friendship network survey either as a nominator (ego) or someone being nominated (alter). In the friendship nomination survey, individuals were invited to nominate up to 5 female close friends and 5 male friends. Throughout this manuscript, the nominator will be referred to as ego and the nominated friend will be referred to as alter for descriptive purposes. (**Supplemental Table 3-1**)

Eligible study participants included in the analysis to assess associations between social network structural characteristics and PA/cigarette smoking behaviors at Wave I included all Wave I study participants that were either an ego or an alter in the friendship nomination survey and had reported behavioral information regarding either cigarette smoking or PA. (**N = 15760**, **Supplemental Figure 3-1**) Individuals that did not participate in the baseline friendship nomination survey, did not nominate any other Add Health participant among their friends or only nominated friends that had an unspecified ID in the Add Health dataset, or had missing information on cigarette smoking or PA were excluded. For additional analyses assessing the association between network characteristics at Wave I and PA/cigarette smoking level at Wave III, we further excluded individuals that did not participate in Wave III and those without cigarette smoking or PA information at Wave III from eligible study participants in the previous analysis. (**N = 14986**, **Supplemental Figure 3-1**)

Definitions of relevant terminology used across this study are listed in **Supplemental Table 3-1**. Use of the data was reviewed and approved by the Institutional Review Board at Emory University and the Add Health study review boards.

Cigarette Smoking. Survey respondents were asked to self-report cigarette smoking behaviors during in-school and in-home interviews during both Wave I and Wave III. Questions regarding lifetime history of cigarette smoking and past 30-day (p30-day) cigarette smoking behavior were asked. In Wave I, the following questions were asked to determine respondents' current smoking status:1) Have you ever tried cigarette smoking, even just 1 or 2 puffs? 2) Have you ever smoked cigarettes regularly, that is, at least 1 cigarette every day for 30 days? 3) During the past 30 days, on how many days did you smoke cigarettes? In Wave III, the following questions were asked: 1) Have you ever tried cigarette smoking, even just one or two puffs? 2) Have you ever smoked an entire cigarette? 3) Have you smoked at all in the past 30 days? 4) During the past 30 days, on how many days did you smoke cigarettes? Based on these sets of questions, respondents were categorized as current smoker and current non-smoker. Current smokers were defined as those that have tried cigarettes and smoked cigarettes in the past 30 days. Otherwise, respondents were categorized as current non-smoker. In addition, whether smokers were present in the household during Wave I was reported as a binary response.

Physical Activity. Study respondents were asked to self-report their weekly frequency (times per week) of a series of standard physical activities including: jogging, walking, karate, jumping rope, gymnastics, dancing, roller-blading, roller-skating, skate-boarding, bicycling, or active sports. Previous studies(109),(110) have frequently used the definition of moderate-vigorous leisure-time physical activity through approximating number of metabolic equivalents. In this study, instead of using number of metabolic equivalents approximated, we generated a physical activity score corresponding to self-reported physical activity frequency of each questionnaire item to account for change in questionnaire designs. During Wave I, the following three questions were asked: During the past week, 1). how many times did you exercise, such as jogging, walking, karate, jumping rope, gymnastics or dancing? 2). how many times did you go roller-blading, roller-skating, skate-boarding, or bicycling? 3). how many times did you play an active sport, such as baseball, softball, basketball, soccer, swimming, or football? During Wave III, a total number of eight questions with a larger number of activities included were asked to determine the weekly physical activity level of an individual: During the past seven days, 1). how many times did you bicycle, skateboard, dance, hike, hunt, or do yard work? 2). how many times did you go to an exercise or fitness center to exercise or work out? 3). how many times did you participate in gymnastics, weight lifting, or strength training? 4). how many times did you participate in individual sports such as running, wrestling, swimming, cross-country skiing, cycle racing, or martial arts? 5). how many times did you participate in strenuous team sports such as football, soccer, basketball, lacrosse, rugby, field hockey, or ice hockey? 6). how many times did you play golf, go fishing or bowling, or play softball or baseball? 7). how many times did you roller blade, roller skate, downhill ski, snow board, play racquet sports, or do aerobics? 8). how many times did you walk for exercise? To standardize across both waves, if reported frequency was zero in the past seven days for each question, then the score was assigned as zero. If frequency was either once or twice in the past seven days, then the score was assigned as 1.5. Otherwise a score of 3.5 was assigned. A summary physical activity score was generated by summing up physical activity scores across all questionnaire items at each wave. Additionally, a standardized physical activity score across all five waves was generated by dividing the summary score by number of activities included in each wave's questionnaire to account for an increased number of activities included in questionnaires starting Wave III. Detailed questionnaires corresponding to tobacco use behavior and physical activity are available on https://www.cpc.unc.edu/projects/addhealth/documentation/publicdata.

Other Variables of Interest

*Socio-demographic Characteristics***.** Socio-demographic variables of interest included biological sex, race/ethnicity, parental education and household income reported at baseline visit. Survey respondents self-identified as White, African American, Hispanic, Asian/Pacific Islander/Native American/Alaska Indians, or Others. Highest level of parental education obtained by 1994 was reported. Respondents' parents were further dichotomized as having received a degree no more than high school or having received a degree beyond high school. In addition, total household pre-tax income including welfare benefits, dividends, and others was reported. A three-level ordinal variable was generated based on tertiles of reported household income.

Baseline Socio-psychological Characteristics. All respondents that participated in the in-home interview at Wave I were asked about whether they felt safe in the current neighborhood and whether they had access to a fitness or recreational center in the neighborhood. Perceived peer support was captured during baseline in-home interviews through questions on whether respondents felt cared for by friends.

*Respondents' Alter Characteristics***.** Friendship nomination survey respondents' alter information was linked to in-person interview surveys at Wave I to obtain alters' cigarette smoking behavior and physical activity score. Since every study participant was invited to nominated up to 10 friends, the mean of nominated friends' physical activity scores was obtained and included in the final analysis. In addition, an indicator variable of having nominated at least one friend who was a smoker was included in the final analysis. Detailed in-person interview and friendship nomination questionnaires are available on https://www.cpc.unc.edu/projects/addhealth/documentation/publicdata.

Social Network Analysis

At Wave I, out of 15760 eligible study participants, there were 14700 ego that nominated a specified friend that also participated in the Add Health study. Based on friendship nominations, an ego-centric social network was constructed to identify in-degree centrality and local clustering coefficient.

- In-degree centrality can be considered a measure of popularity within the network. It was calculated as the total number of nominations directed towards the respondent.
- Local clustering coefficient is a measure that captures the extent to which individuals within a social network tend to cluster together. It was calculated as:

  $\boldsymbol{C}_{node} = \frac{2\boldsymbol{N}(v)}{K(v)(K(v)-1)}$, where N(v) represents total number of network ties of each node and K(v) represents total number of network ties of each node's direct network connection.

Statistical analysis

The association between social network structural characteristics and individual cigarette smoking at Wave I was assessed using binary logistic regression, adjusting for individual socio-demographic characteristics, presence of smoker in the household, perceived peer support, and friend's smoking behavior. We first evaluated the association of network characteristics and individual smoking behavior without adjusting for friend's smoking behavior. Then we additionally assessed the association between network characteristic and individual smoking behavior, adjusting for friend's smoking behavior to see if the observed association between network structural characteristics and individual smoking behavior could be partially explained by friend's smoking behavior. In parallel, the association of social network structural characteristics with PA was assessed with multiple linear regression following a similar adjustment approach. To evaluate the effect of early life network characteristics on cigarette smoking behavior later in life, association between network characteristics at Wave I and cigarette smoking behaviors at Wave III was assessed using binary logistic regression, adjusting for friend's smoking behavior at Wave I, individual's smoking behavior at Wave I and socio-demographic characteristics. Similarly, for physical activity, the association between network characteristics at Wave I and the standardized physical activity score was assessed, adjusting for neighborhood factors, individual socio-demographic characteristics and friend's physical activity level using multiple linear regression. Additional analysis evaluating the association between network characteristics at Wave I and PA level at Wave III was also performed using multivariate linear regression, adjusting for friend's physical activity level at Wave I and individual socio-demographic characteristics. Furthermore, we tested the statistical interactions between network characteristics and friend's health behaviors by including a multiplicative interaction term in the models mentioned above for both PA and cigarette smoking. We hypothesized that both local clustering coefficient and in-degree centrality would be associated with cigarette smoking/PA behavior at Wave I, even after adjusting for friend's behavior. In addition, in-degree centrality and friend's behavior at Wave I would be predictive of individual's health behaviors at Wave III. All analyses were performed in R 3.5.2.

Results

A total number of 15760 individuals were included in the analysis (mean age = 15.6 yrs, **Table 3-1**). Out of the 15760 individuals, 14700 were egos that nominated at least one individual with specified ID in the baseline survey and 1060 were additional alters with matching ID in the baseline survey that did not nominate anyone but were nominated by egos during Wave I friendship nomination survey. The proportion of female versus male was comparable (N = 8025, 51% females) and 54% of respondents were non-Hispanic white. With regards to household information, 42% reported parents receiving an education less than high school. In addition, 45% reported having a smoker in the household. With regards to neighborhood environment, only 21% reported having access to a recreational center in the neighborhood whereas close to 90% individuals reported feeling safe in the neighborhood. Almost everyone reported perceived care from friends (N = 15367, 98%). Mean physical activity level was comparable amongst ego and nominated alter (mean = 0.38, **Table 3-1**). Interestingly, the proportion of smokers was higher in nominated friends (40%) as compared to that in ego (25%, **Table 3-1**).

Association between Social Network Characteristics at Wave I and Cigarette Smoking Behavior/Physical Activity at Wave I

Before adjusting for friend's smoking behavior, in-degree centrality of network at Wave I was associated with respondents' smoking behavior at Wave I. Per unit increase in in-degree centrality was associated with an increased odds of the respondent being a smoker (OR = 1.05, 95% CI: 1.02,1.07, **Table 3-2**). However, local clustering coefficient was not associated with respondent's cigarette smoking behavior at Wave I (OR = 0.96, 95% CI: 0.70, 1.33, **Table 3-2**).

After further adjusting for friend's smoking behavior at Wave I, such findings remained. Only in-degree centrality at Wave I was significantly associated with respondent's smoking behavior when adjusting for friend's smoking behavior, individual socio-demographic information and presence of smoker in the household. (OR = 1.07, 95% CI: 1.03, 1.10).

With regards to PA, both local clustering coefficient and in-degree centrality were associated with respondent's PA level at Wave I when not adjusting for friend's PA level. (**Table 3-3**) However, when additionally adjusting for friend's PA level, only in-degree centrality (0.006, 95% CI: 0.004, 0.008) but not local clustering coefficient (-0.016, 95% CI: -0.060, 0.026) was significantly associated with respondent's PA level at Wave I. Higher in-degree centrality was associated with higher PA level among respondents when adjusting for friend's PA level, individual socio-demographic characteristics and neighborhood characteristics.

In addition to network characteristics, nominated friend's behaviors, both cigarette smoking and PA at Wave I, was consistently associated with respondents' behavior. In the fully adjusted model including network characteristics, having nominated a smoker as a friend was associated with 3 times increased odds of respondent being a smoker (OR = 3.47, 95% CI: 2.94, 4.09). For PA, higher friend's PA level was also associated with an increased PA level of respondent (0.11, 95% CI: 0.08, 0.14).

Association between Social Network Characteristics at Wave I and Cigarette Smoking/Physical Activity at Wave III

We further assessed whether network characteristics and network connections' behavior during adolescence (Wave I) would have an enduring impact on respondents' health behavior in young adulthood (Wave III). Findings from our model showed that neither in-degree centrality nor local clustering coefficient were associated with respondent's smoking behavior as young adult (**Table 3-4**). However, respondents that nominated at least one smoker as a friend at Wave I were 1.5 times more likely to be a smoker at Wave III as compared to those that did not

(OR = 1.48, 95% CI: 1.18, 1.86). After testing for statistical interactions between network characteristics and friend's smoking behavior, we did not find the interaction to be statistically significant (p for interaction between local clustering coefficient and friend' smoking behavior = 0.27, p for interaction between in-degree centrality and smoking = 0.15). Different from findings on cigarette smoking, we did not find an association between friend's PA level during adolescence and individuals' PA level later in life (0.001, 95% CI: -0.003, 0.004; **Table 3-5**). Nevertheless, higher in-degree centrality at Wave I was associated with higher PA level for young adult (0.004, 95% CI: 0.001, 0.007). Similarly, the statistical interactions between network characteristics and friend's PA level were not significant (p for interaction between local clustering coefficient and friend' smoking behavior = 0.76, p for interaction between in-degree centrality and smoking = 0.07)

Discussion

Our results indicate that in-degree centrality, a commonly used network measure indicating one's network position or popularity, was associated with both cigarette smoking and PA among adolescents. Local clustering coefficient, as an indication of whether one's friends also know each other and presence of clique around an individual, was associated with PA but not with cigarette smoking. Notably, friend's smoking behavior during adolescence was predictive of individual's smoking's behavior in young adulthood whereas for physical activity, a higher in-degree centrality in adolescence was associated with higher PA level in young adulthood.

Results from our study did not show a significant association between individual level network clustering and cigarette smoking behavior among adolescents. Even though various studies have shown that adolescent smokers tend to form cliques within the social network due to peer influence(144), one study investigating the role of peer group structure and adolescent cigarette use has also shown that members of friendship groups are more likely to be non-smokers than individuals that do not have a social group membership.(145) In addition, evidence from a systematic review has suggested that social isolation is associated with adolescent smoking.(146)

Collectively, finding from our study might be because some adolescent smokers were not forming cliques due to perception of smoking as a non-normative behavior, thus leading to the observation of null result. Meanwhile, consistent with existing studies(147),(148), our finding showed that higher in-degree centrality is associated an increased likelihood of one being a smoker. One mechanistic hypothesis is that if smoking is the perceived social norm amongst social groups that have popular individuals as smokers, popular individuals are more likely to continue being a smoker in order to maintain their social status. (148,149) For physical activity, even though one study has shown that social group formation contributes to higher PA level due to social support and internalized social identity associated with higher levels, (150) our findings showed that there was no association between local clustering coefficient and PA level. It could be potentially because of higher proportions of individual activities included in the PA questionnaire as compared to group activities. Similar to cigarette smoking, one possible explanation for our finding on association between in-degree centrality and PA is that individuals that engage in PA are more likely to be more socially connected and establish a social identity, which in turn further promotes engagement in PA level. (150) Consistent with our finding, one study has also shown that social position as captured with network centrality measures is associated with PA using an online social network database. However, different from cigarette smoking, extremely limited number of studies have investigated the association between network characteristics and PA overall.(151)

Existing studies and theories in developmental psychology and sociology have suggested that social influence during adolescence is critical to formation of social identity(152,153), which can further influence one's long-term engagement in cigarette smoking and physical activity by promoting positive affect associated with these behaviors. In our study, we hypothesized that in-degree centrality, as a measure that can capture one's social identity within the social network at a younger age might be predictive of one's smoking/PA level later in young adulthood. Consistent with one previous study that found an association between centrality measure and sports engagement later in life(154), our study showed that in-degree centrality was predictive of higher PA level later in life. However, we did not find an association between network characteristics during adolescence and smoking behavior in young adults. It could be because of the complex process of smoker identity development. Peer influence plays an important role in shaping smoker identity during adolescence(155) whereas smoker identity formation during young adulthood is more reliant upon individual rationalization.(156)

Our study is one of the first studies that investigated the association between network structural characteristics and cigarette smoking/PA behaviors in adolescents, as well as long-term impact of network characteristics in adolescence upon health behaviors in young adulthood. However, several limitations need to be acknowledged. First, we excluded individuals that nominated friends with unspecified ID. Adolescents who chose not to disclose their friends' names might be different from those who did in terms of social engagement, which might lead to biased estimates, especially for physical activity. Second, the cross-sectional nature of the analysis investigating the association between network characteristics and both behaviors did not allow us to explore whether observed association could be due to peer friendship selection at Wave I. Last but not the least, due to the self-reported nature of the survey among adolescents, current smoking in the study population could be under-reported. In addition, due to changes in survey design, we used a standardized physical activity score as a measure for physical activity level. The validity of such instrument needs to be further assessed in future research.

Conclusion

Using a large nationally representative sample of adolescents, our study showed that an individual's social position, measured as in-degree centrality, is associated with both adolescent cigarette smoking behavior and physical activity. In addition, one's social position during adolescent is a significant predictor of physical activity level in young adult. Such findings collectively emphasize the potential of using community / network-based intervention for smoking cessation and physical activity promotion as well as the importance of early intervention during adolescence for healthy behavior promotion later in life.

Table 3-1. Baseline Characteristics of Eligible Study Participants

	Study Population Characteristics
N	15760
Age (yrs)	15.6 (1.70)
Socio-demographic Characteristics: N (%)	
Female	8025 (51)
Race/Ethnicity	
Non-Hispanic White	8498 (54)
Non-Hispanic Black	3125 (20)
Hispanic	2506 (16)
Other	1631 (10)
Parental Education (Less than High school)	6343 (42)
Presence of smoker in household	5999 (45)
Access to recreational center in neighborhood	3263 (21)
Feel safe in the neighborhood	14049 (89)
Perceived friends' care	15356 (98)
Current smoker	4016 (25)
Standardized Physical Activity Score: Mean (sd)	0.38 (0.23)
Social Network Characteristics	
Mean Degree Centrality	3.06 (2.99)
# unique ties nominated	34151
Nominated Alters' Characteristics	
Current Smoker	1572 (40)
Standardized Physical Activity Score: Mean (sd)	0.39 (0.22)
*sd: standard deviation	

Table 3-2. Association of Individual and Social Network Characteristics and Cigarette Smoking Behavior at Wave I, the Add Health Study

		Odds Ratio (95% CI)
Model 1	**Female**	0.98 (0.91, 1.05)
	Race (reference = Non-Hispanic White)	
	Non-Hispanic Black	0.27 (0.24, 0.31)
	Hispanic	0.50 (0.45, 0.56)
	Others	0.55 (0.49, 0.63)
	Age	1.22 (1.20, 1.26)
Model 2	**Having at least one nominated friend as a smoker**	3.46 (2.93, 4.08)
	Female	0.95 (0.81, 1.11)
	Age	1.15 (1.10, 1.20)
	Race (reference = Non-Hispanic White)	
	Non-Hispanic Black	0.39 (0.30, 0.50)
	Hispanic	0.69 (0.54, 0.87)
	Others	0.79 (0.61, 1.01)
	Parent education less high school	1.15 (0.97, 1.36)
	Parental income (reference = first tertile)	
	Second Tertile	1.16 (0.95, 1.42)
	Third Tertile	1.05 (0.85, 1.30)
	Perceived care from friends at Wave I	0.84 (0.47, 1.52)
	Presence of smoker in household at baseline	2.32 (1.97, 2.72)
Model 3	**In-degree Centrality at Wave I***	1.05 (1.02, 1.07)
	Local Clustering Coefficient at Wave I**	0.96 (0.70, 1.33)
	Age	1.23 (1.20, 1.26)
	Female	0.97 (0.89, 1.05)
	Race (reference = Non-Hispanic White)	
	Non-Hispanic Black	0.30 (0.26, 0.35)
	Hispanic	0.57 (0.50, 0.65)
	Others	0.67 (0.58, 0.78)
	Parent education less high school	1.17 (1.07, 1.28)
	Parental income (reference = first tertile)	
	Second Tertile	1.01 (0.90, 1.12)
	Third Tertile	1.04 (0.93, 1.16)
	Perceived care from friends at Wave I	0.85 (0.63, 1.15)
	Presence of smoker in household at baseline	2.18 (2.00, 2.37)
Model 4	**In-degree Centrality at Wave I***	1.07 (1.03, 1.10)
	Local Clustering Coefficient at Wave I**	0.82 (0.48, 1.40)
	Having at least one nominated friend as a smoker	3.47 (2.94, 4.09)
	Age	1.14 (1.08, 1.19)

	Female	0.94 (0.81, 1.10)
	Race (reference = Non-Hispanic White)	
	Non-Hispanic Black	0.41 (0.32, 0.53)
	Hispanic	0.70 (0.55, 0.89)
	Others	0.78 (0.61, 1.01)
	Parent education less high school	1.14 (0.97, 1.35)
	Parental income (reference = first tertile)	
	Second Tertile	1.13 (0.92, 1.38)
	Third Tertile	1.03 (0.84, 1.28)
	Perceived care from friends at Wave I	0.80 (0.44, 1.45)
	Presence of smoker in household at baseline	2.31 (1.97, 2.71)

Model 1: Logistic regression adjusted for age, and sociodemographic characteristics
Model 2: Logistic regression adjusted for age, sociodemographic characteristics, alter's smoking behavior
Model 3: Logistic regression adjusted for age, sociodemographic characteristics, and social network characteristics
Model 4: Logistic regression adjusted for age, sociodemographic characteristics, alter's smoking behavior and social network characteristics
**In-degree Centrality has the unit of 1 connection.*
***Local Clustering Coefficient is a ratio bounded between 0 and 1. LCC has a mean of 0.4 and standard deviation of 0.13.*

Table 3-3. Association of Individual and Social Network Characteristics and Physical Activity Level at Wave I, the Add Health Study

		physical activity score (95% CI)
Model 1	**Female**	-0.098 (-0.105, -0.092)
	Race (reference = Non-Hispanic White)	
	Non-Hispanic Black	-0.019 (-0.027, -0.010)
	Hispanic	-0.011 (-0.020, -0.001)
	Others	0.001 (-0.010, 0.012)
	Age	-0.033 (-0.034, -0.031)
Model 2	**Friend's physical activity level**	0.12 (0.09, 0.15)
	Female	-0.079(-0.092, -0.066)
	Age	-0.032 (-0.035, -0.027)
	Race (reference = Non-Hispanic White)	
	Non-Hispanic Black	-0.015 (-0.033, 0.003)
	Hispanic	0.006 (-0.014, 0.025)
	Others	0.021 (0.000, 0.042)
	Parent education less high school	-0.012 (-0.026, 0.002)
	Parental income (reference = first tertile)	
	Second Tertile	0.001 (-0.016, 0.018)
	Third Tertile	0.005 (-0.012, 0.023)
	Perceived care from friends at Wave I	0.008 (-0.041, 0.057)
	Access to recreational center in the neighborhood	0.075 (0.059, 0.091)
	Feeling safe in the neighborhood	0.017 (-0.006, 0.041)
Model 3	**In-degree Centrality at Wave I***	0.006 (0.004, 0.008)
	Local Clustering Coefficient at Wave I**	-0.046 (-0.074, -0.018)
	Age	-0.034 (-0.037, -0.032)
	Female	-0.092 (-0.100, -0.085)
	Race (reference = Non-Hispanic White)	
	Non-Hispanic Black	-0.018 (-0.028, -0.009)
	Hispanic	-0.004 (-0.015, 0.008)
	Others	-0.003 (-0.016, 0.010)
	Parent education less high school	-0.021 (-0.029, -0.013)
	Parental income (reference = first tertile)	
	Second Tertile	0.010 (0.000, 0.019)
	Third Tertile	0.022 (0.013, 0.032)
	Access to recreational center in the neighborhood	0.077 (0.068, 0.086)

	Feeling safe in the neighborhood	0.016 (0.003, 0.029)
	Perceived care from friends at Wave I	0.010 (-0.016, 0.036)
Model 4	**In-degree Centrality at Wave I***	0.007 (0.004, 0.010)
	Local Clustering Coefficient at Wave I**	-0.016 (-0.060, 0.026)
	Friend's physical activity level	0.11 (0.08, 0.14)
	Age	-0.032 (-0.036, -0.028)
	Female	-0.080 (-0.093, -0.067)
	Race (reference = Non-Hispanic White)	
	Non-Hispanic Black	-0.010 (-0.028, 0.008)
	Hispanic	0.008 (-0.012, 0.027)
	Others	0.020 (-0.001, 0.041)
	Parent education less high school	-0.013 (-0.027, 0.001)
	Parental income (reference = first tertile)	
	Second Tertile	-0.001 (-0.018, 0.015)
	Third Tertile	0.004 (-0.013, 0.021)
	Perceived care from friends at Wave I	0.003 (-0.045, 0.052)
	Access to recreational center in the neighborhood	0.072 (0.056, 0.088)
	Feeling safe in the neighborhood	0.018 (-0.010, 0.041)

Model 1: Linear regression adjusted for age, and sociodemographic characteristics
Model 2: Linear regression adjusted for age, sociodemographic characteristics, neighborhood characteristics and alter's physical activity level
Model 3: Linear regression adjusted for age, sociodemographic characteristics, neighborhood characteristics, and social network characteristics
*Model 4: Linear regression adjusted for age, sociodemographic characteristics, neighborhood characteristics, alter's physical activity level and social network characteristic**
**In-degree Centrality has the unit of 1.*
***Local Clustering Coefficient is a ratio bounded between 0 and 1. LCC has a mean of 0.4 and standard deviation of 0.13.*

Table 3-4. Association of Early Life Social Network Characteristics and Cigarette Smoking Behavior in Young Adulthood, the Add Health Study

	Odds Ratio (95% CI)
Age at Wave III	0.87 (0.82, 0.92)
Female	0.61 (0.50, 0.73)
Race (reference = Non-Hispanic White)	
Non-Hispanic Black	0.38 (0.29, 0.50)
Hispanic	0.53 (0.39, 0.71)
Others	0.67 (0.49, 0.91)
Parent education less high school	1.18 (0.96, 1.45)
Parental income (reference = first tertile)	
Second Tertile	0.87 (0.68, 1.11)
Third Tertile	0.96 (0.75, 1.24)
Presence of smoker in household at baseline	1.64 (1.35, 1.99)
Local Clustering Coefficient at Wave I*	1.20 (0.63, 2.27)
In-degree Centrality at Wave I**	0.98 (0.94, 1.02)
Being a Smoker at Wave I	6.77 (5.39, 8.50)
Having at least one friend as a smoker at Wave I	1.48 (1.18, 1.86)

Model: Logistic regression adjusted for age, sociodemographic characteristics, cigarette smoking behavior at Wave I, alter's smoking behavior at Wave I and social network characteristics
**Local Clustering Coefficient is a ratio bounded between 0 and 1. LCC has a mean of 0.4 and standard deviation of 0.13.*
***In-degree Centrality has the unit of 1.*

Table 3-5. Association of Early Life Social Network Characteristics and Physical Activity Level in Young Adulthood, the Add Health Study

	Physical Activity Score (95% CI)
Age at Wave III	-0.003 (-0.008, 0.002)
Female	-0.078 (-0.095, -0.061)
Race (reference = Non-Hispanic White)	
Non-Hispanic Black	0.035 (0.012, 0.058)
Hispanic	0.042 (0.019, 0.066)
Others	0.018 (-0.008, 0.044)
Parent education less high school	-0.006 (-0.024, 0.012)
Parental income (reference = first tertile)	
Second Tertile	0.027 (0.007, 0.048)
Third Tertile	0.026 (0.004, 0.047)
Local Clustering Coefficient at Wave I*	-0.010 (-0.06, 0.04)
In-degree Centrality at Wave I**	0.004 (0.001, 0.007)
Physical activity level at Wave I	0.12 (0.08, 0.16)
Friend's physical activity level at Wave I	0.001 (-0.003, 0.004)

Model: Linear regression adjusted for age at Wave III, sociodemographic characteristics, physical activity level at Wave I, alter's physical activity level at Wave I and social network characteristic at Wave I
**Local Clustering Coefficient is a ratio bounded between 0 and 1. LCC has a mean of 0.4 and standard deviation of 0.13.*
***In-degree Centrality has the unit of 1.*

Supplemental Table 3-1. Terminology and Definition for Social Network Analysis

Terminology	**Definition and Calculation Methods**
Ego	Individuals that nominated others in the friendship network survey
Alter	Individuals that were nominated in the friendship network survey
Node	A vertex used to represent a student on a network graph
Tie	A link/connection between two nodes on a network graph
Edge	A line used to represent a link between a pair of nodes on a network graph
Degree	Number of connections a node has to other nodes on a network graph
Ego-centric Network	The type of network that maps the connections from ego's perspective
Network Structural Features	
In-Degree Centrality	Sum of number of friendship ties directing towards ego
Local Clustering Coefficient	Fraction of a pair of nodes' friends that are also friends with each other

Supplemental Figure 3-1. Inclusion Criteria of Study Population

Wave I Study Participants
(N = 20745, Mean age = 15 yrs)

Eligible study participants Criteria for Cross-sectional Analysis at Wave I
1). Participated in friendship nomination survey and either nominated a friend with a specified matching ID at baseline survey or was nominated with a matching ID at baseline survey
2). Had complete information on either cigarette smoking and physical activity at Wave I
(N = 15760, Mean age = 15 yrs)

Eligible study participants Criteria for Longitudinal Analysis Using Wave I and III of Add Health Study
1). Participated in friendship nomination survey and nominated a specified friend that also participated in baseline survey
2). Had complete information on either cigarette smoking or physical activity at Wave I
3). Had complete information on either cigarette smoking or physical activity at Wave III
(N = 14986, Mean age = 22 yrs)

Chapter 4: From Individual Will to Population Outcomes: a Complex Systems Framework to Evaluate Behavioral Intervention Effectiveness on Social Networks

Introduction

Cardiovascular disease (CVD) remains a leading cause of death in the United States and worldwide.(157) Modifiable behavioral risk factors such as physical activity, tobacco use and sedentary behaviors provide an excellent opportunity for CVD prevention and intervention.(157) Nevertheless, identifying an optimal intervention strategy targeting modifiable health behaviors still faces tremendous obstacles in the public health field, especially due to challenges in effectiveness evaluation.(158)

A comprehensive evaluation of interventions targeting modifiable behaviors is a multifaceted problem. An intervention targeting modifiable behavioral risk factors can be considered effective if it is able to address the following aspects: whether it will reach the targeted population as desired; and, once accessed, whether initiated behavioral change will be sustained to ensure long-term effectiveness. Two important, but often overlooked factors that affect both aspects of intervention effectiveness are social network characteristics and individual behavioral choices. Extensive evidence from sociology and behavioral economics suggests that social networks have a significant impact on information diffusion and behavioral choices.(159–164) Such findings can have important implications for evaluations of intervention effectiveness. Diffusion of interventions across a population can be considered analogous to information diffusion over populations with a heterogeneous network structure(165,166). The type of intervention determines the proportion and characteristics of a population that have initial access to the intervention. Under this framework, whether an intervention is effective will be largely determined by diffusion process and individuals' choices to adopt a behavioral change, given social network characteristics.

Nevertheless, a comprehensive evaluation of intervention effectiveness can be challenging. Certain network structures, such as clustering, can facilitate information diffusion, but might hamper the long-term sustainment of behavioral adoption due to high levels of social inertia within the segregated community.(59,167,168) Leveraging these paradoxical aspects is extremely challenging using traditional statistical methods. In addition, the dynamic interplay between information diffusion and individual behavioral choices within a network can lead to emergence of unforeseen patterns of collective behavior,(68) which is important to consider when assessing the optimal population-level impact of an intervention under limited resources. Therefore, a complex systems approach is needed to evaluate intervention effectiveness on behavioral risk factors, taking into consideration the nonlinear dynamic interactions of inter-dependent factors and emerging collective behavioral patterns on a social network.

To address the research question of how individual incentives and social networks affect behavioral intervention effectiveness respectively and collectively, we constructed several computational experiments using an agent-based modeling (ABM) approach to elucidate the mechanisms linking individual decision-making cascades, social networks and population-level behavioral patterns as well as extrapolating the role of these three factors on behavioral intervention outcomes. We will use smoking behavior as an example. Different from traditional statistical approach and other types of stochastic models, ABM allows for four key features that are crucial to the understanding of how multi-level factors affect the population level outcomes of behavioral interventions: a. agents' decision-making b. feedback and interactions between agents and their social network c. emerging outcomes in the population level, d. "spillover effect" or indirect effect of the social network on individual-targeting interventions. Through these computational experiments, we aim to address the following research questions: a. Do individual incentives for behavior change affect population dynamics of smoking behavioral patterns? b. Does diffusion of interventions across the social network have an effect on network-based interventions? c. How do different combinations of an initial seeding strategy and network structural characteristics affect outcomes of network-based intervention? d. Given a specific network structure, would network-based intervention strategies or individual-targeting intervention strategies be more effective?

Methods

In this study, hypothesized causal mechanisms are built into several computational experiments. We will use smoking as an example. Details of model parameters, model-setup, and key assumptions will be described in each model description section. Given the goal of this study to provide a novel framework and understanding of factors contributing to behavioral intervention outcome instead of identifying the true prevalence of smoking over time under different intervention scenarios, values of model parameters are hypothesized based on literature review and each time-step of the simulation cannot be interpreted as calendar time. Rather, time-steps of simulation are abstract representation of simulation process from t_0 until simulation reaches steady-state.

Overview of Computational Experiment Rationale

We will initialize a network of agents with agent attributes of a binary smoking status ("smoker"/" non-smoker") and an inherent opinion towards smoking. (**Table 4-1**) At each time step of the simulation, agent changes its opinion based on those of its network connections and updates its behavior. To identify the role of individual decision making on population-level behavioral dynamics, we adopted the framework of the DeGroot(169) model of belief for agents' behavior updating function. DeGroot's model of belief has been widely applied in research of information transmission, game theory, and social learning where there is an inherent bias towards a type of behavior, which makes the model ideal for study of behaviors such as smoking, where social norm and social learning are two key components of behavior formation.(170)

In the baseline scenario, we will establish a model that only takes into consideration the concept of peer influence, no individual opinion heterogeneity and smoking behavior as the preferred behavior in the population in absence of intervention. After the baseline model is established, additional heterogeneity is added by relaxing assumptions on agent's individual will to alter their opinion towards smoking, which allows us to explore how individual incentives affect population-level opinion dynamics and ultimately smoking behavioral pattern. To assess the potential effectiveness of network-based intervention on a population, intervention diffusion was additionally introduced to the model incorporating individual incentive heterogeneity. Additional sensitivity analyses were performed by exploring different combinations of initial network structure, network-based intervention seeding strategy and agents' openness to adopt new opinion to identify the optimal intervention strategy, given population network structure and initial prevalence of smoking. The bottom-up model building process will allow us to elucidate the mechanisms in terms of how individual incentive and network-based intervention affect population level prevalence of smoking behavior respectively and collectively.

Baseline Scenario: Model Parameter

1000 agents were initialized on a small-world network, a class of network that most closely represents the network in real-life. (171) We started the baseline simulation with an initial prevalence of smoking as 50%. Additional sensitivity analyses with an initial prevalence of 20% and 80% were later conducted. Agent attribute: opinion towards smoking (δ) was assigned according to agents' smoking status. For smokers, agent's opinion was randomly sampled from a uniform distribution within the range of [6, 10]. For non-smokers, opinion was randomly sampled from a uniform distribution within the range of [1,5]. Descriptions of parameters are listed in **Table 4-1**. For baseline model, no inertia score was considered.

Baseline Scenario: Model Setup

In the baseline scenario, all agents updated their opinion towards smoking based on the average of direct network connections' opinion at the previous time step.

$$\delta_{i(t+1)} = \frac{\sum_{i \neq j,j}^{n} \delta j(t)}{\sum_{i \neq j,j}^{n} \mathbf{j}} \textit{\textbf{(eq.1)}}$$

At each time step, agents first updated their opinion based on ***eq.1***, as modified from DeGroot Averaging Model(169). For smokers, if the updated opinion is greater than 5, then they maintained current smoking status. Otherwise, the status was altered to non-smoker. For non-smokers, if the updated opinion is less than or equal to 5, then they maintained their non-smoking status. There is one key assumption to our baseline model. We assumed that no agent possessed an individual will to decide their behavioral choice. This assumption was used to approximate an extreme condition where peer influence is the only factor affecting individual's behavior. Use of this assumption allowed us to illicit the role of individual level heterogeneity on population level outcomes.

Experiment 1: Impact of individual decision-making to adopt a new idea on population-level smoking behavior pattern

Similar to baseline scenario, 1000 agents were initialized on a small-world network with an initial smoking prevalence of 50%, as well as agent attribute of opinion towards smoking.

$$\delta_{i(t+1)} = \frac{\omega \times \delta\, i(t) + \sum_{i \neq j,j}^{n} (1-\omega) \delta j(t)}{\omega + \sum_{i \neq j,j}^{n} (1-\omega) \mathbf{j}} \textit{\textbf{(eq.2)}}$$

For each time step of the simulation, agents updated their opinion based on a convex combination as shown in ***eq. 2***(169). Different from baseline scenario, individual's incentive to adopt a new idea was factored into agents' updating function. An additional agent attribute – inertia (ω), defined as the reluctance to adopt any new information, was generated as a random variable following a beta distribution of β (3,3). (**Table 4-1**) This combination allowed for exploration of the role of individual's will in facilitating decision making in a social setting, where agents rely on social heuristics to make decisions regarding behavior.(172) In addition, with a sampling distribution of β (3,3), we assumed the population to be neutral towards adopting a new behavior, given the sample mean of inertia around 0.5. Similar to the baseline scenario, agents' smoking status was updated at every time step based on updated opinion. Results from this experiment were then compared to those from the baseline scenario to identify the impact of individual decision-making on population-level smoking behavior pattern. The key assumption to this model is that we assumed agents were able to "quantify" the weight it assigns to its own opinion as supposed to those of its connections. Embedding this assumption in the model allowed us to implicitly embed a decision-making process of weighing agent's initial opinion against other's opinion, which can be extrapolated to a decision cascade in deciding whether one should adopt a new opinion associated with intervention.

Experiment 2: Impact of a network-based targeted intervention on population smoking behavioral patterns

Similar to **Experiment 1**, 1000 agents were initialized on a small-world network with the same distribution of initial prevalence of smoking, opinion towards smoking and levels of inertia. To evaluate the impact of a network-based targeted intervention on population smoking prevalence, we added an additional component of intervention diffusion on the social network to **Experiment 1**. This simulation was comprised of two steps: Step 1, diffusion of intervention; Step 2, agents update behavior. During step 1 of the simulation, 10% of the entire population were assigned as initial seeding population of the intervention. Diffusion of the intervention followed the diffusion of a complex contagion model with a threshold of 40% based on literature(173). That is, at every time step, an agent adopted an intervention if greater than 40% of its network connection adopted an intervention in the prior time step. If the threshold lowers to below 40%, agent will abandon the intervention. This design relaxed the stringent assumption of 100% adherence to the intervention. We set the diffusion of intervention as prior to agents updating their opinion and smoking behavior. Upon the diffusion status of intervention got updated at every time step of the simulation, agents followed the following decision cascade for behavior update:

1). If agent is a smoker and has access to the intervention, then update behavior to non-smoker. Otherwise, update opinion according to equation 2 as mentioned previously.

2) If agent is a non-smoker and has access to the intervention, then maintain its behavior, otherwise, update opinion according equation 2 as mentioned previously.

To further assess how **diffusion of the intervention** affects population level outcomes of network-based intervention strategies, we altered two components of experiment 2: a. initial seeding strategy; b. network structure. With respect to the initial seeding strategy, we compared the scenario of using **10% of randomly selected** initial adopters of the intervention to that of using **10% of initial adopters ranked by top 10% of in-degree centrality**. With respect to network structure, we first compared outcomes of both seeding strategies on an Erdos-Renyi Random network (ER network) with an average of 2 neighbors per node as compared to that of small world network used in previously experiments with an average 2 neighbors per node. ER network has a longer average path length and higher local clustering as compared to that of a small world network. Additionally, we compared outcomes of both seeding strategies on small world networks with different global clustering coefficients. (0.05 and 0.26 respectively) These two sets of simulations allowed us to evaluate the following research questions: a. given the set network structure, which type of initial seeding strategies would be more effective? b. given the set seeding strategy, would local clustering or global clustering of the network matter more to the long-term effectiveness of network-based intervention? There are several assumptions to this experiment. First, we assumed that when an agent becomes an intervention adopter, it will adopt the intervention-biased behavior with a probability of 1. Second, we assumed that diffusion of the intervention, even though occurring within the same time-step of simulation, is sequentially prior to agents' opinion updating.

Experiment 3. The joint impact of social network local-clustering and individual incentives on population smoking behavior pattern.

To further address the collective impact of the node-level network structure and individual incentives on outcomes of the intervention, additional sensitivity analyses on Experiment 2 with ω of varying beta distributions was performed. ω following a sampling distribution of β (2,5), resembles a condition where a proportion of the population is incentivized to be more open to adopt new ideas regarding behaviors whereas ω following a sampling distribution of β (8,2) resembles a relatively conservative population with a mean inertia score closer to 1, meaning a higher social inertia (i.e. higher mean population-level inertia score) for the community to introduce intervention for behavioral change. Interventions that aim to incentivize individuals were operationalized as a differing distribution of inertia score in the population as a proxy. Key assumptions of this experiments were the same as those in experiment 2.

Results

Baseline scenario

In the baseline scenario, we started with a hypothetical cohort of individuals with 50% of smoking prevalence. Each agent possessed an innate opinion towards smoking that was directly correlated with their smoking behavior. At the start of the simulation, the population mean opinion was 5.32. With peer influence on agent's opinion as the only factor affecting population dynamic, given the initial prevalence of smoking of 0.5 in the population and a slight population-level bias towards smoking at model initiation (population mean opinion = 5.32, **Table 4-2**), we observed the convergence of behavior towards smoking as the majority behavior as the model stabilized. This baseline model functioned as the reference and as an intervention-free community of population with an established normative bias towards smoking. Additional sensitivity analyses were done on a hypothetical cohort with an initial prevalence of 20% and 80%. For simulation with an initial prevalence of 0.2, we observed prevalence of smoking stabilized around 0.17 instead of 1. These findings further confirmed the direct effect of peer influence on population prevalence of smoking, conditioning upon the baseline prevalence of smoking behavior. (**Table 4-2**, **Figure 4-1**).

Experiment 1: Impact of individual decision making on the prevalence of smoking in the population

In this simulation, agents weighed their own inherent opinion towards smoking against their network peers' opinion. We tested the hypothesis that given an even mix of 50% of agents that were more reluctant to adopt a new opinion and 50% of agents that were more open to adopt a new opinion with a mean weight of initial opinion centering around 0.53, prevalence of smoking after simulation stabilizes would be around 50%. However, results from this experiment were rather counterintuitive. After 300 simulation iterations, mean population opinion towards smoking remained similar to the initial mean of 5.36. However, prevalence of smoking decreased from 0.50 to 0.38, suggesting the effect of agent's inertia ω heterogeneity to adopt a new opinion is non-trivial on the population-level smoking behavior pattern. (**Figure 4-2)** Experimental condition with ω following sampling distribution of β *(3,3)* resembles a population free of intervention and neutral towards the concept of adopting a new opinion, which serves as an ideal reference point for simulations exploring effectiveness of network-based intervention strategies.

Experiment 2a: Effect of network-based intervention on prevalence of smoking in population

Simulation results from experiment 2 with varying initial seeding strategies showed that given a specific network structure and the condition that agents are not incentivized to adopt new behaviors, the diffusion of an intervention over a social network does have an impact over population-level outcomes of intervention. (**Table 4-3**) A random seeding strategy of a network-based intervention is a good example of any intervention targeting proportions of individuals within the community such as a community-based campaign or school-based intervention. In the meantime, interventions with an initial seeding strategy of selected individuals with highest degree centrality resembles interventions targeting key opinion leaders and popular individuals within the network. Results from simulations over varying network structures suggested that regardless of network structure, interventions targeting central individuals are more effective than randomly selecting targeted individual. (**Table 4-3**, **Figure 4-3**, **Figure 4-4**, **Figure 4-5**).

Experiment 2b**: **Effect of network structural characteristics on intervention outcomes

To assess the effect of network structural characteristics on intervention outcomes, we compared outcomes of network-based interventions on three types of networks: a small world network with a global clustering coefficient of 0.26, a small world network with a global clustering coefficient of 0.05 and an Erdos-Renyi (ER) network. Comparing results from the two small world networks, our results showed that global clustering coefficient did not have an impact on the overall effectiveness of network-based intervention strategies, regardless of the initial seeding strategy. Nonetheless, average path length of the network did have an impact on the overall effectiveness of the intervention. Network-based intervention strategies, especially, randomly targeted interventions, were more effective on networks with shorter average path length (i.e. small-world network vs ER network). (**Table 4-3**, **Figure 4-4** and **Figure 4-5**)

Experiment 3: Effect of individual incentive on outcomes of network-based intervention strategies – from simulation to application

Previous results have shown that regardless of network structural characteristics, targeting most connected individuals within the social network as initial adopters of intervention is more effective than random targeting. Moreover, individual level heterogeneity in opinion towards smoking has an impact on population level behavioral patterns in the long run. Given these two aspects and prior findings that average path length has more effect on intervention outcomes than global clustering of the network, we simulated experiments on an ER network and a small world network with varying distributions of agents' inertia to adopt new behaviors. Results from simulations showed that an individual incentive does have an impact on network-based intervention outcomes. Given the random seeding strategy, largely incentivizing individuals, which was approximated with an inertia distribution of β (2,5), is more effective on an ER network, which has a longer average path length than a small world network. (**Table 4-4**, **Figure 4-6**) To our surprise, the effect of heterogeneity in an agent's inertia for opinion updating on prevalence of smoking differs by local structural characteristics. For a highly clustered network such as a small-world network, it appears that the directionality of effect of incentivizing individuals to alter their individual inertia was not directly correlated with intervention outcomes. However, incentivizing agents was particularly effective for a random network. Such finding provides helpful insights for future intervention design. Firstly, it is extremely challenging to accurately identify most connected individuals in the population, which makes network-based random seeding strategy more realistic than targeting most connected individuals. Moreover, ER networks are more prevalent in the real-world as compared to small-world networks. Given these conditions, when data on network characteristics are not available, it might be more effective to provide individual-targeting incentivizing interventions such as individual counselling.

Discussion

Using agent-based computational experiments, our study incorporated individual decision making and diffusion over social networks into an established social learning model to evaluate how individual decision making and diffusion over social networks collectively affect population-level behavioral outcomes and behavioral intervention effectiveness. Results from our study suggested that individual decision making and social network structural characteristics are both critical to shaping of population level behavior patterns as well as intervention effectiveness. Ensuring community's openness to adopt novel opinions alone can lead to reduced prevalence of harmful behaviors. When taking into consideration the presence of social network, presence of highly clustered network does not imply more necessity or higher effectiveness of network-based intervention. Paradoxically, network-based intervention did not appear to be more effective for more clustered network, regardless of the initial seeding strategy. Moreover, in more scattered networks with longer average path length, incentivizing individuals to lower their inertia might be more effective. Findings from our model highlighted the importance of individual inertia in decision making in shaping the outcomes of diffusion of interventions on a population.

Our results showed that a network-based intervention strategy does provide additional advantages, especially when targeting most central individuals as the initial adopters of the intervention. Such finding is consistent with outcomes from various existing network-based online behavioral intervention studies and randomized trials.(174,175) In the Center for Disease and Control (CDC)'s best practice guideline for tobacco cessation programs, community-based interventions have been listed as the core element of successful intervention for tobacco cessation.(176) Even though such interventions did not specify the use of social network, establishment of social norms through facilitation of social networks and the concept of social influence are the fundamental concepts behind these community-based interventions. Various empirical studies have shown the success of such intervention. (177–179) In addition, our simulation suggested that targeting popular individuals at the initiation of intervention would provide optimal outcome as compared to randomly selecting individuals on the social network. This theoretic design approach has been previously incorporated into various popular opinion leader (POL) intervention programs for smoking cessation and CDC-led POL HIV risk behavior interventions. Results from these studies showed that under the framework of intervention diffusion, targeting well-connected individuals and POLs could be effective.(180,181) However, generalizability of these results is limited due to short duration of follow-up of these randomized trial studies.(180,181) The real-world applicability of such theoretic concepts is also hindered by the difficulty in obtaining information regarding most central individuals prior to intervention. This leads us to the discussion of a more important aspect of social network that has an impact on intervention outcomes – network structural characteristics. Our findings showed that average path length, instead of global clustering coefficient, has a significant impact on outcomes of network-based intervention. Most importantly, when taking into consideration individual level inertia, lowering proportions of individuals' inertia could compensate for the weakness of longer average path length, thus leading to more optimal intervention effectiveness. This finding could potentially be explained through amplified network externality through lowered population-level mean inertia, while incentivizing individuals to be more open minded throughout the intervention. Extremely limited numbers of studies have investigated the role of individual decision making or individual inertia in shaping the outcomes of diffusion of intervention.(182–184) However, individual behavioral choice has been widely recognized as the key component that drives behavioral changes in numerous interventions targeting tobacco cessation and physical activity. Prominent examples of these interventions include counter-marketing tobacco cessation programs and Community Healthy Activities Model Program for Seniors.(185),(186) Given the resource-intensive character of randomized trials for interventions that involve altering individual-level behavior choice, extremely few numbers of randomized trials have been in place these days, especially ones that would allow researchers to conduct follow-up studies over an extended period of time. Our study might be able to provide an alternative to explain the success of individual-based incentive programs as compared to group-based programs without the prior knowledge of network structure for smoking cessation.(187)

Our study provided a novel framework in epidemiology to identify potential effective design strategies for behavioral intervention. However, several limitations of the study need to be acknowledged. First, due to the nature of the study, our model had several strong assumptions regarding adoption of network-based intervention. We assumed that conversion probability from smoker to non-smoker based on network-based intervention is different from that of altering individual's opinion towards smoking. However, model setup in this way allowed us to differentiate the effect of diffusion on behavioral change versus the effect of altering the opinion on behavioral change. Second, due to limited data availability, we were not able to calibrate the model to the real-world scenario. Nonetheless, to our knowledge, this is one of the first studies that take into consideration both individual decision making and network-based diffusion when evaluating effectiveness of behavioral interventions. More importantly, the hypothesis-generating nature of this model provided several implications for future study directions and intervention design. First, the effectiveness of the behavioral intervention not only relies on the intervention alone. Baseline prevalence of the behavior in the population is also an important determinant of intervention outcomes, especially when social influence and social learning play an important role in behavioral formation. Second, collecting information on changes in individual attitude towards behavior is crucial to estimating the effect of network-based interventions on behavioral outcomes. Last, with unknown network structure, especially highly clustered areas, individual-based intervention, such as providing incentives for individuals, might be a more cost-effective alternative.

Conclusion

To conclude, our study indicates both individual behavioral choice and social network structural characteristics are key to design of behavioral interventions. More importantly, with limited resources and unknown network structure, it might be more cost-effective to carry out individual-targeting interventions.

Table 4-1. Model Parameter Definition for Computational Experiments

Model Parameter	
Agent Attribute	
Behavior Status	Binary (smoker, non-smoker)
Opinion towards Smoking (δ)	Integer $\in$ [1,10], sampling distribution determined by initial prevalence of behavior; Higher value translates to higher preference for smoking.
Inertia for opinion updating (ω)	Numeric variable $\in$ (0, 1), sampling distribution determined by intervention type. Higher value translates to higher the inertia is for opinion updating.
Network attribute	
Network structure	Pre-specified network at model initiation
Population-level	
Initial prevalence of behavior ($prev_0$)	Initial prevalence of smoking at t_0

Table 4-2. Prevalence of Smoking and Mean Opinion for Baseline Simulation

Experimental Condition	Final Prevalence of Smoking
Condition 1. (Figure 4-1A, Figure 4-1B)	
$prev_0$= 0.50, initial mean opinion = 5.32	1.00
Condition 2 (Figure 4-1C, Figure 4-1D)	
$prev_{0=}$ 0.80, initial mean opinion = 6.98	1.00
Condition 3 (Figure 4-1E, Figure 4-1F)	
$prev_0$= 0.20, initial mean opinion = 3.72	0.17
Baseline Simulation Network Characteristics: Global clustering coefficient: 0.26, Average Path Length: 7	

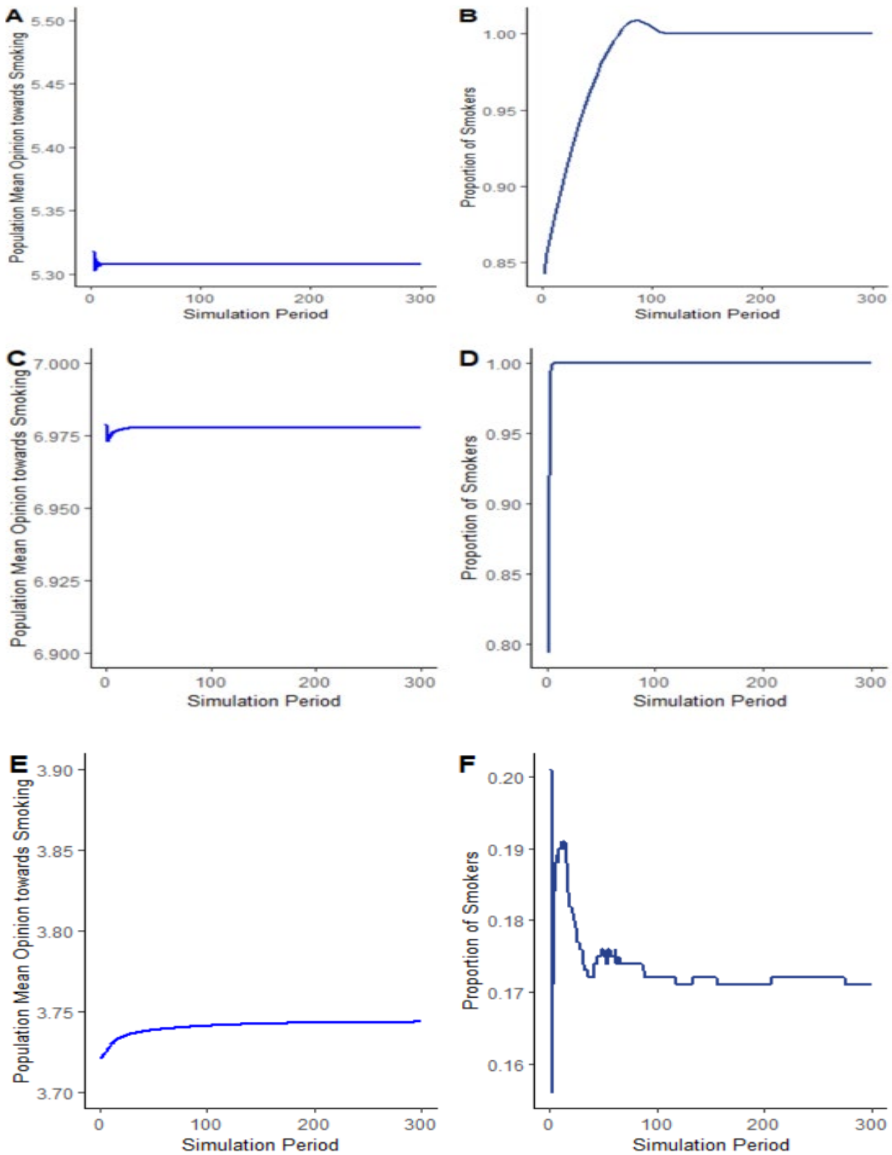


Figure 4-1. Baseline Scenario

Fig 4-1A): Population Mean of Opinion towards Smoking Behavior over Simulation Period($prev_0$ = 0.5)
Fig 4-1B): Prevalence of Smoker over Simulation Period($prev_0$ = 0.5)
Fig 4-1C): Population Mean of Opinion towards Smoking Behavior over Simulation Period($prev_0$ = 0.8)
Fig 4-1D): Prevalence of Smoker over Simulation Period($prev_0$ = 0.8)
Fig 4-1E): Population Mean of Opinion towards Smoking Behavior over Simulation Period($prev_0$ = 0.2)
Fig 4-1F): Prevalence of Smoker over Simulation Period($prev_0$ = 0.2)

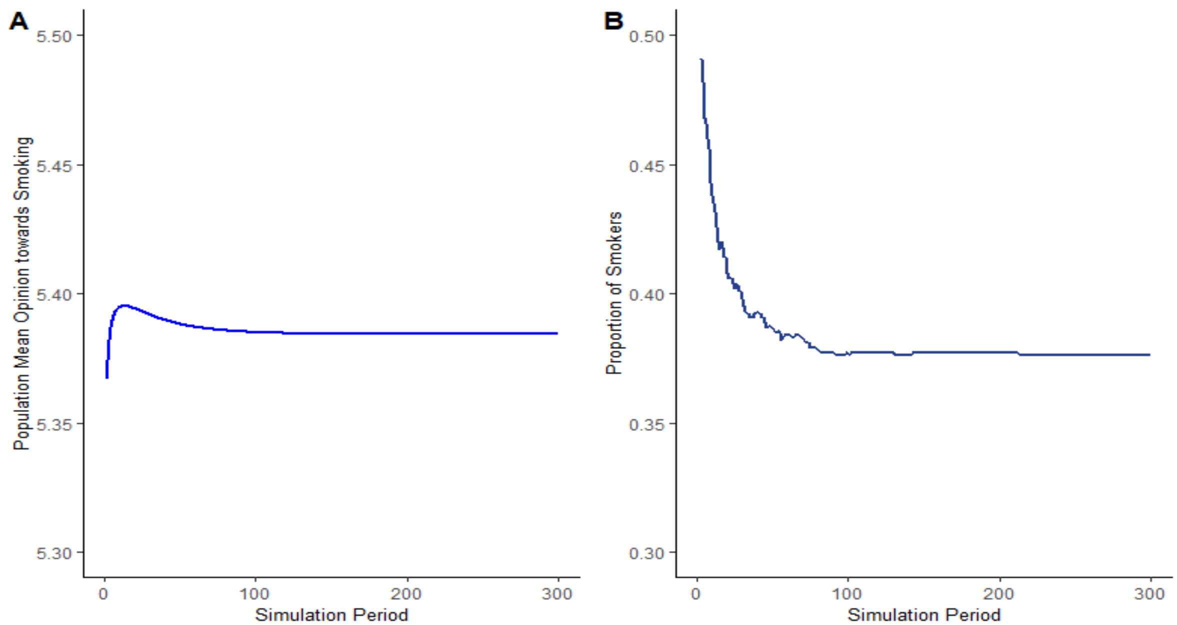


Figure 4-2. Computational Experiment with Individual Inertia for Opinion Updating Following Sampling distribution of β (3,3), $prev_0$ = 0.50

Fig 4-2A): Population Mean of Opinion towards Smoking Behavior over Simulation Period
Fig 4-2B): Prevalence of Smoker over Simulation Period

Table 4-3. Prevalence of Smoking with Network-based Intervention and No Alteration on Individual Incentive

Network structure	Seeding Strategy	Final Prevalence of Smoking
No Network-based Intervention (Fig 4-3A)	--	0.38
Small-world Network (Fig 4-3B & Fig 4-3C) Global Clustering Coefficient = 0.26	Random Seeding	0.36
	Highest Degree	0.33
Small-world Network (Fig 4-5A & Fig 4-5B) Global Clustering Coefficient = 0.05	Random Seeding	0.37
	Highest Degree	0.32
Erdos-Renyi Network (Fig 4-4A & Fig 4-4B)	Random Seeding	0.42
	Highest Degree	0.35

**p refers to probability of tie formation in the network*
**Number of initial adopters of intervention = 100 for all simulations*
**For all simulations, individual inertia follows sampling distribution of β (3,3), initial prevalence of smoking - $prev_0$ = 0.50*

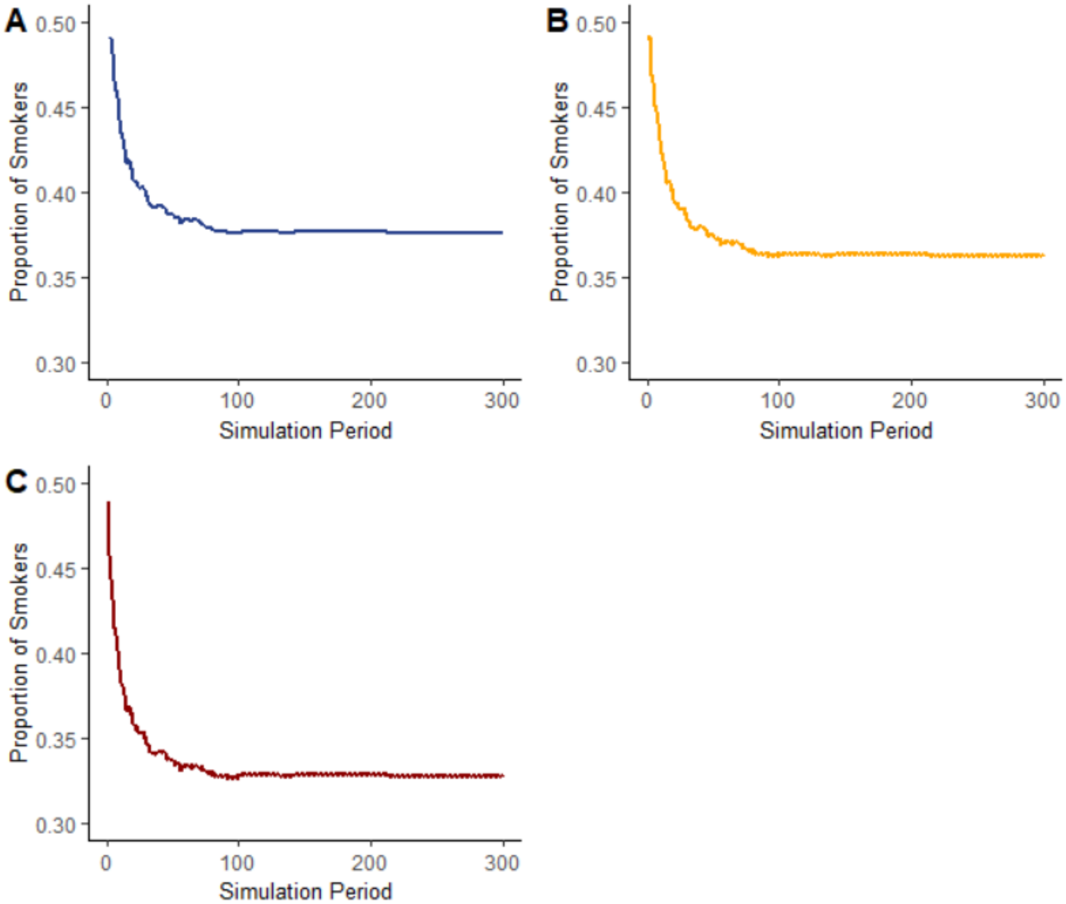


Figure 4-3. Network-based Intervention with Individual Inertia for Opinion Updating Following Sampling Distribution of β (3,3), $prev_0$ = 0.50, Varying Initial Seeding Strategy

Fig 4-3A): No network-based intervention diffusion
Fig 4-3B): Small-world network with 10% randomly selected initial seeding population on the network
Fig 4-3C): Small-world network with with top 10% of individuals ranked based on degree centrality selected as initial seeding population

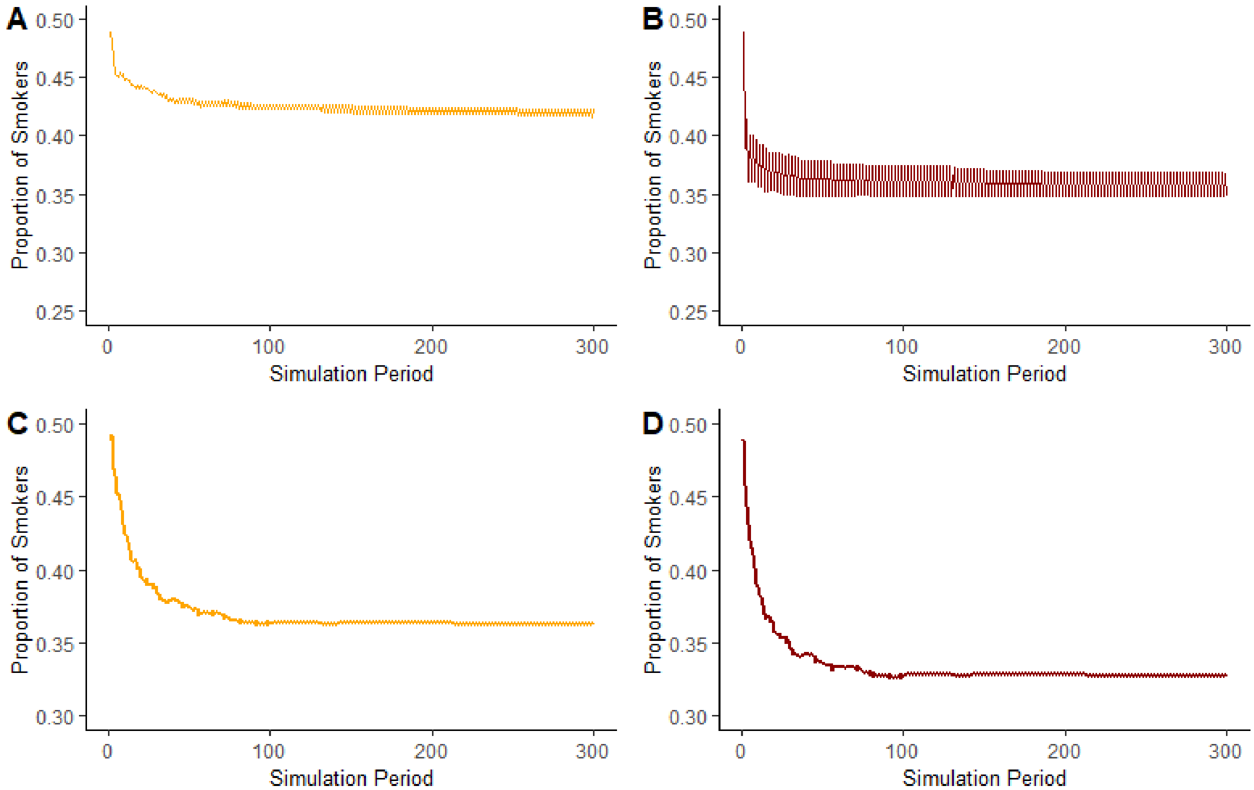


Figure 4-4. Network-based Intervention with Individual Inertia for Opinion Updating Following Sampling distribution of β (3,3), $prev_0$ = 0.50, Varying Network Average Path Length

Fig 4-4A): Erdos-Renyi network with 10% randomly selected initial seeding population on the network
Fig 4-4B): Erdos-Renyi network with with top 10% of individuals ranked based on degree centrality selected as initial seeding population
Fig 4-4C): Small-world network with 10% randomly selected initial seeding population on the network
Fig 4-4D): Small-world network with with top 10% of individuals ranked based on degree centrality selected as initial seeding population

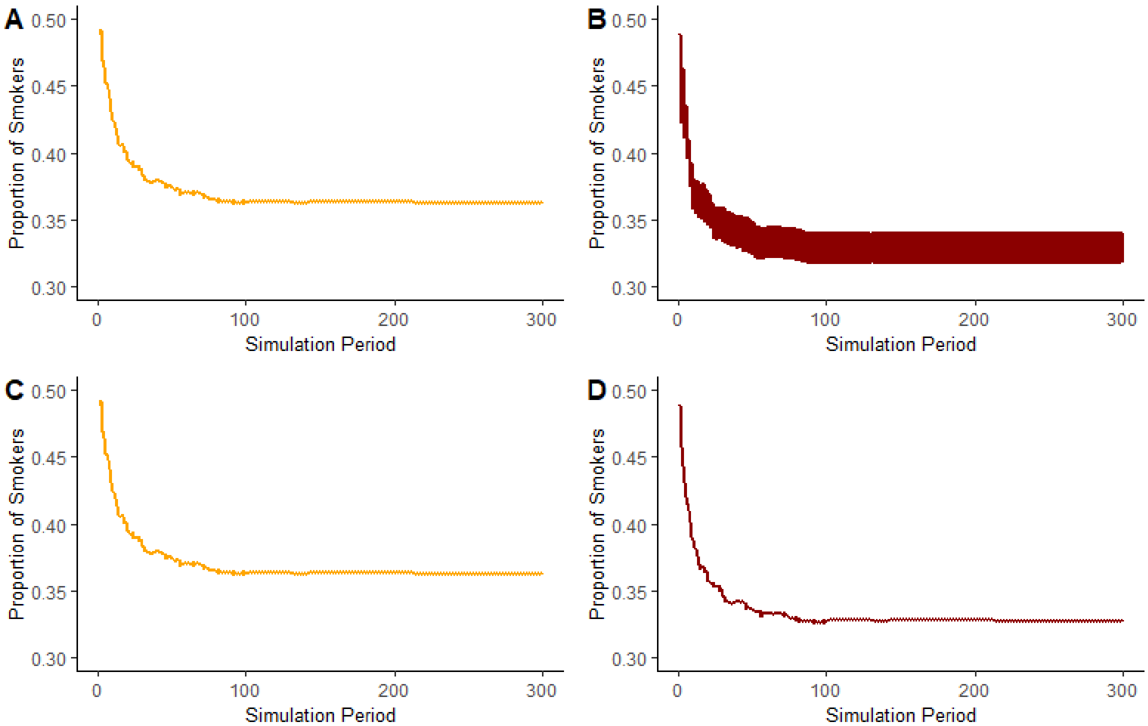


Figure 4-5. Network-based Intervention with Individual Inertia for Opinion Updating Following Sampling Distribution of β (3,3), $prev_0$ = 0.50, Varying Network Global Clustering Coefficient

Fig 4-5A): Small-world network with 10% randomly selected initial seeding population, network global clustering = 0.05
Fig 4-5B): Small-world network with with top 10% of individuals ranked based on degree centrality selected as initial seeding population, network global clustering = 0.05
Fig 4-5C): Small-world network with 10% randomly selected initial seeding population, network global clustering coefficient = 0. 26
Fig 4-5D): Small-world network with with top 10% of individuals ranked based on degree centrality selected as initial seeding population, network global clustering coefficient = 0. 26

Table 4-4. Prevalence of Smoking with a Network-based Intervention and Random Seeding Strategy, Varying Network Average Path Length and Sampling Distribution of Individual Inertia for Opinion Updating

	Sampling Distribution of ω	Final Prevalence of Smoking
Small-world Network (p=0.1) (Figure 4-6A, 4-6B & 4-6C)	*β (3,3)*	0.36
	β (2,5)	0.41
	β (8,2)	0.31
Erdos-Renyi Network (p=0.1) (Figure 4-6D, 4-6E & 4-6F)	*β (3,3)*	0.42
	β (2,5)	0.35
	β (8,2)	0.42

***p refers to probability of tie formation in the network*
**Number of initial adopters of intervention = 100 for all simulations*
**For all simulations, initial prevalence of smoking - $prev_0$ = 0.50, seeding strategy for network-based intervention follows random seeding of initial adopters of intervention.*

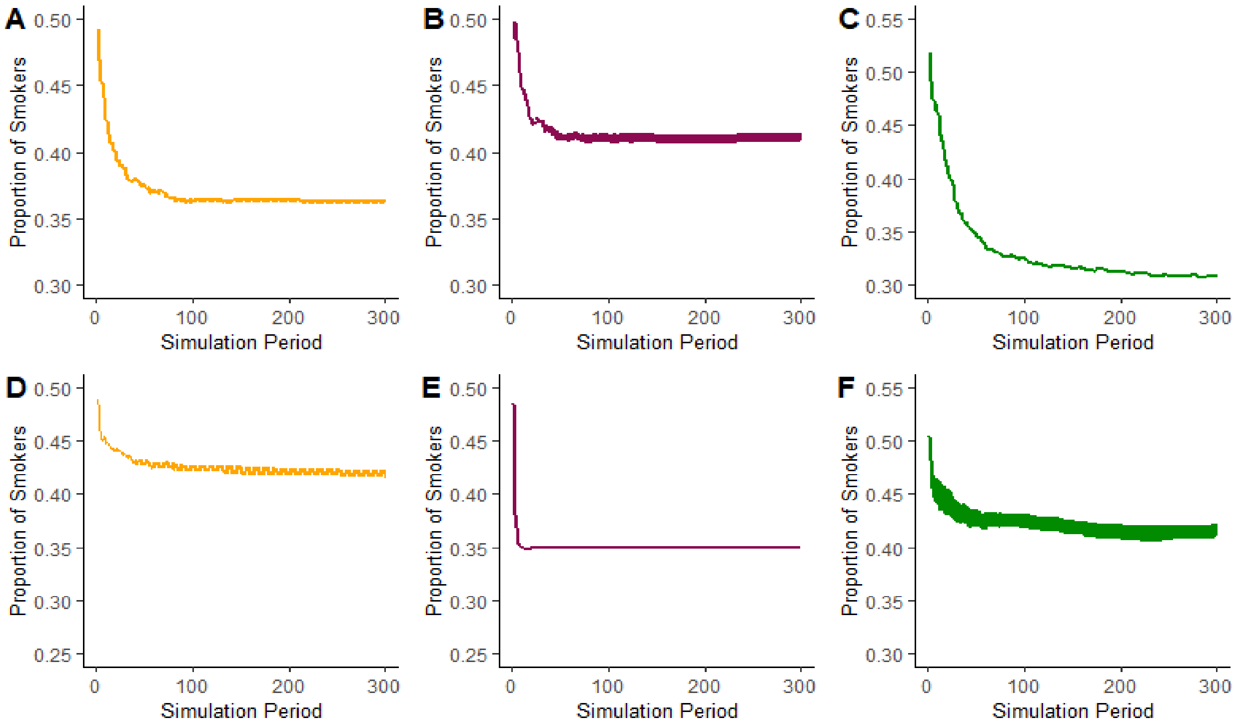


Figure 4-6. Computational Experiment with Individual Incentive Altering: Altering Structure, Random Seeding, 50% Initial Prevalence

Fig 4-6A): Small-world network (p = 0.1) with 10% randomly selected initial seeding population on the network, $\omega \sim \beta\ (3,3)$
Fig 4-6B): Small-world network (p = 0.1) with 10% randomly selected initial seeding population on the network, $\omega \sim \beta\ (2,5)$
Fig 4-6C): Small-world network (p = 0.1) with 10% randomly selected initial seeding population on the network, $\omega \sim \beta\ (8,2)$
Fig 4-6D): Erdos-Renyi network (p = 0.1) with 10% randomly selected initial seeding population on the network, $\omega \sim \beta\ (3,3)$
Fig 4-6E): Erdos-Renyi network (p = 0.1) with 10% randomly selected initial seeding population on the network, $\omega \sim \beta\ (2,5)$
Fig 4-6F): Erdos-Renyi network (p = 0.1) with 10% randomly selected initial seeding population on the network, $\omega \sim \beta\ (8,2)$

Chapter 5: Public Health Implications and Future Research Direction

Public Health Implications

Given the recognized importance of modifiable health behaviors in cardiometabolic disease prevention and challenges involved in designing effective behavioral interventions, this dissertation empirically identified multi-level risk factors associated with cigarette smoking and physical activity, as well as trajectories of both behaviors throughout life-course. In addition, this dissertation developed a novel framework to evaluate population level outcomes of different behavioral interventions.

Aim 1 of this dissertation used latent class growth mixture model that allowed identification of subgroups in the population, given individual level heterogeneity. Results showed that there are three subgroups of individuals sharing similar patterns of physical activity and past 30-day cigarette smoking behavior from early adolescence to adulthood. Age, socio-demographic and early-life psychological factors are all important predictors of trajectories for both behaviors. Findings from Aim 1 showed that multi-level risk factors collectively shape individual's health behavioral pattern throughout the entire life course. More importantly, early-life intervention might have an enduring impact on individual health behaviors, even later in life. Additionally, results from this aim implied that future behavioral interventions targeting physical activity and cigarette smoking behaviors need to take into consideration both timing and target population characteristics to be effective.

Aim 2 of this dissertation used social network analysis and regression methods to evaluate the association between social network characteristics and physical activity/cigarette smoking behaviors during adolescence and the adolescence to young adulthood transition. Results from our study showed that individuals' health behaviors at younger age are the strongest predictors of health behaviors during young adulthood. In addition, an individual's social position during adolescence is a predictor for physical activity level during young adulthood but not for cigarette smoking. Consistent with findings from Aim 1, results from Aim 2 of this dissertation suggested the importance of early life behavioral interventions for promoting healthy behaviors later in life. Moreover, our findings showed that social network characteristics, especially individual's social position, have lasting impact on individual's health behavior.

Using computational models, aim 3 of this dissertation successfully recapitulated the empirical findings in Aim 2, showing that network-level clustering, whether it be local or global, does not impact long-term effectiveness of behavioral intervention. Meanwhile, average path length of the network, an indicator of how reachable individuals are on a social network as well as individuals' behavioral incentive are key to behavioral intervention outcomes. Ensuring community's openness to adopt novel opinions alone can lead to reduced prevalence of harmful behaviors. When taking into consideration diffusion of intervention-related information within a network, a highly clustered network does not imply more necessity of network-based intervention. Paradoxically, network-based interventions did not appear to be more effective for more clustered network, regardless of whether popular opinion leaders were targeted. Moreover, in more scattered networks with longer average path length or unknown network structure, incentivizing individuals might be more effective.

Expanding from existing simulation and observational studies on the effect of intervention on modifiable health behaviors as well as the effect of social network on behavior outcomes, results from this dissertation highlighted the importance of both social network and individual-level heterogeneity are both key to shaping the population level distributions of modifiable health behaviors and behavioral intervention outcomes. In addition, this dissertation highlighted the possibilities that complex systems methods such as agent-based modeling can bring.

Epidemiology has long been identified as a key source to inform the decision making process in public health.(188),(189) To assist with the decision-making process, observational studies are often conducted on effectiveness of intervention strategies by comparing population outcomes at the initiation of intervention and post-intervention.(190,191) Over the course of its development, the definition of Epidemiology has largely extended from its very original formal definition of “a discipline studying the population level distribution of determinants of health”.(192) In the recent decade, increasing amount of research has taken on a consequentialist approach to epidemiology, which focuses on placing more emphasis on maximizing the desired outcome of interest and aiming to bridge the gap between observational studies and implementation science.(193) As a result, the number of studies using a simulation study approach such as agent-based modeling has drastically increased in the past five years or so.(193) In epidemiologic studies that apply agent-based modeling to evaluate intervention outcomes on a social network, agent-level behavioral heterogeneity is often handled as the following: a. match agent attribute to real-world data such as race/ethnicity b. allow for node-level agent behavior change such as formation and dissolution of network ties c. set-up probabilistic behavior transition rules at each time step, which assimilates a Markov process at individual level, as a function of agent attribute specified from real world data. Even though these modeling approaches would allow for recapitulation of real-world setting and network-level heterogeneity, they often overlook the mechanistic hypotheses linking individual level behaviors to population-level outcomes. Models and results from this dissertation showed that in addition to recapitulating real-world scenarios and forecasting, agent-based models are well suited to explore research questions in behavioral epidemiology and facilitate decision making through incorporation of social science theories into agent behavioral rules. Such bottom-up model process would be particularly valuable when research questions are dependent on strict assumptions on individual-level interactions or variables that are difficult to measure during the study design and data collection phase.

Collectively, results from this dissertation emphasize a need to embrace a complex systems lens incorporating individual heterogeneity in behavioral choice, especially when addressing research question involving behavioral determinants of health.

Future Research Direction

Given the highlighted importance of behavior choice in shaping population level distributions of health behaviors and intervention outcomes targeting health behaviors, future research in epidemiology should take into consideration behavioral choice throughout research question design, data collection, study design and analyses stages. Individual level measures such as attitudes towards behavioral outcome of interest can be collected and analyzed as a proxy to individual behavioral choice in epidemiologic studies.

Methodologically, future research to evaluate behavioral intervention outcomes could further explore different assumptions on individual decision making, given different network topologies to address questions. For example, analytical game theoretical approaches can be adopted in the future where we analyze the dynamics of adolescent health behavior as a Bayesian game with incomplete information on the network. The model presented in this dissertation can also be additionally adapted to introduce more agent-level heterogeneity such as race/ethnic profile to explore whether introduction of certain types of intervention on a given network would reduce or exacerbate outcome disparities in the population.

Funding

Data used in this dissertation came from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a study designed and led by teams at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the National Institute of Child Health and Human Development.

This dissertation was partially funded by the Center for the Assessment of Tobacco Regulations at University of Michigan (U54CA229974).

References

1. UN General Assembly. Political Declaration of the High-level Meeting of the General Assembly on the Prevention and Control of Non-communicable Diseases.
2. Principled Promotion of Health: Implementing Five Guiding Health Promotion Principles for Research-Based Prevention and Management of Diabetes. Societies. 2017;7(2):10.
3. U S Department of Health and Human Services. HHS News - Press Release. 2010;(202):1–2.
4. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics—2020 update: A report from the American Heart Association. Circulation. 2020. 139–596 p.
5. American Heart Association. Cardiovascular Disease: A Costly Burden For American - Projection Through 2035. 2017;
6. Gurenlian JR. The power of prevention. Int J Dent Hyg. 2014;12(1):1–1.
7. Buchan DS, Ollis S, Thomas NE, Baker JS. Physical activity behaviour: An overview of current and emergent theoretical practices. J Obes. 2012;2012.
8. Simpson V. Models and Theories to Support Health Behavior Intervention and Program Planning. Heal Hum Sci. 2015;1–5.
9. Kotier P, Zaltman G. Social marketing: An approach to planned social change. Soc Mar Q. 1996;3(3–4):7–20.
10. Kahn EB, Ramsey LT, Brownson RC, Heath GW, Howze EH, Powell KE, et al. The Effectiveness of Interventions A Systematic Review. 2010;22(02).
11. Winkleby MA, Taylor B, Jatulis D, Fortmann SP. The long-term effects of a cardiovascular disease prevention trial: The stanford five-city project. Am J Public Health. 1996;86(12):1773–9.
12. Farquhar JW, Fortmann SP, Maccoby N, Haskell WL, Williams PT, Flora JA, et al. The stanford five-city project: design and methods. Am J Epidemiol. 1985 Aug 1;122(2):323–34.
13. Stanford T, Farquhar JW, Fortmann SP, Flora JA, Taylor CB, Haskell WL, et al. Effects of Communitywide Education on Cardiovascular Disease Risk Factors Five-City Project. 2015;(5).
14. Carleton RA, Lasater TM, Assaf AR, Feldman HA, McKinlay S. The Pawtucket Heart Health Program: community changes in cardiovascular risk factors and projected disease risk. Am J Public Health. 1995 Jun;85(6):777–85.
15. Eaton CB, Lapane KL, Garber CE, Gans KM, Lasater TM, Carleton RA. Effects of a community-based intervention on physical activity: The Pawtuckett Heart Health Program. Am J Public Health. 1999;89(11):1741–4.
16. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int. 2014 Oct 15;30(1):174–83.
17. Cooper RN. Living with global imbalances: A contrarian view. Vol. 28, Journal of Policy Modeling. 2006. 615–627 p.
18. Chaddha A, Jackson EA, Richardson CR, Franklin BA. Technology to Help Promote Physical Activity. Am J Cardiol. 2017;119(1):149–52.
19. Baranowski T, Abdelsamad D, Baranowski J, O'Connor TM, Thompson D, Barnett A, et al. Impact of an active video game on healthy children's physical activity. Pediatrics. 2012;129(3).
20. Doll R, Bradfordhill A. Themortalityofdoctorsin Relation Totheirsmokinghabits Apreliminaryreport. Br Med J. 1954;1(4877):1451–5.
21. Hopkins DP. Recommendations regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke. Am J Prev Med. 2001;20(2 SUPPL.):10–5.
22. Schlam TR, Baker TB. Interventions for tobacco smoking. Vol. 9, Annual Review of Clinical Psychology. 2013. 675–702 p.

23. Chang SS. Re: Smoking Cessation: A Report of the Surgeon General. J Urol. 2020;
24. Stratton K, Shetty P, Wallace R, Bondurant S, Institute of Medicine (US) Committee to Assess the Science Base for Tobacco Harm Reduction. Tobacco Smoke and Toxicology. Clearing the Smoke: Assessing the Science Base for Tobacco Harm Reduction . 2001. 283–308 p.
25. Department of Health. Introduction, Conclusions, and Historical Background Relative to E-Cigarettes (in E-Cigarette Use Among Youth and Young Adults: A Report of the Surgeon General). Dep Heal Us Serv Hum Dis Control Centers Cent Chronic Dis Prev Natl Promot Heal Smoking, Off. 2016;1–24.
26. Epidemic TE, Youth A. Surgeon General ' s Advisory on E-cigarette Use Among Youth The E-cigarette Epidemic Among Youth E-cigarettes Come in Many Shapes and Sizes You Can Take Action. 2018;2–5.
27. Werner AK, Koumans EH, Chatham-Stephens K, Salvatore PP, Armatas C, Byers P, et al. Hospitalizations and deaths associated with EVALI. N Engl J Med. 2020;382(17):1589–98.
28. Huang J, Duan Z, Kwok J, Binns S, Vera LE, Kim Y, et al. Vaping versus JUULing: How the extraordinary growth and marketing of JUUL transformed the US retail e-cigarette market. Tob Control. 2019;28(2):146–51.
29. Marynak K, Gentzke A, Wang TW, Neff L, King BA. Exposure to Electronic Cigarette Advertising Among Middle and High School Students — United States, 2014–2016. MMWR Morb Mortal Wkly Rep. 2018;67(10):294–9.
30. East KA, Hitchman SC, McNeill A, Thrasher JF, Hammond D. Social norms towards smoking and vaping and associations with product use among youth in England, Canada, and the US. Drug Alcohol Depend. 2019;205(May):107635.
31. FDA. Public Law 111-31-June 22,2009. Family smoking prevention and tobacco control and federal retirement reform. Public Law. 2009;111(22):1–84.
32. Ford P, Clifford A, Gussy K, Gartner C. A systematic review of peer-support programs for smoking cessation in disadvantaged groups. Int J Environ Res Public Health. 2013;10(11):5507–22.
33. Mermelstein R. Teen smoking cessation. Tob Control. 2003;12(SUPPL. I):25–34.
34. National Prevention Council. National prevention strategy. Natl Prev Strateg. 2011;(June).
35. Bauer JE, Hyland A, Li Q, Steger C, Cummings KM. A longitudinal assessment of the impact of smoke-free worksite policies on tobacco use. Am J Public Health. 2005;95(6):1024–9.
36. American Heart Association. Reducing Sugar-Sweetened Beverage Consumption: A Focus on Sugar-Sweetened Beverage Taxes. 2016;(June):1–9.
37. Liverani M, Hawkins B, Parkhurst JO. Political and institutional influences on the use of evidence in public health policy. A systematic review. PLoS One. 2013;8(10).
38. Levy DT, Tam J, Kuo C, Fong GT, Chaloupka F. The Impact of Implementing Tobacco Control Policies : The 2017 Tobacco Control Policy Scorecard. 2018;24(5):448–57.
39. Laverack G. The Challenge of Behaviour Change and Health Promotion. Challenges. 2017 Oct 17;8(2):25.
40. Visscher TLS, Bell C, Gubbels JS, Huang TTK, Bryant MJ, Peeters A, et al. Challenges in lifestyle and community interventions research; a call for innovation. BMC Obes. 2014;1(1):29–32.
41. Mittelmark MB, Hunt MK, Heath GW, Schmid TL. Realistic Outcomes: Lessons from Community-Based Research and Demonstration Programs for the Prevention of Cardiovascular Diseases. J Public Health Policy. 1993;14(4):437.
42. Nowak D. Smoking cessation in groups — who benefits in the long term ? 2013;28(5):869–78.

43. Durkin S, Brennan E, Wakefield M. Mass media campaigns to promote smoking cessation among adults: An integrative review. Tob Control. 2012;21(2):127–38.
44. Abioye AI, Hajifathalian K, Danaei G. Do mass media campaigns improve physical activity? a systematic review and meta-analysis. Arch Public Heal. 2013 Dec 2;71(1):20.
45. Yun L, Ori EM, Lee Y, Sivak A, Berry TR. A Systematic Review of Community-wide Media Physical Activity Campaigns: An Update From 2010. J Phys Act Heal. 2017 Jul;14(7):552–70.
46. O'Reilly GA, Spruijt-Metz D. Current mHealth technologies for physical activity assessment and promotion. Am J Prev Med. 2013 Oct;45(4):501–7.
47. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth Physical Activity Intervention: A Randomized Pilot Study in Physically Inactive Pregnant Women. Matern Child Health J. 2016 May;20(5):1091–101.
48. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: A Randomized Clinical Trial of an Automated mHealth Intervention for Physical Activity Promotion. J Am Heart Assoc. 2015 Nov 9;4(11).
49. Harrison RA, McNair F, Dugdill L. Access to exercise referral schemes - A population based analysis. J Public Health (Bangkok). 2005;27(4):326–30.
50. Arsenijevic J, Groot W. Physical activity on prescription schemes (PARS): do programme characteristics influence effectiveness? Results of a systematic review and meta-analyses. BMJ Open. 2017;7(2):e012156.
51. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the Use of Mobile Health Technology by Patients: An Observational Study in Primary Care Clinics. JMIR mHealth uHealth. 2016 Apr 19;4(2):e41.
52. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. Stat Med. 2013 Feb 20;32(4):556–77.
53. Ritter C. Social supports, social networks, and health behaviors. Health behavior: Emerging research perspectives. New York, NY, US: Plenum Press; 1988. p. 149–61.
54. Valente TW, Pitts SR. An Appraisal of Social Network Theory and Analysis as Applied to Public Health: Challenges and Opportunities. Annu Rev Public Health. 2017 Mar 20;38(1):103–18.
55. McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. Annu Rev Sociol. 2001;27:415–44.
56. Ball K, Jeffery RW, Abbott G, Mcnaughton SA, Crawford D. Is healthy behavior contagious : associations of social norms with physical activity and healthy eating. Int J Behav Nutr Phys Act. 2010;7(1):86.
57. Huang GC, Ph D, Unger JB, Ph D, Soto D, H MP, et al. Peer Influences : The Impact of Online and Of fl ine Friendship Networks on Adolescent Smoking and Alcohol Use. 2014;54:508–14.
58. Centola D. The Social Origins of Networks and Diffusion. AJS. 2015 Mar;120(5):1295–338.
59. Guilbeault D, Becker J, Centola D. Complex Contagions: A Decade in Review. In 2018. p. 3–25.
60. Centola D. How Behavior Spreads: The Science of Complex Contagions. Vol. 361, Science. 2018. 1320.2-1320.
61. Lelarge M. Diffusion and cascading behavior in random networks. Games Econ Behav. 2012 Jul 9;75(2):752–75.
62. Easley D, Kleinberg J. Cascading Behavior in Networks. Networks, Crowds, Mark Reason about a Highly Connect World. 2010;563--609.
63. Kershaw KN, Albrecht SS. Racial/ethnic residential segregation and cardiovascular disease risk. Curr Cardiovasc Risk Rep. 2015;9(3):1–19.
64. Mayne SL, Hicken MT, Merkin SS, Seeman TE, Kershaw KN, Do DP, et al.

Neighbourhood racial/ethnic residential segregation and cardiometabolic risk: the multiethnic study of atherosclerosis. J Epidemiol Community Health. 2019 Jan 1;73(1):26 LP – 33.
65. de la Sablonnière R. Toward a Psychology of Social Change: A Typology of Social Change. Front Psychol. 2017 Mar 28;8:397.
66. Rice T. The Behavioral Economics of Health and Health Care. Annu Rev Public Health. 2013 Mar 18;34(1):431–47.
67. Barkley-Levenson EE, Van Leijenhorst L, Galván A. Behavioral and neural correlates of loss aversion and risk avoidance in adolescents and adults. Dev Cogn Neurosci. 2013;3(1):72–83.
68. Granovetter M. Threshold Models of Collective Behavior. Am J Sociol. 1978;83(6):1420–43.
69. Tian Q. The Complex Systems Approach to Policy Analysis. In 2017. p. 123–42.
70. Wallace R, Geller A, Ogawa VA, editors. Assessing the Use of Agent-Based Models for Tobacco Regulation. Washington, D.C.: National Academies Press; 2015.
71. Chan S. Complex Adaptive Systems. ESD83 Res Semin Eng Syst. 2001;31:1–9.
72. Jackson JC, Rand D, Lewis K, Norton MI, Gray K. Agent-Based Modeling. Soc Psychol Personal Sci. 2017 May 13;8(4):387–95.
73. Nowak S, Matthews L, Parker A. A General Agent-Based Model of Social Learning. RAND Corporation; 2017.
74. Rahmandad H, Sterman J. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. Manage Sci. 2008;54(5):998–1014.
75. Bruch E, Atwell J. Agent-Based Models in Empirical Social Research. Sociol Methods Res. 2015 May 24;44(2):186–221.
76. Fletcher JM, Ross SL. Estimating the effects of friends on health behaviors of adolescents. Health Econ. 2018;27(10):1450–83.
77. Card D, Giuliano L. Peer Effects and Multiple Equilibria in the Risky Behavior of Friends. Rev Econ Stat. 2013 Oct 1;95(4):1130–49.
78. Ali MM, Dwyer DS. Estimating peer effects in adolescent smoking behavior: A longitudinal analysis. J Adolesc Heal. 2009;45(4):402–8.
79. Forouzanfar MH, Afshin A, Alexander LT, Biryukov S, Brauer M, Cercy K, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1659–724.
80. Jackson SE, Brown J, Ussher M, Shahab L, Steptoe A, Smith L. Combined health risks of cigarette smoking and low levels of physical activity : a prospective cohort study in England with 12-year follow-up; BMJ-open, 2019 Nov 27;9(11):e032852.
81. Andrew T, Stephen R, Roger C. Smoking and Physical Activity : A Systematic Review. 2008;
82. Audrain-mcgovern J, Rodriguez D, Moss HB. Smoking Progression and Physical Activity. 2010;12(November 2003):1121–9.
83. Conway TL. Smoking , and Physical Fitness ’. 1992;734:723–34.
84. Dijkstra A, Vries H De, Kok G, Rouackers J. Self-Evaluation and Motivation To Change : Social Cognitive Constructs In Smoking Cessation Self-evaluation And Motivation to Change: Social Cognitive Constructs in Smoking Cessation. 2007; 0446
85. Bidstrup PE, Frederiksen K, Siersma V, Mortensen EL, Ross L, Vinther-larsen M, et al. Social-Cognitive and School Factors in Initiation of Smoking among Adolescents : A Prospective Cohort Study. 2009;18(February):384–93.
86. Young MD, Plotnikoff RC, Collins CE, Callister R, Morgan PJ. Social cognitive theory and physical activity a systematic review and meta-analysis.

87. Golden SD, Earp JAL. Social Ecological Approaches to Individuals and Their Contexts : Twenty Years of Health Education & Behavior Health Promotion Interventions. 2012;
88. King JL, Merten JW, Wong T, Pomeranz JL. Applying a Social – Ecological Framework to Factors Related to Nicotine Replacement Therapy for Adolescent Smoking Cessation. 2018;32(5):1291–303.
89. Anette M, Mehtälä K, Sääkslahti AK, Inkinen ME, Eija M, Poskiparta H. A socio-ecological approach to physical activity interventions in childcare : a systematic review. 2014;
90. Haasova M, Warren FC, Ussher M, Rensburg KJ Van, Faulkner G, Cropley M, et al. The acute effects of physical activity on cigarette cravings : systematic review and meta-analysis with individual participant data. 2012;26–37.
91. Benadjaoud MA, Menai M, Hees VT Van, Zipunnikov V. The association between accelerometer-assessed physical activity and respiratory function in older adults differs between smokers and non-smokers. 2019;(February):1–9.
92. Dugral E. Effects of smoking and physical exercise on respiratory function test results in students of university. 2003;2003.
93. Bricker JB, Comstock BA, Peterson A V, Kealey KA, Marek PM. Social Cognitive Mediators of Adolescent Smoking Cessation: Results from a Large Randomized Intervention Trial Jonathan. 2011;24(3):436–45.
94. Ekkekakis P, Zenko Z. Escape From Cognitivism: Exercise as Hedonic Experience." Sport and Exercise Psychology Research. 2016. 389–414 p.
95. Stults-kolehmainen MA, Blacutt M, Bartholomew JB, Gilson TA, Ash GI, Mckee PC, et al. Motivation States for Physical Activity and Sedentary Behavior : Desire , Urge , Wanting , and Craving. 2020;11(November):1–17.
96. Piercy KL. Recent Trends in Adherence of Physical Activity and Sedentary Behavior — We Need to Move More and Sit Less. 2022;2(7):10–2.
97. Achttien RJ, van Lieshout J, Wensing M, Nijhuis-Van der Sanden M, Staal JB. The decline in physical activity in aging people is not modified by gender or the presence of cardiovascular disease. Eur J Public Health. 2020;30(2):333–9.
98. Kwon S, Janz KF, Letuchy EM, Burns TL, Levy SM. Developmental trajectories of physical activity, sports, and television viewing during childhood to young adulthood: Iowa bone development study. JAMA Pediatr. 2015;169(7):666–72.
99. Howie EK, McVeigh JA, Smith AJ, Zabatiero J, Bucks RS, Mori TA, et al. Physical activity trajectories from childhood to late adolescence and their implications for health in young adulthood. Prev Med (Baltim). 2020;139(April):106224.
100. Lounassalo I, Salin K, Kankaanpaä A, Hirvensalo M, Palomäki S, Tolvanen A, et al. Distinct trajectories of physical activity and related factors during the life course in the general population: A systematic review. BMC Public Health. 2019;19(1):1–12.
101. Gordon-Larsen P, McMurray RG, Popkin BM. Adolescent physical activity and inactivity vary by ethnicity: The National Longitudinal Study of Adolescent Health. J Pediatr. 1999;135(3):301–6.
102. Brook JS, Ed D. Developmental Trajectories of Cigarette Smoking from Adolescence to the Early Thirties: Personality and Behavioral Risk Factors. 2013;10(8):1283–91.
103. Lenk KM, Erickson DJ, Forster JL. Trajectories of Cigarette Smoking from Teens to Young Adulthood: 2000–2013. 2019;32(5):612–24.
104. Meza R, Jimenez-mendoza E, Levy DT. Trends in Tobacco Use Among Adolescents by Grade , Sex , and Race , 1991-2019. 2022;3(12):1–14.
105. Ahun MN, Lauzon B, Sylvestre M, Eltonsy S, Loughlin JO, Bergeron-caron C. International Journal of Drug Policy A systematic review of cigarette smoking trajectories in adolescents. Int J Drug Policy. 2020;83:102838.
106. Nguena Nguefack HL, Pagé MG, Katz J, Choinière M, Vanasse A, Dorais M, et al. Trajectory modelling techniques useful to epidemiological research: A comparative

narrative review of approaches. Clin Epidemiol. 2020;12:1205–22.
107. Ram N, Grimm KJ. Methods and Measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. Vol. 33, International Journal of Behavioral Development. 2009. p. 565–76.
108. Harris KM, Udry JR, Bearman PS. The Add Health Study: Design and Accomplishments. Chapel Hill Carolina Popul Center, Univ North Carolina Chapel Hill. 2013;1–22.
109. Gordon-Larsen P, Nelson MC, Popkin BM. Longitudinal physical activity and sedentary behavior trends: Adolescence to adulthood. Am J Prev Med. 2004;27(4):277–83.
110. Kwon S, Janz KF, Letuchy EM, Burns TL, Levy SM. Developmental Trajectories of Physical Activity, Sports, and Television Viewing During Childhood to Young Adulthood Iowa Bone Development Study. 2021;169(7):666–72.
111. Oberski DL. Mixture models : latent profile and latent class analysis.
112. Masyn KE. Latent Class Analysis and Finite Mixture Modeling. The Oxford.
113. Barnett TA, Gauvin L, Craig CL, Katzmarzyk PT. Distinct trajectories of leisure time physical activity and predictors of trajectory class membership: A 22 year cohort study. Int J Behav Nutr Phys Act. 2008;5:1–8.
114. Koning M, Hoekstra T, De Jong E, Visscher TLS, Seidell JC, Renders CM. Identifying developmental trajectories of body mass index in childhood using latent class growth (mixture) modelling: associations with dietary, sedentary and physical activity behaviors: a longitudinal study. BMC Public Health. 2016;16(1):1–12.
115. Farooq MA, Parkinson KN, Adamson AJ, Pearce MS, Reilly JK, Hughes AR, et al. Timing of the decline in physical activity in childhood and adolescence: Gateshead Millennium Cohort Study. Br J Sports Med. 2018;52(15):1002–6.
116. Wiium N, Breivik K, Wold B. Growth trajectories of health behaviors from adolescence through young adulthood. Int J Environ Res Public Health. 2015;12(11):13711–29.
117. Lounassalo I, Hirvensalo M, Palomäki S, Salin K, Tolvanen A, Pahkala K, et al. Life-course leisure-time physical activity trajectories in relation to health-related behaviors in adulthood: the Cardiovascular Risk in Young Finns study. BMC Public Health. 2021;21(1):1–13.
118. Saint-Maurice PF, Coughlan D, Kelly SP, Keadle SK, Cook MB, Carlson SA, et al. Association of Leisure-Time Physical Activity across the Adult Life Course with All-Cause and Cause-Specific Mortality. JAMA Netw Open. 2019;2(3):1–12.
119. Oura P, Rissanen I, Junno JA, Harju T, Paananen M. Lifelong smoking trajectories of Northern Finns are characterized by sociodemographic and lifestyle differences in a 46-year follow-up. Sci Rep. 2020;10(1):1–10.
120. The Lancet Public Health. Time to tackle the physical activity gender gap. Lancet Public Heal. 2019;4(8):e360.
121. Devries MC, Jakobi JM. Importance of considering sex and gender in exercise and nutrition research. Appl Physiol Nutr Metab. 2021 Jun;46(6):iii–vii.
122. Barkley JE, Salvy S-J, Sanders GJ, Dey S, Von Carlowitz K-P, Williamson ML. Peer Influence and Physical Activity Behavior in Young Children: An Experimental Study. J Phys Act Heal. 2014 Feb;11(2):404–9.
123. Draper CE, Grobler L, Micklesfield LK, Norris SA. Impact of social norms and social support on diet, physical activity and sedentary behaviour of adolescents: A scoping review. Child Care Health Dev. 2015;41(5):654–67.
124. Rice EL, Klein WMP. Interactions among perceived norms and attitudes about health-related behaviors in U.S. adolescents. Heal Psychol. 2019;38(3):268–75.
125. Perkins KA. Smoking Cessation in Women. CNS Drugs 2001;15(5):391–411.
126. Smith PH, Bessette AJ, Weinberger AH, Sheffer CE, McKee SA. Sex/gender differences in smoking cessation: A review. Prev Med (Baltim). 2016 Nov;92(12):135–40.
127. Trinidad DR, Pérez-Stable EJ, White MM, Emery SL, Messer K. A nationwide analysis of

US racial/ethnic disparities in smoking behaviors, smoking cessation, and cessation-related factors. Am J Public Health. 2011;101(4):699–706.
128. Kandel DB, Kiros G-E, Schaffran C, Hu M-C. Racial/Ethnic Differences in Cigarette Smoking Initiation and Progression to Daily Smoking: A Multilevel Analysis. Am J Public Health. 2004 Jan;94(1):128–35.
129. Flint AJ, Yamada EG, Novotny TE. Black-White Differences in Cigarette Smoking Uptake: Progression from Adolescent Experimentation to Regular Use. Prev Med (Baltim). 1998 May;27(3):358–64.
130. Griesler PC, Kandel DB, Davies M. Ethnic differences in predictors of initiation and persistence of adolescent cigarette smoking in the National Longitudinal Survey of Youth. Nicotine Tob Res. 2002 Feb 1;4(1):79–93.
131. Ivory VC, Blakely T, Richardson K, Thomson G, Carter K. Do changes in neighborhood and household levels of smoking and deprivation result in changes in individual smoking behavior? A large-scale longitudinal study of New Zealand adults. Am J Epidemiol. 2015;182(5):431–40.
132. Mitchell JC. Social Networks. 1974;3(May 2022):279–99.
133. Berkman Lf, Syme Sl. Social Networks, Host Resistance, And Mortality: A Nine-Year Follow-Up Study Of Alameda County Residents. Am J Epidemiol. 1979 Feb;109(2):186–204.
134. Seo D-C, Huang Y. Systematic Review of Social Network Analysis in Adolescent Cigarette Smoking Behavior*. J Sch Health. 2012 Jan;82(1):21–7.
135. Voorhees CC, Murray D, Welk G, Birnbaum A, Ribisl KM, Johnson CC, et al. The role of peer social network factors and physical activity in adolescent girls. Am J Health Behav. 2005;29(2):183–90.
136. Christakis NA, Fowler JH. The Collective Dynamics of Smoking in a Large Social Network. N Engl J Med. 2008 May 22;358(21):2249–58.
137. Aysola J, Rewley J, Xu C, Schapira M, Hubbard RA. Primary Care Patient Social Networks and Tobacco Use: An Observational Study. J Prim Care Community Heal. 2022;13.
138. Prochnow T, Patterson MS. Assessing Social Network Influences on Adult Physical Activity Using Social Network Analysis: A Systematic Review. Am J Health Promot. 2022;36(3):537–58.
139. NELSON EE, LEIBENLUFT E, McCLURE EB, PINE DS. The social re-orientation of adolescence: a neuroscience perspective on the process and its relation to psychopathology. Psychol Med. 2005 Feb 21;35(2):163–74.
140. Page RM, Ihasz F, Simonek J, Klarova R, Hantiu I. Cigarette Smoking, Friendship Factors, and Social Norm Perceptions among Central and Eastern European High School Students. J Drug Educ. 2006 Sep 21;36(3):213–31.
141. Ali MM, Dwyer DS. Estimating Peer Effects in Adolescent Smoking Behavior: A Longitudinal Analysis. J Adolesc Heal. 2009 Oct;45(4):402–8.
142. Roberts ME, Nargiso JE, Gaitonde LB, Stanton CA, Colby SM. Adolescent social networks: General and smoking-specific characteristics associated with smoking. J Stud Alcohol Drugs. 2015;76(2):247–55.
143. Yun EH, Kang YH, Lim MK, Oh JK, Son JM. The role of social support and social networks in smoking behavior among middle and older aged people in rural areas of South Korea: A cross-sectional study. BMC Public Health. 2010;10.
144. Liu J, Zhao S, Chen X, Falk E, Albarracín D. The influence of peer behavior as a function of social and cultural closeness: A meta-analysis of normative influence on adolescent smoking initiation and continuation. Psychol Bull. 2017 Oct;143(10):1082–115.
145. Dyal SR, Valente TW. A Systematic Review of Loneliness and Smoking: Small Effects, Big Implications. Subst Use Misuse. 2015 Nov 10;50(13):1697–716.

146. Dyal SR, Valente TW. A Systematic Review of Loneliness and Smoking: Small Effects, Big Implications. Subst Use Misuse. 2015 Nov 10;50(13):1697–716.
147. Littlecott HJ, Moore GF, McCann M, Melendez-Torres GJ, Mercken L, Reed H, et al. Exploring the association between school-based peer networks and smoking according to socioeconomic status and tobacco control context: a systematic review. BMC Public Health. 2022;22(1):1–22.
148. Valente TW, Unger JB, Johnson CA. Do popular students smoke? The association between popularity and smoking among middle school students. J Adolesc Heal. 2005;37(4):323–9.
149. Valente TW, Gallaher P, Mouttapa M. Using social networks to understand and prevent substance use: A transdisciplinary perspective. Subst Use Misuse. 2004;39(10–12):1685–712.
150. Golaszewski NM, LaCroix AZ, Hooker SP, Bartholomew JB. Group exercise membership is associated with forms of social support, exercise identity, and amount of physical activity. Int J Sport Exerc Psychol. 2022;20(2):630–43.
151. Liu S, Hachen D, Lizardo O, Poellabauer C, Striegel A, Milenković T. Network analysis of the NetHealth data: exploring co-evolution of individuals' social network positions and physical activities. Appl Netw Sci. 2018;3(1):45.
152. Haslam SA, Jetten J, Postmes T, Haslam C. Social identity, health and well-being: An emerging agenda for applied psychology. Appl Psychol. 2009;58(1):1–23.
153. Karelina K, Devries AC. Modeling social influences on human health. Psychosom Med. 2011;73(1):67–74.
154. Graupensperger S, Panza M, Evans MB. Network centrality, group density, and strength of social identification in college club sport teams. Gr Dyn Theory, Res Pract. 2020 Jun;24(2):59–73.
155. Hertel AW, Mermelstein RJ. Smoker identity and smoking escalation among adolescents. Heal Psychol. 2012;31(4):467–75.
156. Poole R, Carver H, Anagnostou D, Edwards A, Moore G, Smith P, et al. Tobacco use, smoking identities and pathways into and out of smoking among young adults: a meta-ethnography. Subst Abus Treat Prev Policy. 2022;17(1):1–19.
157. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. Vol. 145, Circulation. 2022. 153–639 p.
158. Gitlin LN, Czaja SJ. Promises and Challenges of Behavioral Intervention Research. Behav Interv Res Des Eval Implement. 2018;3–18.
159. López-Pintado D. Diffusion in complex social networks. Games Econ Behav. 2008;62(2):573–90.
160. Yi Y, Zhang Z, Gan C. The effect of social tie on information diffusion in complex networks. Phys A Stat Mech its Appl. 2018;509:783–94.
161. Arnaboldi V, Conti M, Passarella A, Dunbar RIM. Online Social Networks and information diffusion: The role of ego networks. Online Soc Networks Media. 2017;1:44–55.
162. Centola D. Supporting Online Material for "The spread of behavior in an online social network experiment." Science. 2010;329(5996):1194–7.
163. Kim J, Rasouli S, Timmermans HJP. Social networks, social influence and activity-travel behaviour: a review of models and empirical evidence. Transp Rev. 2018 Jul 4;38(4):499–523.
164. Jackson MO. Networks and the Identification of Economic Behaviors. SSRN Electron J. 2014;28(4):3–22.
165. Leskovec J. Diffusion and Cascading Behavior in Networks. Min Massive Data Sets Secur - Adv Data Mining, Search, Soc Networks Text Mining, their Appl to Secur Proc {NATO} Adv Study Inst Min Massive Data Sets Secur Gazzada (Varese), It.2007;19:169–

85.
166. Dyal SR. Network Influences on Behavior: A Summary of Tom Valente's Keynote Address at Sunbelt XXXV: The Annual Meeting of the International Network for Social Network Analysis. Connections. 2016;35(2):52–7.
167. Blume L, Easley D, Kleinberg J, Kleinberg R, Tardos É. Which Networks are Least Susceptible to Cascading Failures? In: 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science. 2011. p. 393–402.
168. Olof Hedman N, Johansson R, Rosenqvist U. Clustering and inertia: structural integration of home care in Swedish elderly care. Int J Integr Care. 2007 Sep 12;7:e32–e32.
169. DeGroot MH. Degroot Consensus.Pdf. Vol. 69, Journal of the American Statistical Associtation. 1974. p. 118–21.
170. Fudenberg D. Rules of Thumb for Social Learning Author (s): Glenn Ellison and Drew Fudenberg Published by : The University of Chicago Press
171. Telesford QK, Joyce KE, Hayasaka S, Burdette JH, Laurienti PJ. The Ubiquity of Small-World Networks. Brain Connect. 2011;1(5):367–75.
172. Pereira G, Prada R, Santos PA. Integrating social power into the decision-making of cognitive agents. Artif Intell. 2016;241:1–44.
173. Valente TW. Social network thresholds in the diffusion of innovations. Soc Networks. 1996;18(1):69–89.
174. Centola D, Macy M. Long Ties 1. Am J Sociol. 2007;113(3):702–34.
175. Id RFH, De K, Id H, Id JMM, Id JB, Id TWV, et al. Social network interventions for health behaviours and outcomes : A systematic review and meta-analysis. 2019;46:1–25.
176. Interventions C. I . State and Community Interventions.
177. Institute NC. Smoking and Tobacco Control Monograph No. 6. 11(1):vii–viii.
178. Roeseler A, Burns D. The quarter that changed the world. Tob Control. 2010;19(SUPPL. 1):3–15.
179. Farrelly MC, Loomis BR, Han B, Gfroerer J, Kuiper N, Couzens GL, et al. A comprehensive examination of the influence of state tobacco control programs and policies on youth smoking. Am J Public Health. 2013;103(3):549–55.
180. Kelly JA, Murphy DA, Sikkema KJ, McAuliffe TL, Roffman RA, Solomon LJ, et al. Randomised, controlled, community-level HIV-prevention intervention for sexual-risk behaviour among homosexual men in US cities. Lancet. 1997 Nov;350(9090):1500–5.
181. Holliday J, Audrey S, Campbell R, Moore L. Identifying Well-Connected Opinion Leaders for Informal Health Promotion: The Example of the ASSIST Smoking Prevention Program. Health Commun. 2016;31(8):946–53.
182. Ye M, Zino L, Mlakar Ž, Bolderdijk JW, Risselada H, Fennis BM, et al. Collective patterns of social diffusion are shaped by individual inertia and trend-seeking. Nat Commun. 2021;12(1).
183. Hunter RF, Montes F, Murray JM, Sanchez-Franco SC, Montgomery SC, Jaramillo J, et al. MECHANISMS Study: Using Game Theory to Assess the Effects of Social Norms and Social Networks on Adolescent Smoking in Schools—Study Protocol. Front Public Heal. 2020;8(August):1–14.
184. Barnes W, Gartland M, Stack M. Old habits die hard: Path dependency and behavioral lock-in. J Econ Issues. 2004;38(2):371–7.
185. Mass-Reach Health Communication Interventions. Best Pract Compr Tob Control Progr. :30–9.
186. Tuzzio L, Verboncoeur C, Stewart AL, Ph D, Fineman N, Ph D. Community Healthy Activities Model Program for Seniors COMMUNITY HEALTHY ACTIVITIES MODEL PROGRA M Community Healthy Activities Model Program for Seniors. 2018;2003(Champs Ii).
187. Halpern SD, French B, Small DS, Saulsgiver K, Harhay MO, Audrain-McGovern J, et al.

Randomized Trial of Four Financial-Incentive Programs for Smoking Cessation. N Engl J Med. 2015;372(22):2108–17.
188. Prost A. The role of epidemiology in the process of decision-making. Sante. 7(1):61–4.
189. Ahern J, Jones MR, Bakshis E, Galea S. Revisiting rose: Comparing the benefits and costs of population-wide and targeted interventions. Milbank Q. 2008;86(4):581–600.
190. Health EC, Rychetnik L, Building VC. interventions. 2006;119–27.
191. Schneeweiss S, Seeger JD, Jackson JW, Smith SR. Methods for comparative effectiveness research/patient-centered outcomes research: From efficacy to effectiveness. J Clin Epidemiol. 2013;66(8 SUPPL.8):S1.
192. Evans As. Re: “Definitions Of Epidemiology.” Am J Epidemiol. 1979 Mar;109(3):379–82.
193. Tracy M, Cerdá M, Keyes KM. Agent-Based Modeling in Public Health: Current Applications and Future Directions. Annu Rev Public Health. 2018 Apr 1;39(1):77–94.