Title

Causal and Agent-Based Modeling of Obesity and its Life-Course Risk Factors and Outcomes in Children and Adults

Author

Nianogo, Roch Arnaud Kibsa

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Causal and Agent-Based Modeling of Obesity and its Life-Course Risk Factors and Outcomes in Children and Adults

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Epidemiology

by

Roch Arnaud Kibsa Nianogo

2017

ABSTRACT OF THE DISSERTATION

Causal and Agent-Based Modeling

of Obesity and its Life-Course Risk Factors and Outcomes

in Children and Adults

by

Roch Arnaud Kibsa Nianogo

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2017

Professor Onyebuchi Aniweta Arah, Chair

For decades, obesity has been a major public health problem in the US and has been one of the most predominant players in the increase of type 2 diabetes (T2DM) incidence. Obesity is thought to be the result of the interplay between individual and environmental factors which can occur early and throughout an individual's life course. Despite major ongoing prevention efforts, obesity is still on the rise and this has warranted its description as a complex health problem

calling for the use of systems science methods to disentangle such complexity. The overarching goal of this dissertation was to apply systems science and causal analytical approaches to study the life-course development of obesity and its effects on T2DM. Specifically, we developed an agent-based model of a cohort of children born in Los Angeles county—ViLA (i.e. Virtual Los Angeles Cohort) and followed from birth into adulthood in order (i) to forecast the incidence and trends of obesity and T2DM, (ii) to investigate the mechanisms through which childhood obesity affects T2DM and (iii) to evaluate the effectiveness of key health interventions on obesity and T2DM in ViLA. We used simulated data from 98,230 individuals in ViLA and observational data from 1054 children enrolled in the Special Supplemental Nutrition Program for Women, Infants and Children and applied the g-computation algorithm to estimate causal quantities. Our results suggest that the incidence and prevalence of obesity and T2DM are generally high with notable racial disparities and will continue rising over time and with age at an alarming rate. Furthermore, much of the effect attributable to childhood obesity in the development of incident T2DM was due to pathways other than through adult obesity. Additionally, engaging in moderate-to-vigorous physical activity and eliminating fast-food consumption were the most effective interventions for preventing obesity and T2DM. For maximum effectiveness, interventions have to be implemented in combination with one another and virtually at every critical life stages throughout the life span. Agent-based simulation models could be used as virtual laboratories for integrating best existing evidence, gaining new insights, exploring new mechanisms and evaluating intervention effectiveness in obesity and diabetes research.

The dissertation of Roch Arnaud Kibsa Nianogo is approved.

Beate Ritz R.

ZuoFeng Zhang

May-Choo Wang

Tony Yu-Huong Kuo

Onyebuchi Aniweta Arah, Committee Chair

University of California, Los Angeles

2017

DEDICATION

To the God the Father, Jesus Christ the Son and the Holy Spirit,

To the Blessed Virgin Mary, Saint Joseph, Saint Michael and my Guardian Angel,

To my wonderful parents Aimé and Georgette Nianogo,

To my supportive brothers William and Thierry Nianogo,

And to my better half and beloved wife Annie Nianogo

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF TABLES

Chapter 5

Chapter 6

LIST OF ACRONYMS

ACS	American community survey
ALC	Alcohol consumption
BMI	Body mass index
CDC	Center for disease control and prevention
CDE	Controlled direct effect
CDEsto	Stochastic controlled direct effect
CHIS	California health interview survey
CI	Confidence interval
DAG	Directed acyclic graph
DE	Direct effect
EBF	Exclusive breast-feeding
EnvFFD	Environmental or neighborhood fast-food density
EnvPRK	Environmental or neighborhood park Access
EnvSMD	Environmental or neighborhood supermarket density
EnvWLK	Environmental or neighborhood walkability
FFD	Fast-food consumption
FFV	Fresh fruit and vegetable consumption
FPL	Federal poverty level
HR	Hazard Ratio
MD	Mean difference
MSM	Marginal structural model
MVPA	Moderate-to-vigorous physical activity
NDE	Natural direct effect
NHANES	National health and nutrition examination survey
NIE	Natural indirect effect
OBE	Obese
OR	Odds ratio
P0	Prevalence of outcome when the predictors are set to 0
PDE	Pure direct effect
PE	Portion eliminated
Pest	Point estimate
PIE	Pure direct effect
RERI	Relative excess risk for interaction
RIE	Reference interaction effect
RR	Risk Ratio
SD	Standard deviation
SE	Standard error
SES	Socio-economic status
SMK	Smoking

SSB	Sugar-sweetened beverage consumption
SUTVA	Stable unit treatment value assumption
T2DM	Type 2 diabetes
TDE	Total direct effect
TE	Total effect
TIE	Total indirect effect
WHO	World Health Organization
WIC	Special supplemental nutrition program for women, infants and children

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Dr. Onyebuchi Arah who has been and is for me a mentor, an advisor, my dissertation committee chair and a friend. I am grateful to Dr. Arah for having chosen me to undertake such wonderful and inspiring work and for giving me the great opportunity to work with him. I am thankful to Dr. Arah for believing in me, for always pushing me to surpass my limits and for the constant guidance and support. I am also very grateful to Dr. Arah for always paying attention to my needs even before I express them and for looking after me. Working closing with Dr. Arah has truly been a blessing in my journey and I am thankful to him for opening new career opportunities through such an exclusive and highly technical training. He has been an inspiration and a role model for me throughout this journey.

I am also very thankful to Dr. Beate Ritz for all her support during this journey and for encouraging and helping me to pursue my career goals. I am also thankful to her for her constant mentorship and for giving me the opportunity to meet new collaborators. I would like to express my gratitude to Dr. Zuo-Feng Zhang for encouraging me during my master's and for his kind support during the doctoral program. I am also very thankful to Dr. May Wang for giving me the opportunity to work with her and for proving timely support during my training. I am deeply thankful to her for her mentorship, invaluable advice and encouragement and also for giving me the opportunity to learn new methods and to work with new collaborators. Lastly, I wanted thank Dr. Tony Kuo for being a wonderful collaborator and mentor and for all the kind support that he has provided during the doctoral training.

I want to express my deepest gratitude to great faculty and staff within the school of public health without whom this journey to UCLA would not have been possible. I am particularly greatly indebted to Dr. Anne Rimoin, Kirstin Chickering, Dr. Michael Prelip, Dr. Paula Tavrow, Dr. Anne Pebley for their compassion and help in making my dream come true. Likewise, I wanted thank Josie Wei, Sagoo Skhuwinder and David Clarck for all their advice and assistance during my years of training at UCLA. I also want to thank Dr. Shehnaz Hussain for her support and for providing assistance during my transition to the doctoral program in Epidemiology. I want to thank the UCLA Department of Epidemiology staff and faculty and especially Joy Miller, Dr. Roberta Malmgren, Dr. Lisa Smith, Dr. Roger Detels and Dr. Karin Michels, for their constant support and advice before and during my training. I also wanted thank the late Dr. Toni Yancy who provided great support during my master's and who was a great mentor, a role model and an inspiration for physicians engaged in preventive medicine and health promotion. I would also like to thank my uncle and Godfather Dr. Alain Damiba who inspired me and greatly contributed to my choosing of public health and coming to the United States to further my education. A particular thanks to my previous mentors and current collaborators from Burkina Faso who have believed in me and encouraged me throughout this journey: Dr. Nicolas Meda, Dr. Laurent Ouedraogo, Dr. Seni Kounda, Dr. Alain Bougouma, Dr. Bibiane Koné, Dr. Maimouna Outtara, Dr. Arsene Ouedraogo, Dr. Joseph Drabo, Dr. Arnaud Dienderé and Dr. Nayi Zongo.

I greatly appreciate and acknowledge the Early Childhood Obesity Systems Science Study (ECOSys) and PHFE-WIC and Shannon Whaley for providing resources including but not limited to access to the NETS and WIC data and the technical training and facilities that helped

complete this project. I would also like to acknowledge and thank the UCLA Lewis Center and Norman Wong for providing the WalkScore data and the Division of Chronic Disease and Injury Prevention and the Office of Health Assessment and Epidemiology in the Los Angeles County Department of Public Health for providing us with the LAHANES data and for their inputs in this project. I would like to acknowledge the UCLA Institute for Digital Research and Education and Michael Tzen for providing statistical training and support. I would also like to thank Dr. Nathaniel Osgood who taught me how to build agent-based models in AnyLogic. Lastly, I would like to thank Maral DerSarkissian, Aolin Wang, Solomon Makgoeng, Amy Chai, Mekdes Gebremariam, Tabashir Nobari and Darren Ho for their various help during this project.

I would like to thank my friends, family and church community both in Burkina Faso, in Los Angeles and around the U.S. for providing help and emotional support during all these years. I would especially like to thank Eugene, Yves, the McFadyen Family, the Traoré family, Michael, Denise, William, Heather, Claire, Nelly, Eden, Tess, Charlie, Sunday and Toyin. I am grateful that God has put them in my life. Lastly, I would like to express my heartfelt and deepest gratitude to my parents Aimé and Georgette for instilling in me my faith in God, the love and compassion towards others, the integrity and diligence in work and for making this training possible, to my brothers William and Thierry for their constant support and to my beloved companion, my better half and best friend and wife Annie for her faith in me, her patience, compassion and unwavering support in this journey.

This dissertation was supported by the UCLA Department of Epidemiology, the UCLA Fielding Epidemiology Fellowship, the Burroughs Wellcome Fund Inter-school Training in Chronic

Diseases (BWF-CHIP) and the Dissertation Year Fellowship (DYF). I would also like to acknowledge that I benefited from facilities and resources provided by the California Center for Population Research at UCLA (CCPR), which receives core support (R24-HD041022) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).

Chapter 3 is a version of "*Nianogo, R. A. and Arah, O. A. (2017) An Agent-Based Model of Obesity and Diabetes in the Virtual Los Angeles Cohort: The ViLA-Obesity Simulation Suite*" which is in preparation for publication.

Chapter 4 is a version of "*Nianogo, R. A. and Arah, O. A. (2017) Modeling the Role of Childhood Adiposity in the Development of Adult Type 2 Diabetes in a 64-Year Follow-Up Study in Los Angeles: An Agent-Based Simulation Study*" which is in preparation for publication.

Chapter 5 is a version of "*Nianogo, R. A. and Arah, O. A. (2017) Evaluating the Effectiveness of Key Health Interventions on Obesity and Diabetes throughout the Life Course in the Virtual Los Angeles Cohort: An Agent-Based Simulation Study*" which is in preparation for publication.

Chapter 6 is incorporated from Nianogo, R. A., Wang, M. C., Wang, A., Nobari, T. Z., Crespi, C. M., Whaley, S. E., and Arah, O. A. (2016) Projecting the impact of hypothetical early life interventions on adiposity in children living in low-income households. *Pediatric Obesity*, doi: 10.1111/ijpo.12157. © John Wiley and Sons with permission from John Wiley and Sons

Permission to reuse and reproduce **Figure 2.3** and **Figure 2.4** was obtained from © Springer and © Elsevier prior to publication of this dissertation.

VITA

EDUCATION

2013	Master's of Public Health, *Public Health* University of California, Los Angeles, CA, USA
2010	Medical Doctorate, *Medicine* Université de Ouagadougou, Ouagadougou, Burkina Faso (BF)

PROFESSIONAL AND ACADEMIC EXPERIENCE

Research Experience

2013-2017	*Research Assistant* -Department of Epidemiology -Early Childhood Obesity Systems Science Study (ECOSyS) University of California, Los Angeles, CA, USA
2011-2013	*Research Assistant* Working Out Regularly Keeps Individuals Nurtured and Going study (WORKING) UCLA Jonson Comprehensive Cancer Center, Division of Cancer Prevention and Control Research University of California, Los Angeles, CA, USA

Clinical Experience

2009-2010	*Intern* HIV-infected Care unit, Department of Internal Medicine Centre Hospitalier Universitaire Yalgado Ouedraogo Ouagadougou, BF

Teaching Experience

2015-2016	*Teaching Assistant* Courses: Statistical Methods in Epidemiology Statistical Modeling in Epidemiology UCLA, Dept. of Epidemiology, Los Angeles, CA
2005- 2009	*Teaching Assistant* Course: Osteology, Anatomy Université de Ouagadougou, Ouagadougou, BF

PEER-REVIEWED PUBLICATIONS

Wang A, **Nianogo RA**, Arah OA. G-Computation of Average Treatment Effects on the Treated and the Untreated. *BMC Med Res Methodol*. 2016; 17(3). DOI: 10.1186/s12874-016-0282-4

Nianogo RA, Wang MC, Wang A, Nobari TZ, Crespi CM, Whaley SE, Arah OA. Projecting the impact of hypothetical early life interventions on childhood adiposity in children living in low-income households. *Pediatric Obesity*. 2016. DOI: 10.1111/ijpo.12157

Nianogo RA, Kuo T, Smith L V, Arah OA. Associations between self-perception of weight, food choice intentions, and consumer response to calorie information: a retrospective investigation of public health center clients in Los Angeles County before the implementation of menu-labeling regulation. *BMC Public Health*. 2016;16(1):60. doi:10.1186/s12889-016-2714-9.

Young SD, **Nianogo RA**, Chiu CJ, Menacho L, Galea J. Substance use and sexual risk behaviors among Peruvian MSM social media users. *AIDS Care*. 2016;28(1):112–8. doi:10.1080/09540121.2015.1069789.

Young SD, Cumberland WG, **Nianogo RA**, Menacho LA, Galea JT, Coates T. The Hope Social Media Intervention For Global HIV Prevention In Peru: A Cluster Randomized Controlled Trial. *Lancet HIV*. 2015;2(1):e27–e32. Doi:10.1016/s2352-3018(14)00006-x.

Nianogo RA, Arah OA. Agent-Based Modeling of Non-Communicable Diseases: A Systematic Review. *Am J Public Health*. 2015;105(3):e20–31. Doi:10.2105/ajph.2014.302426.

ORAL* AND POSTER PRESENTATIONS

***Nianogo RA**, Wang MC, Wang A, Nobari T, Crespi K, Whaley SE, Arah OA. Using the parametric g-formula to estimate the potential effects of intervening on risk factors of increased adiposity in a longitudinal cohort of preschool-aged children in Los Angeles
Society of Epidemiologic Research (SER) annual meeting, Denver, Colorado, June 15-19th, 2015

***Nianogo RA**, Wang MC, Wang A, Nobari T, Crespi K, Whaley SE, Arah OA. Early life exposure to behavioral risk factors of obesity and subsequent weight of preschool-aged children in Los Angeles
American Public Health Association (APHA), October 31st-November 4th, 2015 Chicago, Illinois

Nianogo RA, Arah OA. Using current evidence and the parametric g-formula to forecast the mean body mass index and diabetes prevalence in 48 low- and middle-income countries
Epidemiology Congress of the Americas, Miami, Florida June 21st- June 24th, 2016

Chapter 1. General Introduction

1.1. Background

Obesity is a global health problem affecting worldwide about 1.5 billion adults in 2008 and over 40 million children in 2011.[1] In the United States, obesity has reached epidemic levels with two in three adults and one in three children and adolescents considered obese or overweight.[2] This condition disproportionately affects lower-income minority and disadvantaged groups giving rise to health disparities.[2] In 2012, the prevalence of overweight or obesity among adults older than 20 years was higher in African-Americans and Hispanics compared to other racial groups. Overall, males were more likely to be obese than their female counterparts with some variations by race.[3] The younger segment of the society is not spared and has in fact seen its obesity prevalence drastically increase. Over the past three decades, childhood obesity (often defined as having a body mass index (BMI) $\geq$ 95th percentile of age-and sex-specific reference values) has more than doubled in children and quadrupled in adolescents, rising from 7% to about 18% among children aged 6-11 years and from 5% to about 21% among adolescents aged 12-19 years in 2012.[3] Even more worrisome is the economic toll of childhood obesity estimated at $14.1 billion annually in direct health expenses, costs that have dramatically increased in the past decades and expected to add to the burden of overall obesity-related health care costs since today's obese children are prone to become tomorrow's obese adults.[4–7] Childhood obesity is, therefore, a major public health problem that requires immediate action in order to halt and reverse the current trends.

Fundamentally, obesity is thought to be due to a greater energy imbalance between energy intake and energy expenditure than is expected for normal growth and development, which results in the accumulation of energy stores in the body and formation of excessive adipose tissue.[1,8,9] The increase in obesity rates has been seen as a result of an increased energy intake and/or an increase in physical inactivity.[1] These changes have their roots in individual and environmental factors.[10] Individual or personal factors include genetic, biologic and physiologic factors (e.g. genetic predisposition, neuronal regulation, appetite and satiety mechanism, metabolism rate, adipose tissue metabolism, lipoprotein and glucose regulation, intrauterine and early life programming) as well as cognitive and psychological factors (e.g. knowledge, beliefs, attitude, taste preference). Environmental factors, on the other hand, consist of physical (e.g. food environment, built-environment, home, recreational facilities, schools, and community) and social factors (e.g. social networks, cultural norms, advertising and food marketing).[11,12] All these factors may influence and are influenced by individual dietary behaviors (e.g. food portion sizes, consumption of sugar-sweetened beverages [SSB], consumption of fruits and vegetables, consumption of fast-foods) and physical activity related behaviors (e.g. walking, TV watching and sedentary lifestyle).[12,13] Lastly, intergenerational inheritance of obesity through epigenetic factors (i.e. heritable influence on genes that occurs without a change in the DNA sequence) can also play a critical role in obesity development. For instance, children whose mothers are diabetic or develop gestational diabetes or who smoke during pregnancy are more likely to become obese.[14,15] In addition, during infancy, a shortened period of breastfeeding as well as a diminished amount of sleep can increase a child's risk of being obese.[15] Environmental chemicals such as endocrine disrupting chemicals (EDCs such as Bisphenol-A) have been observed to be associated with the development of obesity in animal

studies and may play an important etiologic role in human obesity. A review by Newbold et al suggests that a brief exposure early in development to estrogenic EDCs during critical periods can increase body weight gain with age by altering weight homeostasis. Another example is infancy exposure to chemicals found in soy-based formula (e.g. genistein).[16]

Obesity has been linked to a number of chronic conditions including but not limited to cardiovascular (e.g. atherosclerosis, hypertension, stroke, coronary heart diseases, left ventricular hypertrophy), metabolic (e.g. metabolic syndrome, diabetes mellitus, dyslipidemias), cancers (e.g. postmenopausal breast, colon, kidney and endometrial cancers), pulmonary (e.g. obstructive sleep apnea, asthma) hepatic (e.g. nonalcoholic fatty liver disease, nonalcoholic steatohepatitis) orthopedic (e.g. osteoarthritis, Blount's disease) and psychological conditions (e.g. poor quality of life, reduced self-esteem, depression) as well as mortality.[15,17] Through these conditions, obesity may affect disability risk medically (e.g. reduced mobility, breathing problems) and socially (e.g. reduced employment) which in turn may further increase adiposity level.[18]

1.2. Obesity-related diabetes mellitus and pathophysiology

Most notably, overweight and obesity are strongly associated with increased risk of type 2 diabetes. . Almost 80 to 90% of type 2 diabetes patients are overweight or obese.[19–21] The recent increase in diabetes prevalence has been concomitant to the rise in overweight and obesity prevalence.[15,21] In 2012, approximately 29 million Americans (i.e. 9.3% of the population) had diabetes; about 1.7 million new cases are diagnosed each year.[22] Diabetes mellitus (DM) is a metabolic disorder characterized by prolonged hyperglycemia (i.e. high levels of blood glucose, clinically seen as fasting plasma glucose $\geq$ 7.0 mmol/L or a glucose tolerance test $\geq$ 11.1 mmol/l or percent of $HbA_{1c} \geq 6.5$).[21] Two types are of interest in this research: type 1 diabetes mellitus (T1DM), previously known as insulin-dependent diabetes or juvenile onset diabetes due to a

complete insulin deficiency and type 2 diabetes (T2DM), formerly known as non-insulin-dependent diabetes mellitus or adult-onset diabetes due to insulin resistance and relative insulin deficiency. [21] Although, some systematic reviews have reported on a likely association between obesity and type 1 diabetes,[23] most of the increased risk due to obesity has been seen with type 2 diabetes.[21] Obesity is thought to cause type 2 diabetes through inflammatory, hormonal, and metabolic pathways via the production of biochemical factors such as hormones (e.g. leptin, adiponectin), adipokines (e.g. tumor necrosis factor α [TNF-α], interleukin-6 [IL-6]) and free fatty acids (FFA) (See **Figure 1.1**). Adiponectin, in particular, one of the most predominant hormones secreted by fat cells increases insulin sensitivity and decreases inflammation and atherosclerosis [24] However, during obesity, changes in the adipocyte function occur resulting in a decreased adiponectin level. This reduced adiponectin level is an independent risk factor for the metabolic syndrome and diabetes mellitus through pathways involving insulin resistance and inflammation. [24,25] In addition, the level of FFA from adipocyte lipolysis is increased during obesity and is responsible for inducing insulin resistance and type 2 diabetes via accumulation of triglyceride and release of pro-inflammatory adipokines (i.e. interleukin-6, TNF) and adiponectin. [21]

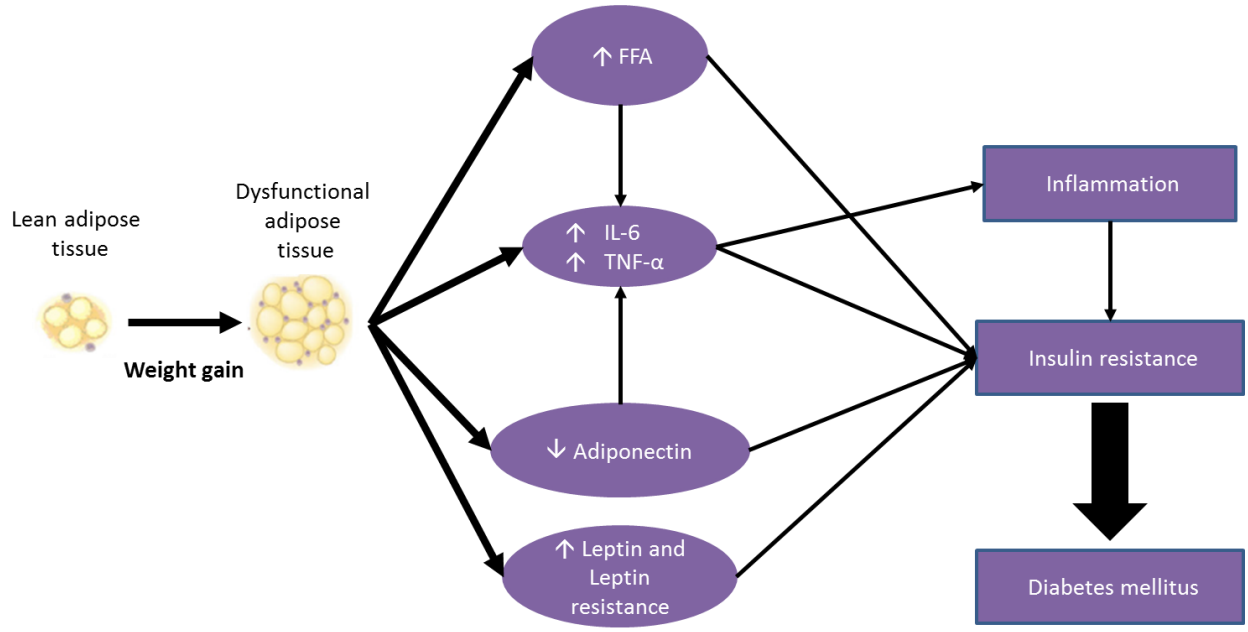


Figure 1.1 Pathophysiology of obesity-related diabetes adapted from [21,26,27]. TNF-α : tumor necrosis factor α; IL-6: interleukin-6; FFA: free fatty acids

1.3. Life course development of obesity and long term health effects

During the life course, individuals who are obese or overweight in their childhood or adolescence may experience obesity-related long-term health effects later in their lives. Besides the short term effects (e.g. high cholesterol levels, high blood pressure, high triglyceride levels, high insulin level and impaired blood glucose, fatty liver disease, high rates of alcohol and tobacco consumption, premature sexual behaviors, unhealthy dietary habits, sedentary lifestyles and poor school performance) associated with obesity in childhood ,[28] a number of studies have reported long term health effects. [7,15] Children and adolescents who are obese are at higher risk of becoming obese as adults, thus putting them further at higher risk for cardio-metabolic diseases as well as some types of cancers.[28,29] A study found that up to four out of five overweight children or adolescents who were overweight become obese as adults.[30] In addition, obese adults who were also obese in childhood had higher risk of hypertension, metabolic

syndrome, coronary heart disease and dyslipidemia.[31] Other conditions associated with being obese as a child include sleep apnea, asthma, hepatic steatosis, musculoskeletal disorders, gall bladder diseases and potentially a shorter life span.[9] These findings suggest that effective prevention during childhood and adolescence could substantially reduce subsequent risk of cardio-metabolic diseases and related morbidity and mortality in adulthood.

1.4. Overview of the conceptual frameworks of obesity etiology and prevention

The complex and multi-factorial nature of the obesity epidemic suggests that tackling one risk factor at a time may not be adequate for in curbing the epidemic. This has led public health researchers and practitioners to seek behavioral health theories to help explain and understand mediating variables and relationships that give rise to obesity.[12] These theories and frameworks have also provided the basis for designing and evaluating health interventions.[32] Common ones include the Trans Theoretical Model (TTM) (also called Stages of Change (SOC)), the Social Cognitive Theory (SCT) and the Socio-Ecological Model (SEM).[12,33] Two of the most frequently used are the SCT and the SEM.[33,34] Briefly, the SCT is a model that emphasizes the dynamic ongoing process in which personal factors (e.g. cognitive and biological), environmental (e.g. physical and social) and human behavior exert influence upon each other. [35] This framework theorizes that behavior changes as a result of an ongoing interaction between personal factors and environmental factors which also change as a result of new behavior adoption.[35] The SEM on the other hand, is a multilevel and interactive approach that emphasizes the interaction between factors within and across all levels of health problems (individual, interpersonal, institutional, community, public policy).[35] The basic premise is that individual behaviors both shape and are shaped by the physical and social environment. [35]

Nevertheless, a systematic review of obesity-related interventions reported that few interventions have indeed incorporated multilevel approaches.[36] What is more, available theoretical models have only been able to explain less than 50% of the variability in behavior change.[37] In light of these shortcomings, some authors have suggested that the effects of physical activity and dietary behaviors on obesity risk may be non-linear and sensitive to initial conditions and may involve multiple dynamic interactions between individuals and their environment.[37,38] This issue underlines the need to use novel and innovative approaches such as systems science methods to incorporate the dynamic complex and non-linear nature of human behavior in efforts to the obesity epidemic.[12,13,39]

1.5. Racial/ethnic health disparities in obesity

Racial/ethnic health disparities in obesity abound. While the trends and prevalence of obesity are high in all racial/ethnic groups, the prevalence is constantly higher among African-Americans and Hispanic sub-populations.[2,40] Although the exact mechanism for these disparities are not clear, a study based on the California Behavioral Risk Factor Survey showed that after adjusting for socio-economic status, gender, age and co-morbidities the relative risk of obesity was attenuated but remained non-null among minorities such as African-Americans and Hispanics whereas being Asian was associated with a lower risk of being obese.[40] The concept of health disparities is one that is not new but one that steers discussion.[41] The National Institute of Health defines disparities as differences in the incidence, prevalence, mortality, and burden of diseases and related adverse health conditions that exist among specific population groups in the United States[41,42] Nevertheless, not all differences are considered a health disparity but rather differences that systematically and negatively impact less advantaged

groups.[41,43]These differences are deemed avoidable, unnecessary, unjust and remediable by social or policy interventions.[41,43] There is no real consensus on what is responsible for such health disparities but they are believed to occur in groups that have persistently experienced social and/or economic disadvantage or discrimination (e.g. low socio-economic position, female sex, sexual minorities).[41,44]In this dissertation, we will investigate these health disparities in terms of heterogeneity in the target causes and consequences of obesity due to race/ethnicity and present the findings stratified by racial/ethnic subpopulations.

1.6. Practice and Research gaps and how this dissertation is going to close the gap

1.6.1. Gap in framing childhood obesity as a complex health issue

Many authors now think that the rising rates of obesity—in particular in childhood—is not just the result of individual behaviors but rather a corollary of complex dynamic interactions and influences between the relevant and heterogeneous factors within and across all levels making it a complex health issue to address.[13,39] Such conclusion stems from the realization that focusing only on individual behavioral factors has shown limited success in preventing further obesity.[45] Some researchers have posited that to adequately address complex problems, researchers need to use a systems-oriented approaches (e.g. agent-based modeling).[13] Until recently, these methods as well as computer power to model such complex systems were not available.[46] Therefore, in the absence of systems science technologies, the tendency has been to simplify complex problems to simpler more manageable problem that could be handled with less powerful computers.[46]

This research intends to study childhood obesity as a complex health issue that emerges from the dynamic interactions between individuals and their environment, by using a promising

sophisticated systems science method: agent-based modeling (ABM). An increasing number of public health researchers have indeed started to embrace systems science methods to untangle questions pertaining to childhood obesity.[47,48]

1.6.2. Gap in assessing long-term effects of obesity

As mentioned before, obesity is one of the most potent risk factor for diabetes.[20,21] Globally, 44% of the diabetes burden is attributable to obesity as compared to 23% and 7-41% of ischemic heart diseases and cancer burden, respectively. Studies have shown that being obese in adulthood is associated with increased risk of diabetes later in life.[15,19,20,49] Likewise, being obese as a child has been related to increased chances of being obese in adulthood.[4,7] What is less clear is the contribution of childhood obesity to diabetes risk later in life cumulatively and independently of adult adiposity.[9,50] Perhaps, one of the barriers of studying such long-term effects in an individual's life course is the relatively small number of birth cohorts with a long follow-up of over 30 years. Another potential obstacle is that most of these cohorts were started when obesity had not reached epidemic levels. One way to circumvent this limitation and follow children born when obesity rates are as high as today in order to assess subsequent diabetes risks to create a virtual (i.e. simulated) cohort that integrates the best available knowledge and evidence.

This research plans to study the life-course development of obesity on diabetes using an ABM implemented in a simulated birth cohort followed up to age 65 and empirically calibrated to vulnerable populations. Studying such long-term effects of obesity is of critical importance as it has implications for the timely primary prevention of chronic diseases during the life course.

1.6.3. Gap in large-scale intervention evaluation

Recent studies of childhood obesity have reported that the prevalence of obesity among children aged 2-19 years may be stabilizing, although questions remain about what intervention has led to this apparent stabilization.[51] Today, there is strong empirical evidence linking healthy behavior (e.g. exercising more or eating healthier) to lower risk of obesity.[52] What is less known are the long term effects of health interventions or policies aimed at improving those healthy behaviors on obesity in large-scale populations.[53–55]

This research proposes to evaluate in a large-scale population, key interventions aimed at halting or reversing childhood obesity using agent-based modeling and simulations. In an ideal world, one could conduct a randomized controlled trial which provides the best way to establish causality, though such endeavor may not always be feasible because of ethical considerations and cost in time and money. On the other hand, ABM allows one to conduct a virtual experimentation *in silico* using empirical knowledge on (i) intervention effects on behavior change and (ii) behavior effects on change in adiposity. In addition, an ABM gives the opportunity to test intervention trials for safety and efficacy before its implementation in the real-world. Above all, this research will give further insights into which interventions are likely responsible for the stabilizing rates and which one or combinations thereof can yield large-scale substantial and sustainable decreases in obesity prevalence and incidence. To illustrate the usefulness of an ABM in evaluating health interventions, we focused this research on Los Angeles County, one of the most populous and diverse counties of the US. We did so for the following reasons: (i) Los Angeles County has seen its rates of obesity almost double in the last two decades[56] (ii) there have been marked racial/ethnic disparities in obesity and[56] (iii) the Los Angeles County Department of Public Health has recently deployed major efforts to counter the

epidemic and address racial disparities in obesity.[57] Examples of potential interventions implemented in Los Angeles County and considered in this dissertation include those targeting nutrition-related behaviors such as breastfeeding promotion, and reduction of sugar-sweetened beverages and those targeting physical activity-related behavior such as access to recreational facilities and a pedestrian friendly community design.[58]

1.7. Overall and specific research aims

In this dissertation, we aim to apply systems science and causal analytical approaches to investigate policy-relevant life-course causes and consequences of obesity. This dissertation will (i) investigate the life-course effects of obesity on diabetes mellitus; (ii) explore the racial/ethnic differences in the effects of these interventions; and (iii), evaluate the overall effects of key interventions aimed at reversing or halting childhood obesity using agent-based modeling and implemented in a simulated birth cohort empirically calibrated to vulnerable populations in Los Angeles County.

Specific aim 1: To develop an agent-based model of a cohort of children born in Los Angeles County followed into adulthood in order to study the life-course development of obesity and of its effects on diabetes mellitus.

Specific aim 2: To investigate the contribution of childhood obesity to incident diabetes not due to adult adiposity overall and in different racial/ethnic groups.

Specific aim 3: To assess the overall impact and racial/ethnic differences in the impact of breastfeeding, reducing sugar-sweetened beverage consumption, increasing access to parks and recreational facilities and having a pedestrian-friendly community on obesity and diabetes throughout the life course.

Specific aim 4: To quantify the potential impact of various hypothetical and plausible behavioral interventions early in life on adiposity in a cohort of children aged 1-5 years enrolled in the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) in Los Angeles County.

1.8. Dissertation structure

We addressed this topical and methodological opportunity as follows. This dissertation is structured into seven chapters. In the first chapter (general introduction), we gave an overview of the obesity problem and its long-term effects on diabetes mellitus and related outcomes. We then identify the practice and research gaps in the literature and describe how this dissertation proposes to close these gaps. In chapter 2 (general methods), we provided an overview of the conceptual frameworks underlying our methodologies and describe the two main methods used in this dissertation: agent-based modeling and the g-computation algorithm. We then addressed the four aims of the dissertation in the subsequent chapters 3, 4, 5 and 6. In the final chapter (general discussion) we summarized our findings and implications of this research. An overview of the chapters is described in **Table 1.1**.

Table 1.1 Overview of the chapters of the dissertation showing the study characteristics

Chapters	Title	Exposures/Interventions	Outcomes	Target population	Method
3	An agent-based model of obesity and diabetes in the Virtual Los Angeles Cohort: The ViLA-Obesity simulation suite	Exclusive breastfeeding for ≥ 6 months Sugar-sweetened beverage consumption Physical activity Fast-food consumption Fresh fruit and vegetable consumption Neighborhood access to parks Neighborhood walkability Neighborhood supermarket density Neighborhood fast-food density	Obesity (2-65 years) Type 2 diabetes (18-65 years)	98230 simulated individuals aged 0-65 followed from 2009 to 2074 from the Virtual Los Angeles Cohort (ViLA)	Simulation
4	Modeling the role of childhood adiposity in the development of adult type 2 diabetes in a 64-year follow-up study: an agent-based simulation study	Childhood obesity	Type 2 diabetes (40-49 years)	98230 simulated individuals aged 6-49 followed from 2009 to 2074 from ViLA	Causal mediation analysis using the g-formula

Table 1.1 Overview of the chapters of the dissertation showing the study characteristics (continued)

Chapters	Title	Exposures/Interventions	Outcomes	Target population	Method
5	Evaluating the effectiveness of key health interventions on obesity and diabetes throughout the life course in the Virtual Los Angeles Cohort	Exclusive breastfeeding for ≥ 6 months Sugar-sweetened beverage consumption Physical activity Fast-food consumption Fresh fruit and vegetable consumption Neighborhood access to parks Neighborhood walkability	Obesity in childhood (2-65 years) Type 2 diabetes (18-65 years)	98230 simulated individuals aged 2-65 followed from 2009 to 2074 from ViLA	g-formula
6	Projecting the impact of early life interventions on adiposity in children living in low income households	Exclusive breastfeeding for ≥ 6 months Sugar-sweetened beverage consumption TV viewing Physical activity Fast-food consumption Fresh fruit and vegetable consumption	Mean weight-for-height Z score (1-5 years)	799 individuals aged 1-5 followed from 2008 to 2010 from the Special supplemental nutrition program for women, infant and children	g-formula

Chapter 2. General Methods

2.1. Conceptual frameworks

In this dissertation, a number of frameworks/theories were used to guide our modeling and analysis: the socio-cognitive theory (SCT), the socio-ecological model (SEM) and the life course perspective (LCP). Our modeling was also aided by using a systems science approach and our assumptions of hypothesized mechanisms were depicted using causal diagrams.

2.1.1. Social Cognitive Theory

A theory that has been used successfully to explain obesity-related behaviors and design interventions to address the obesity epidemic is the social-cognitive theory. In a systematic review conducted by Nixon et al, Bandura's Social Cognitive Theory (SCT)/Social Learning Theories (SLT) was the most common theory used in successful efforts to change dietary and physical activity behaviors.[59] This theory emphasizes that behavior changes as an interaction between personal, behavioral and environmental factors, a concept also referred to as reciprocal determinism. Other concepts in this theory that explain and help change behaviors include the behavioral capability (i.e. skills), anticipated outcomes (i.e. expectations), self-efficacy, and observational learning (e.g. modeling to encourage exploration to new foods).[33,35] The representation of the reciprocal determinism in social sciences parallel that of the epidemiological triad of causal factors in epidemiology wherein (i) the environment has the same meaning as in the former (ii) the agent's virulence, infectivity represent the behavioral factors such as eating calorie-dense foods, and (iii) the host (genetic susceptibility, nutritional status) has the same meaning as personal factors in the SCT model.[60] An example of the social cognitive theory applied to obesity is shown in **Figure 2.1**.

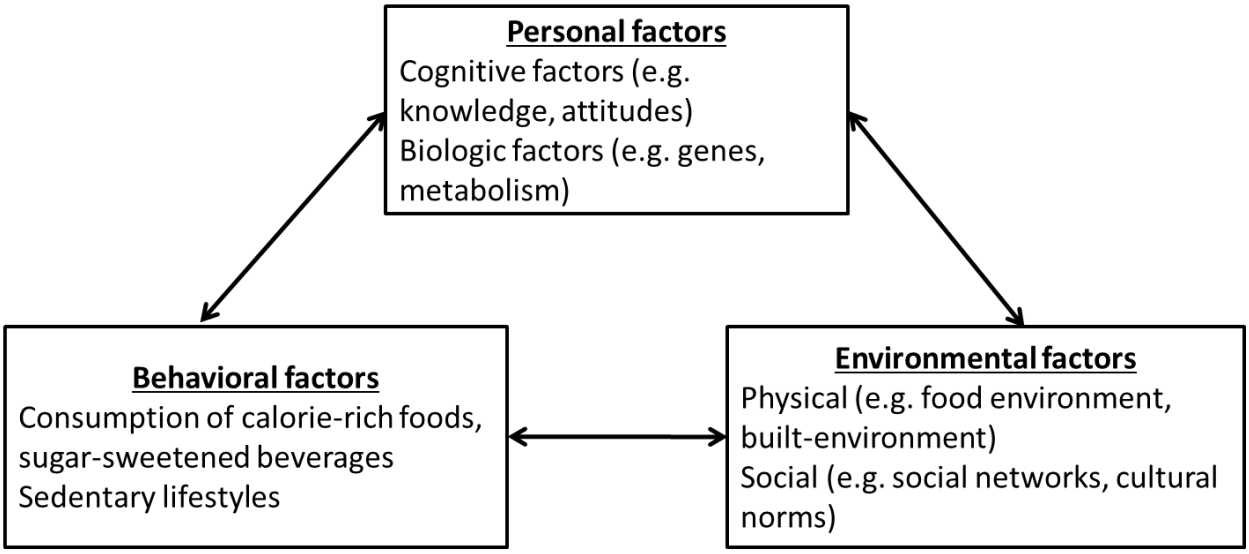


Figure 2.1 Social Cognitive theory: Reciprocal determinism occurring in behavior change during obesity

2.1.2. Socio-ecological model

In conceptualizing our agent-based modeling for exploring risk factors that lead to obesity in childhood, we also considered the socio-ecological model (SEM) of McLeroy and colleague (which is based on Bronfenbrenner's model)[61,62] which offers a multilevel and interactive approach that emphasizes the interaction between, and interdependence of, factors within and across all levels of health problems.[35] This framework suggests that the obesity epidemic not only results from individual behaviors as it relates to physical activity and dietary behaviors but also from the larger ecology within which the individual lives (e.g. food environment, built environment). The basic premise is that individual behaviors both shape and are shaped by the physical and social environment and this concept is termed as "reciprocal causation". [35] Typically, five levels are included: intrapersonal or individual factors (e.g.

knowledge, attitude, beliefs), interpersonal factors (e.g. family, friends), institutional or organization factors (e.g. schools, healthcare organization), community (e.g. social networks) and public policy factors (e.g. local policies) [35] A number of authors have used the SEM to study the obesity epidemic and design multi-level interventions pertaining to physical activity and dietary behaviors in an effort to reverse the current trends.[34,63,64] An example of the socio-ecological model applied to obesity is shown in **Figure 2.2**.

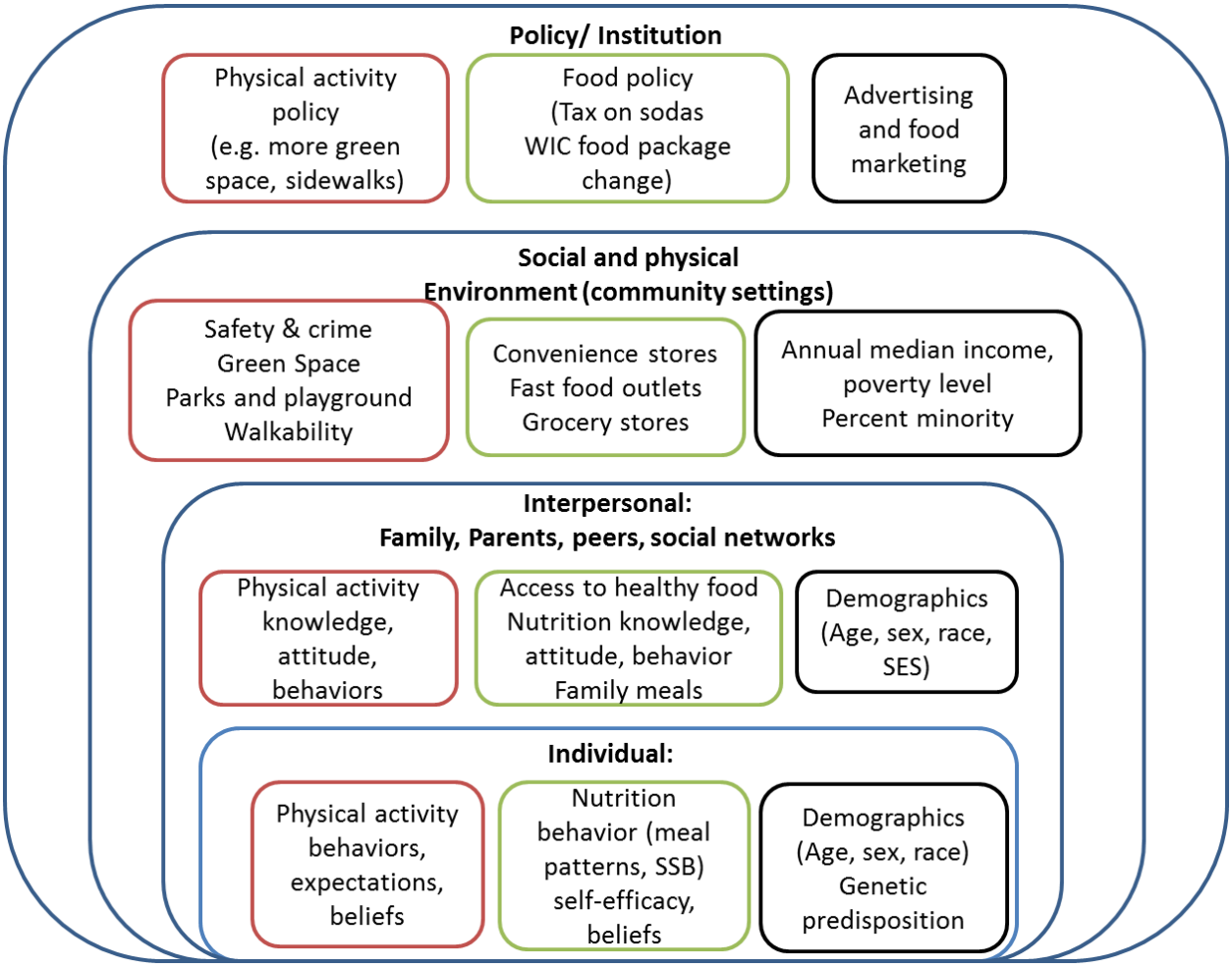


Figure 2.2 Example of the socio-ecological Model applied to obesity

2.1.3. Life course perspective

We used a life-course approach[65] of obesity and its risk factors since we were interested in the long-term effects of childhood obesity. Life-course approaches have been used for years in other fields but are only beginning to be used more frequently in epidemiology.[66] Specifically, Shlomon and Kuh define this field of life-course epidemiology as "the study of long-term effects on chronic disease risk of physical and social exposures during gestation, childhood, adolescence, young adulthood and later adult life.[67] This field or approach emphasizes that early exposures in the life course at different levels (e.g. individual, environmental) can have later life effects on disease risk. It emphasizes that (i) cumulative exposure can impact health over time (i.e. timeline concept) as is seen in obese children who remain obese in their adulthood; (ii) health trajectories can be affected during critical periods in life (i.e. the timing concept) such that earliest exposures in the life course at different levels (e.g. individual, environmental) can have later life effects on disease risk. An example of the latter is illustrated when children who are breastfed for a shortened period of time experience an elevated risk of becoming obese later in their lives.[15] Hence, we simulated a birth cohort to examine the long term effects of childhood obesity on diabetes risk. An example of the life-course perspective applied to obesity is shown in **Figure 2.3**.

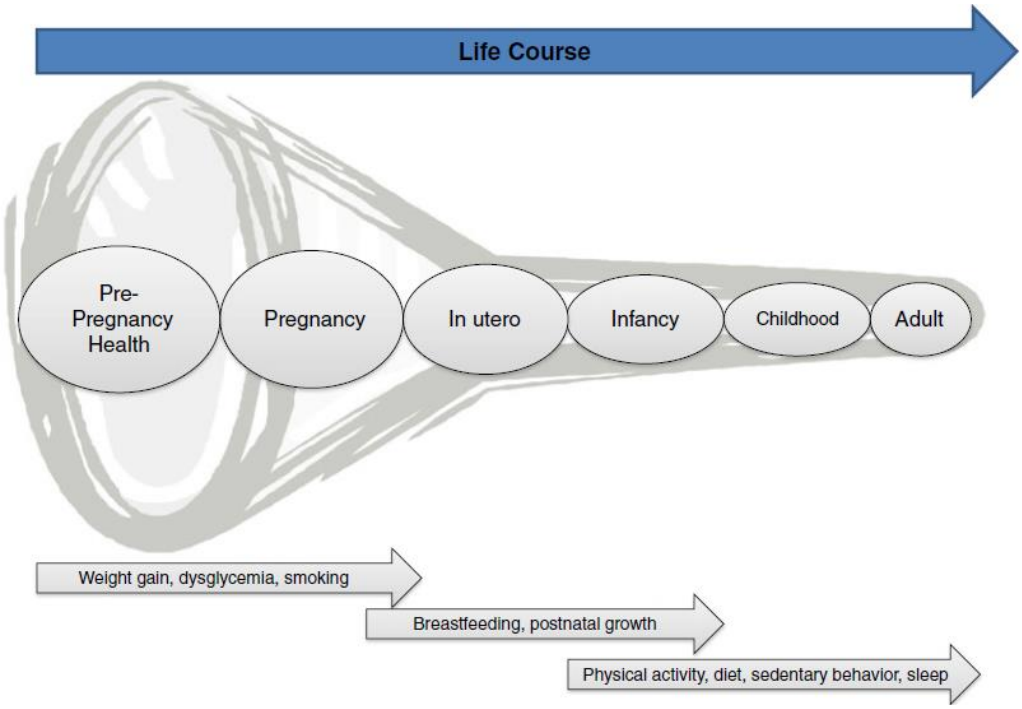


Figure 2.3 A life course approach to obesity: key health behaviors and risk factors from reference[68]

2.1.4. Pulling all theories together: A systems science approach

Because no single factor or theory can explain at one point in time the entire complexity of the obesity epidemic, finding a novel approach to investigate, and reverse this epidemic is critical. To-date, few obesity treatment approaches have been successful in sustaining weight loss. Hence, there is urgency to use population approaches to curb the obesity epidemic.[60] Such approaches are comprehensive and holistic, and allow the consideration of (i) multiple levels of scales (e.g. individuals, social and physical environment) described in the socio-ecological model; (ii) bidirectional interactions (i.e. feedback loops) between personal factors, behavioral factors and environmental factors as described in the socio-cognitive theory and (iii) a life-course approach (i.e. health trajectories and critical periods) to investigating these interactions over time from conception to death. An umbrella term for the study of such complex systems is

systems science. A number of public health researchers and practitioners have now begun to incorporate systems thinking into their understanding of the obesity epidemic.[13,39,46,48,69] To this end, Glass and McAtee designed a systems science model that incorporated (i) individual (genes, metabolism) and environmental (e.g. social and physical) influences on behaviors (e.g. dietary and physical activity behaviors) (ii) a vertical axis depicting a nested hierarchy of systems (e.g. biological, social, environmental) and (ii) a horizontal axis depicting the life course at the individual or environmental level).[13,70] Other more complex depictions of the obesity complexity exist but are far more involved and would not fit in this document.[71] An example of a systems-oriented approach to the study of obesity is shown in **Figure 2.4**.

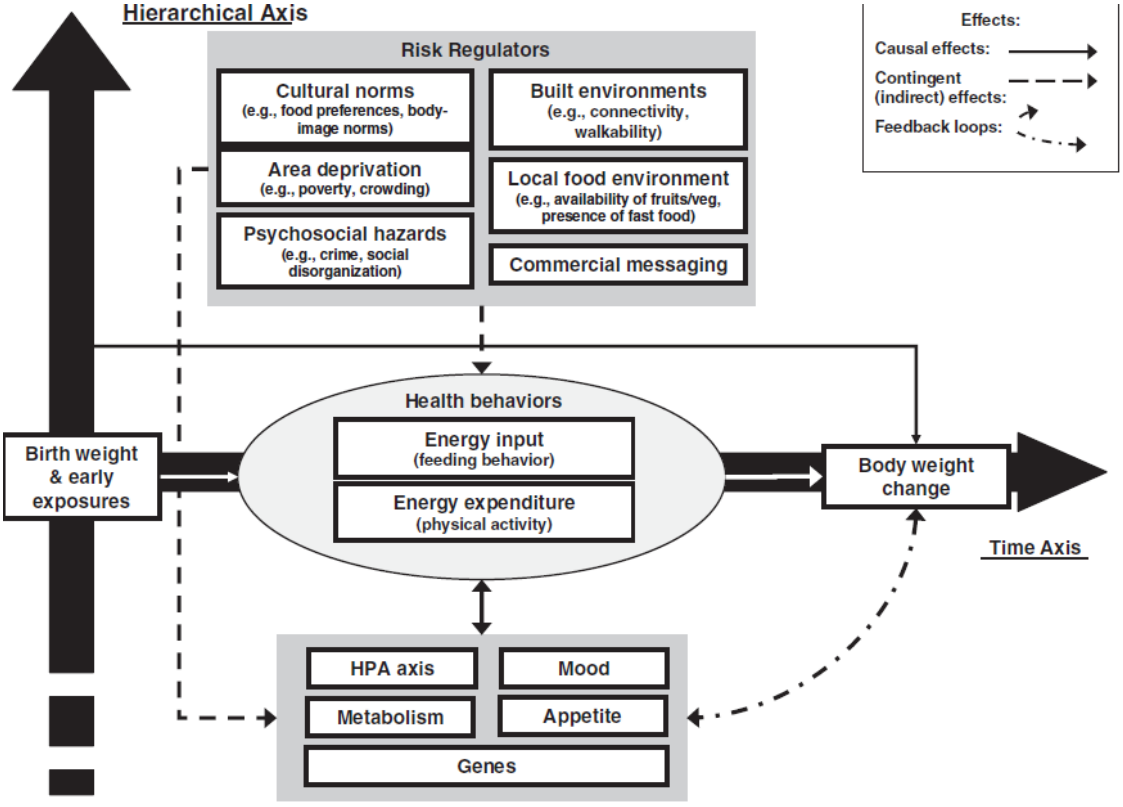


Figure 2.4 Systems-oriented, multilevel model applied to the study of obesity from reference [70]

2.2. Causal graphs

We used the causal diagram framework or Directed Acyclic Graph (DAG) developed by Pearl[72] to encode causal relationships between two variables X and Y. DAGs are acyclic (i.e. no feedback loops) path diagrams that depict our assumptions about relationships between variables and convey a set of mathematical and probabilistic rules. Biases such as confounding bias, collider bias as well as mediators can be depicted using DAGs. Absence of an arrow between two variables encodes the absence of a relationship between the two variables. Extensive descriptions of DAGs are available elsewhere but key points are highlighted below (see **Figure 2.5**) .[72–74] Let A_0 = exposure variable; M = mediator variable; Y = outcome; S = set of variables affected by both exposure A_0 and outcome Y; L = set of variables affecting both exposure A_0 and outcome Y. L is said to be a parent (or ancestor) of Y and Y is said to be a child (or descendent) of L.

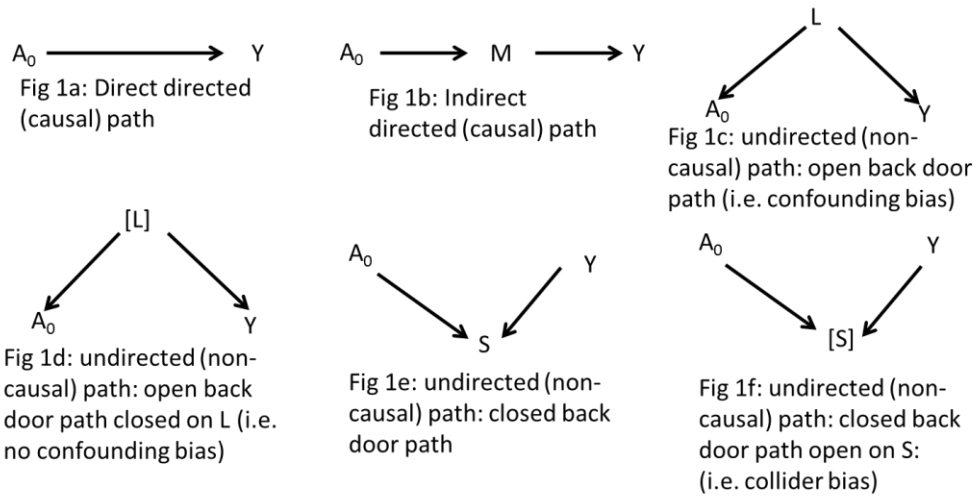


Figure 2.5 Directed acyclic graphs (DAGs) depicting different types of paths and relationships. The square brackets around a variable represent the conditioning or controlling of the variable.

2.3. General approach

We carried out this dissertation in four phases. First, we created a simulated birth cohort empirically calibrated to vulnerable populations in Los Angeles County using an agent-based model. This phase, grounded in the socio-ecological model and the social cognitive theory used a systems science approach. Second, within this virtual cohort we decomposed the effect of childhood adiposity on adult diabetes using the parametric g-computation algorithm (and based on the life-course perspective[65]). Third, we conducted within the simulated cohort a virtual experimentation of contextual and individual interventions relevant to Los Angeles County using the parametric g-computation formula. Last, we applied the g-formula to longitudinal data obtained from the Special Supplemental Program for Women, Infants and Children (WIC) to quantify the impact of various hypothetical and plausible interventions on adiposity.

2.3.1. Overview of agent-based modeling

Agent-based models (ABM) are computer representations of systems consisting of a collection of microentities (referred to as agents) interacting and changing over time and whose interactions give rise to macrosystems (e.g. obesity epidemic).[69,75–78] ABM is one of the most widely used systems science methods and is particularly useful when the system being modelled is one that is complex i.e. one that involves agents (e.g. individuals, organizations) (i) that are autonomous and heterogeneous (ii) whose environment (e.g. neighborhood) is crucial and not fixed and (iii) whose dynamic interactions between agents and with their environment give rise to an emergent phenomena that is complex and non-linear with feedback loops. [69,79] In particular, the ability of ABM to naturally describe obesity from the bottom-up[75] and its flexibility, confers to ABM its added value compared to other reductionist approaches [79,80]

which tend to simplify complex systems. These techniques often assume linearity (at least on some scale), normality, homogeneity, independence between individuals and over time and is most concerned with variables representing a single-level system.[69].In contrast, in agent-based modeling, the researcher relaxes these assumptions and allows the model to account for non-linearity and possible dependence between individuals and their environment. Nevertheless, ABM and conventional statistical analyses are complementary and were both used in this dissertation. For instance, parameter abstraction and effect estimation made use of multivariate regressions.

ABM has been used in a number of studies to study health behaviors such as physical activity, drinking and diseases such as diabetes and obesity.[81–83] Also termed individual-based modeling in ecology,[77] and extensively described by many authors [75,84,85], ABMs are composed of three key elements:

1. ***A set of agents*** that compose the system. Each agent has specific attributes (e.g. age, gender) and behaviors (e.g. going to school). An agent can be of different nature (e.g. individuals, communities, organizations).
2. *A* ***set of agent relationships*** and underlying conceptual models outlining how agents interact with each other and with their environment.
3. The ***agent's environment*** or topology (e.g. spatial location, lattice). This environment can be passive or active with its own dynamic properties and behavioral rules.

To calibrate, verify and validate (V&V) agent-based models, several iterative and successive steps are typically undertaken. The main steps typically taken in an ABM project are as follow: [85,86]

1. **Calibration (or parameterization):** This is the first step in ABM that is concerned with assigning baseline or trend characteristics (i.e. input parameters) to the virtual neo-system. Data are abstracted from real-world data or taken from other publicly available surveys and the literature.
2. **Verification:** In this step, we check the model for errors in coding and make sure that the model does what it is intended to do (e.g. correct calculations). Baseline output operations of the codes are then compared to the expectations stated in the design documents.
3. **Validation:** In this step, the investigator is concerned with how accurate the virtual system reflects the real-word system. For instance, *predictive validation*, (comparing predicted output data to the original real-world data) was used to validate our model.

2.3.2. Overview of the g-formula computation

In this study, we used the parametric g-formula,[87] a generalization of the standardization method for time-varying exposures and confounders, to estimate the effect of childhood adiposity on adult diabetes and to estimate the prevalence or incidence of diabetes mellitus under hypothetical interventions. The g-formula is particularly attractive for assessing the impact of hypothetical interventions on disease risk in longitudinal settings.[88–90]

The g-computation formula or algorithm (or g-formula for short) method is a general method, often seen as an alternative to the so-called inverse-probability-weighted (IPW) fitting of marginal structural models (MSMs) and g-estimation of structural nested mean model (SNMM).[91] Altogether, these methods are referred to as "g-methods" and have been developed by Robins to deal with time-varying exposures in complex longitudinal data. They are general

methods for estimating causal effects by deactivating biasing paths while leaving alone the desired paths.

One benefit of using the g-formula over conventional methods is that in addition to adjusting appropriately for time-varying confounding affected by prior exposures, it (i) naturally handles interventions on multiple risk factors (joint interventions) and interventions dependent on evolving risk factor values (dynamic interventions); (ii) can be used to estimate multiple parameters and (iii) compute population estimates of potential outcomes. On the other hand, it requires models for covariates and for the outcome and may be more sensitive to violations of assumptions of no unmeasured confounding, no measurement error and no model misspecification.[88,90] As described by Robins and illustrated by Daniel et al, we provide a brief description of the g-formula.[87,92]

- ❖ Illustrative DAGs

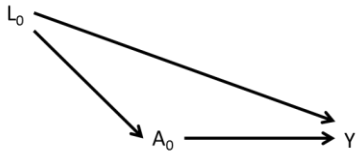


Figure 2.6 DAG with time-fixed confounding

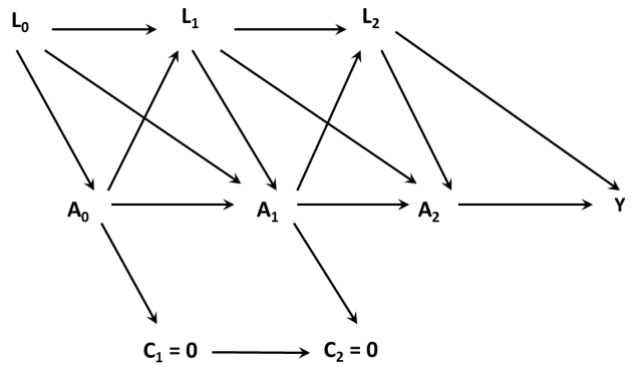


Figure 2.7 DAG for time-varying exposure and confounding. The subscripts represent the time points 0, 1 and 2; L = set of confounder or vector of covariate; A = exposure; Y = outcome; C_t = Censored in the period (t-1, t) due to loss-to-follow-up or competing risk (e.g. death to other causes)

❖ Mathematical expression and description

Let Y^{a0} be the expected value of Y in the population under the hypothetical intervention that every member of the population receives treatment a_0.

In a time-fixed confounding scenario as displayed in **Figure 2.6**, the g-computation formula is:

$$E(Y^{a0}) = \sum_{l0 \in L0} E(Y|A_0 = a_0, L_0 = l_0)P(L_0 = l_0)$$

In time-varying exposures and confounding as depicted by **Figure 2.7**, the g-computation formula is:

- For binary covariates

$$E(Y^{\bar{a}}) = \sum_{\bar{l} \in \bar{L}} E\left(Y\middle|\bar{A} = \bar{a}, \bar{L} = \bar{l}\right) \prod_{t=0}^{T} P(L_t = l_t \mid \bar{A}_{t-1} = \bar{a}_{t-1}, \bar{L}_{t-1} = \bar{l}_{t-1})$$

for binary covariates and including the censoring variable:

$$E(Y^{\bar{a}}) = \sum_{\bar{l} \in \bar{L}} E\left(Y\middle|\bar{A} = \bar{a}, \bar{L} = \bar{l}, \bar{C}_t = 0\right) \prod_{t=0}^{T} P(L_t = l_t \mid \bar{A}_{t-1} = \bar{a}_{t-1}, \bar{L}_{t-1} = \bar{l}_{t-1}, C_t = 0)$$

- For continuous covariates

$$E(Y^{\bar{a}}) = \int_{\bar{l}\in\bar{L}} E\left(Y\middle|\bar{A}=\bar{a},\bar{L}=\bar{l}\right)\prod_{t=0}^{T}\int L_t|\bar{A}_{t-1},\bar{L}_{t-1}\,(l_t\ ,\bar{a}_{t-1},\bar{l}_{t-1})d\bar{l} \quad (4)$$

The overbars indicate history for any time-dependent variable such that $\bar{A}_t$ to denote treatment history and $\bar{L}_t$ the covariate history up to and including time t. Capital letters denote random variables and lower case letters their realized values.

❖ The parametric estimation of the g-formula

Step 1: Parametric modeling

We fitted parametric models using the method described by Westreich et al [(90)]as follows:

- The density covariates measured at time *t* -1 (l_t) conditional on past covariate history through *t* – 1 surviving and remaining uncensored to time t.
- The probability of Y at time *t* conditional on past covariate history through *t*-1 following the intervention through t, surviving to time *t* and remaining uncensored to time *t* + 1.

Step 2: Monte Carlo simulation

To estimate the g-computation formula, one needs to sum over all possible $\bar{l}_t$. Due to the fact, that each $\bar{l}_t$ is a high-dimensional vector of covariates, and that $\bar{l}_t$ can be multivariate and/or non-binary , as *T* becomes large, it becomes computationally impractical to perform a direct calculation based on equations (1-4). We perform a Monte Carlo simulation at each time point t on a random sample from the total population under a given intervention to overcome the computational difficulty.

❖ Assumptions

We assume the presence of the following conditions whenever we used the g-formula:

- Conditional exchangeability i.e. no uncontrolled confounding conditional on the measured covariates.[93]
- The loss-to-follow-up mechanism follows a missingness-at-random (MAR) pattern, which means that conditional on staying in the study up to and including time *t*, and on $\bar{A}_t$ and $\bar{L}_t$ the probability that each subject remains in the study until at least time t + 1 is independent of all future variables. This is the same thing as assuming no selection bias.
- Positivity[94] (i.e. for every level of *L*, $P(A_t|L_t) > 0$) and consistency[95] (i.e. for every individual whose exposure status is A=a, his potential outcome Y_a under the intervention doA=a is equal to his observed outcome)
- No model misspecification.[88]

Chapter 3. An agent-based simulation model of obesity and diabetes in the Virtual Los Angeles Cohort: The ViLA-Obesity Simulation Suite

3.1. Abstract

Background: Obesity is the result of a complex interplay between individual and environmental factors that can occur early and throughout an individual's life course giving rise to many chronic conditions including type 2 diabetes mellitus. We developed an agent-based model of children born in Los Angeles County and followed from birth into adulthood to study the life course incidence and trends of obesity and its effect on type 2 diabetes mellitus risk.

Methods: We built the Virtual Los Angeles cohort (ViLA), a stochastic, dynamic, discrete time, agent-based model informed by various data sources and calibrated to the population of Los Angeles County in California. We simulated 98,230 inhabitants spread out in 235 neighborhoods in Los Angeles County. Each agent was followed from birth to middle adulthood, exhibited healthy and unhealthy behaviors and became obese and/or developed type 2 diabetes throughout their lifetime with specified predictive probabilities.

Results: The obesity age-specific incidence was generally increasing from about 10% to about 30% across the individual life span with two notable peaks at age 6-12 and 40-50; the age-specific incidence of type 2 diabetes was generally increasing from less than 2% at age 18-24 to reach a peak of about 25% at age 40-50. The 16-year risk and 48-year risk of obesity was 32.1%, 95%CI (31.8%, 32.4%) for children aged 2-17 years and 81%, 95%CI (80.8%, 81.3%) for adults aged 18-65. The 48-year risk of type 2 diabetes in the ViLA-Obesity model was 53.4% , 95%CI (0.53.1%, 0.53.7%) for adults aged 18 to 65 years and the average incidence rate was about 13 cases per 1000 persons-years: 95%CI (12.679, 12.897).

Conclusions: Our findings suggest that the incidence and prevalence of obesity and type 2 diabetes will continue to increase over time and with age. This experiment illustrates the usefulness of agent-based models in forecasting the future burden of disease within a population over time and reinforces the need for timely preparedness and effective interventions to curb the epidemic.

Keywords: agent-based model, simulation, obesity, diabetes, life-course, cohort, Los Angeles

3.2. Introduction

Obesity is a major public health problem affecting millions of Americans with two in three adults and one in three children considered overweight or obese. [2]This condition disproportionately affects lower-income minority and disadvantaged groups [2] giving rise to health disparities. Obesity has been on the rise for the past few decades [2,96] despite ongoing prevention efforts warranting its description as a pervasive and complex phenomenon.[13,39] The proximal cause of obesity is an imbalance between energy intake and energy expenditure.[8] Energy intake and energy expenditure are dependent on dietary and physical activity behaviors respectively which are determined by a multitude of factors in a complex system. These factors can be investigated using the socio-ecological framework[62] which identifies three issues at the core of the pervasiveness and the complexity of the obesity problem: (i) behaviors (e.g. unhealthy eating and physical inactivity, etc.) in childhood and adolescence can influence our behaviors later in life [97,98] and the weight we put on as a result of these past unhealthy behaviors is more likely to track into adulthood (i.e. feedback loop or time-dependency complexity) [7,18,99]; (ii) the neighborhood we live in (aka built-environment) can influence our ability to make healthy or unhealthy choices (i.e. cross-level or hierarchical complexity) [100–102] and lastly (iii) our peers and family can also influence what we eat and whether we exercise regularly (i.e. social network, or interference complexity).[103]

These individual and environmental factors can occur early and throughout an individual's life course affecting the individual's health trajectory and later health outcomes, one of the most predominant being type 2 diabetes.[104] In fact, obesity (and overweight) is considered one of the most potent risk factor for type 2 diabetes.[104] Almost 80 to 90% of type 2 diabetes patients are overweight or obese. This is alarming as type 2diabetes is a disabling

disease that imposes considerable burden on individuals, families, communities and the health system. In 2002, the direct medical and indirect expenditures attributable to diabetes amounted to approximately $132 billion.[105]

As many authors have recognized today, the complex and pervasive nature of obesity calls for equally complex methods namely the use of an agent-based model.[13,39] An agent-based model (ABM) has been defined as a computer representation of systems consisting of a collection of micro-entities (referred to as agents) interacting and changing over time and whose interaction give rise to macro-systems.[69,79] In other words, an ABM is a simulation model in which individuals represented by agents are given certain characteristics and whose behavior can be influenced by (i) their past behavior, (ii) the environment they live in (iii) and the agents around them.[106] Essentially, building an agent-based simulation model is akin to creating a virtual world that resembles reality using our best available knowledge about what governs individual behaviors and how these behaviors are influenced by our past behaviors, the place we live in and people around us. An example of such a virtual world is the Archimedes diabetes model,[107] a simulation model aimed at addressing clinical problems and questions that has been validated against and replicated the results of eighteen randomized controlled trials.[108] This type of simulation endeavor could be instrumental in forecasting the future state of a particular outcome given current knowledge.

Therefore, to study the obesity epidemic, its effect on health during an individual's life course and health disparities in obesity, we focused our attention on one of the most populous and most ethnically diverse counties in the United States namely, Los Angeles County, California.[109] With its rates of obesity on the rise and its marked racial/ethnic disparities in obesity,[57] Los Angeles County represents a suitable candidate for implementing our study. It is

also timely since the Los Angeles County Department of Public Health has been leading major efforts to curb the epidemic by identifying obesity-related risk factors and consequences and promoting healthy behaviors both at the individual level and through environmental changes. [57] The overarching goal of this study is to develop an agent-based model of a cohort of children born in Los Angeles County and followed into adulthood in order to study the life-course development of obesity and of its effects on diabetes mellitus. Specifically, we will describe the model and forecast the burden of obesity and type 2 diabetes in our Virtual Los Angeles population.

3.3. Methods

We developed the ViLA–Obesity model, a stochastic, dynamic, discrete time, agent-based model informed by various data sources and calibrated to the population of Los Angeles County in California to explore the incidence and trends in obesity and type 2 diabetes.

- ***Simulated population: ViLA***

According to the 2010 US Census, Los Angeles County was inhabited by 9,818,605 individuals who lived in 2,346 census tracts. [110] In this model, as it is the case in some other studies,[111] we considered a census tract to represent a neighborhood. We simulated 235 neighborhoods with 418 inhabitants per neighborhood for a total simulated population of 98,230. These numbers were obtained by dividing the actual population and neighborhood size by a 100 so that the virtual population would represent a 100th of the Los Angeles County total population rounded to the nearest ones) (See **Table 3.1**). Simulated individuals in the model are referred to as agents. Each agent lived in a specific neighborhood and was simulated from birth to middle adulthood in 10 discrete time-steps (See **Table 3.2**).

While a full-fledged ABM ideally incorporates all three levels of complexities as described in the introduction (i.e. time-dependency, cross-level interaction and interference), we have at this stage of the model building incorporated two of the three domains of complexities (time-dependency and cross-level interaction). We did so for the following reasons. While there have been some suggestions that obesity can spread through social networks (i.e. induction or person-to-person spread),[103] other authors have demonstrated that such effects may in fact be the result of confounding by contextual exposures (e.g. food environment, built-environment).[112] In fact, after properly accounting for environmental exposures, the social network effects in obesity almost vanished.[112] This finding, however, did not mean that peer support could not enhance the effectiveness of certain prevention efforts.[113]

Nevertheless, future iterations of the model may explore in sensitivity analyses, the added insights gained when incorporating an interference component. Currently, ViLA has three nested hierarchical level: the neighborhood or environment where the agents live, the agent itself and the time points.

Model specification

Agent

Each simulated agent had three domains of attributes. The first domain was the agent's socio-demographics (i.e. age, sex, socio-economic status [SES], race/ethnicity and marital status) representing the individual's inherent susceptibility which was not allowed to change with the exception of age (i.e. time-invariant variables). We assumed that individuals born in a certain SES group will remain in that group until the end of the simulation (i.e. inherit their parents' SES). Agents could get married after their eighteenth birthday (with a certain probability

obtained from external sources) but did not become single after being married (see **Table 3.3**).The second domain was the agent's behaviors divided into: (i) dietary behaviors (breastfeeding, fast-food consumption, sugar-sweetened beverage consumption [SSB], fresh fruit and vegetable consumption); (ii) physical activity behaviors (moderate-to-vigorous physical activity) and (iii) other behaviors (smoking, alcohol consumption). (See **Table 3.4**) The last domain was the agent's outcomes (BMI, and type 2 diabetes status).

Agents were only allowed to engage in smoking, alcohol consumption and develop type 2 diabetes after their eighteenth birthday. Both behavior and outcome domains were considered time-varying variables. For children aged 0-19, we defined overweight and obesity using the WHO BMI Z-score international child cutoffs. BMI Z-scores are more appropriate for defining obesity in children because they consider age-related biological changes in growth patterns and body composition and recognize that the relationship between BMI and adiposity varies because of these biological changes.[114] We calculated BMI Z-scores using CDC's SAS codes. [115] Based on the WHO growth charts, a child with a BMI Z-score (BMIz) less than -2 was classified as underweight; a BMIz greater or equal to -2 but less than 1 was classified as normal-weight; a BMIz greater or equal to 1 but less than + 2 was classified as overweight and a BMIz greater or equal to 2 was classified as obese.[116].

Similarly, an individual with a BMI less than 18.5 was classified as underweight; a BMI greater or equal to 18.5 but less than 25 was classified as normal-weight; a BMI greater or equal to 25 but less than 30 was classified as overweight and a BMI greater or equal to 30 was classified as obese. [117]

❖ *Neighborhood (built-environment)*

The neighborhood where the agents lived in also had three domains. The first domain was the neighborhood socio-demographics encompassing the proportion of individual self-identified as non-White, the proportion of individuals living below the federal poverty level (FPL) and the proportion of individuals who had a bachelor's degree or higher (See **Table 3.3**).The second domain was the neighborhood physical activity opportunities that comprise the neighborhood walkability and access to parks. Walk Score® is a validated commercial walkability measurement tool that has been developed to measure neighborhood walkability and pedestrian friendliness by analyzing population density and road metrics.[118–120] Briefly, the walk score was calculated based on the distance from a specific address to various amenities such that amenities within a 5-minute walk (.25 miles) were given maximum points whereas no points were given after a 30-minute walk. It has also been shown to be related to health indicators such as physical activity.[121] Typically, scores at or above 70 were considered very walkable to "walker's paradise" (i.e. most to all errands could be accomplished on foot) and scores below 70 were considered car-dependent to somewhat walkable (almost all to most errands required a car).[118] Park access was defined as the percent of population living within a quarter-mile buffer as was done in.[122] (See **eTable 3.5** in appendix)

The third domain was the neighborhood food environment comprising the supermarket and the fast-food density. These data were obtained from the business listings developed by Dun and Bradstreet and available through Walls and Associates' National Establishment Time-Series (NETS). The densities were calculated by dividing the count of supermarkets and fast-food outlets per census tracts by the census tract square mileage.[111,123] (See **eTable 3.5** in appendix)

❖ *Conceptual model, equations and decision rules*

The decision rules underlying this model were mainly based on mathematical equations. Completely exogenous variables in this model were few and limited to individual- and neighborhood-level socio-demographics. Except at birth (time t=0), all behavior equations (e.g. SSB, physical activity) had a common form whereby the dependent variable would be a function of the following: intercept, lagged version of the dependent variables and socio-demographics. Likewise, the outcome equations (e.g. BMI, type 2 diabetes) had in addition to the previous ones listed all age-specific behaviors (e.g. SSB, physical activity, smoking). Linear and logistic regressions were used for modelling continuous and binary dependent variables, respectively. Accordingly, the inverse of the link functions used in the regression modeling were used for simulation (i.e. identity and expit functions respectively). The built-environment with its attributes is first simulated, then agents with their attributes by time period are simulated within neighborhoods. These will engender a change in BMI and will subsequently affect diabetes risk. Most endogenous variables allow for time-dependency (i.e. previous behavior affecting future behavior). Features of feed-back were also allowed. For instance, when BMI changed, it affected subsequent ability to exercise which subsequently affected future BMI and so on.[124] A detailed description of the equation structure are presented in the appendix (See **eTable 3.1**)

- ***Model calibration, verification and validation***

We undertook several iterative steps to build the ViLA-Obesity model. First, we obtained parameters (i.e. proportions, means, standard deviations of each variable and the regression

coefficients relating any two variables) from multiple studies and datasets. Second, we calibrated and internally validated our model while verifying the program for errors throughout the process.

- *Data sources and parameters*
  - *Proportions, means and standard deviations*

The proportions, means and standard deviations of the individual-level socio-demographics and those of the neighborhood-level socio-demographics were obtained from the American Community Survey (ACS) (see **Table 3.3**). The individual-level race and income group were derived respectively from the neighborhood-specific race percentage and percent below federal poverty level (FPL). The proportions, means and standard deviations of the individual-level exposures and outcomes (breastfeeding, SSB, physical activity, fast-food consumption and fruit and vegetable consumption, smoking, alcohol consumption, body mass index [BMI], type 2 diabetes) were obtained from the California Health Interview Survey (CHIS), the Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO) (see **Table 3.4**).

  - *Parameters for effect and association measures*

These regression coefficients were taken from various sources detailed in the appendix (See **eTable 3.2** for a general outline described in the appendix)

For clarity, we defined three levels of evidence. 'Evidence level 1'parameters are defined as parameters that are directly taken, in this order of preference, from published systematic reviews and meta-analyses, randomized control trial studies or cohort studies. 'Evidence level 2' parameters are defined as parameters that are directly taken from cross-sectional studies from the

peer-reviewed literature. The third level of evidence represents parameters computed (indirectly obtained) by our research team using merged publicly and privately available data (e.g. American Community Survey, National Establishment Time-Series (NETS), Walkscore.com, WHO, WIC, National Health and Nutrition Examination Survey [NHANES]). Ideally, all parameters would be coming from 'evidence level 1' but because most studies do not report on the relationships between covariates such as age, sex, race, SES and the outcome and between the covariates and the exposures, we identified other sources of evidence.

In addition, in some cases, we have mathematically converted estimates obtained from the literature to fit our model needs using simulations where necessary. These instances include but are not limited to: converting a mean difference into a risk ratio or an odds ratio; converting a rescaled effect into an appropriate scaled effect; converting an estimate obtained using a continuous predictor into an estimate that would be obtained using a categorical version of the predictor; converting proportions obtained from a contingency table into an effect measure such as an odds ratio and converting a weight difference into a BMI difference by dividing it by a common age-group specific height.

To obtain the effect or association regression coefficients between the individual-level exposures (breastfeeding, SSB, physical activity, fast-food consumption and fruit and vegetable consumption) and the individual-level outcomes (BMI, type 2 diabetes), we used parameters from 'evidence level 1' parameters (See **eTable 3.3** in appendix). To obtain the effect or association regression coefficients between the neighborhood-level exposures (walkability, park access, supermarket density, fast-food density) and the individual-level outcome (physical activity, fast-food consumption, fruit and vegetable consumption), we used parameters from our defined 'evidence-level 2' parameters (see **eTable 3.4** in appendix).

To obtain the effect or association regression coefficients between the neighborhood-level socio-demographics (percent non-White, percent below FPL, percent bachelor graduates) and the neighborhood-level exposures (walkability, park access, supermarket density, fast-food density), we used parameters from our defined 'evidence-level 3' parameters (see **eTable 3.5** in appendix).

Lastly, to obtain the effect or association regression coefficients between the individual-level covariates (sex, race, marital status, SES, smoking, alcohol consumption, family history of diabetes) and the individual-level outcomes (BMI, type 2 diabetes) and between the individual-level covariates (sex, race, marital status, SES) the individual-level exposures (breastfeeding, SSB, physical activity, fast-food consumption and fruit and vegetable consumption), we used parameters from our defined 'evidence-level 3' parameters (see **eTable 3.6**, **eTable 3.7** in appendix).

❖ *Calibration and validation*

Many commonly used validation techniques [125] could not be used here because we did not have a base cohort in Los Angeles that followed individuals from birth to adulthood and studied our exposures and outcomes of interests. Nevertheless, we used a rigorous approach to calibrate our model to the population of Los Angeles.

We used a "calibration-in-the-large" technique to calibrate and validate our model.[125] The "calibration-in-the-large" is a calibration whereby one ensures that the mean predicted outcome equals the mean observed outcome (i.e. mean($\hat{Y}_{predicted}$) = mean($Y_{observed}$)) through the fine tuning of the intercept.[125] To get an estimate of the unknown intercept α, we used Robins' g-testing method. [126] Through simulation and "grid search" (i.e. testing a range of potential α

values), the value at which mean($\hat{Y}_{predicted}$) = mean($Y_{observed}$) was retained as an estimate of the intercept α.

A similar approach was used to obtain the regression coefficients of the lagged dependent variables. Briefly, we first predicted the intercept at time *t=1* from an initial form of the equation at time *t=0* as the intercept at time *t=0* plus one unit age effect. Note that at time *t=0*, there would be no lagged dependent variable. We subsequently estimated the regression coefficient of the lagged dependent variables using a technique similar to the "calibration-in-the-large" technique to ensure that mean($\hat{Y}_{predicted}$) = mean($Y_{observed}$) through the fine tuning of the regression coefficient. Lastly, we verified the programs to check for errors in coding and ensure that the models computed the intended outputs.

The finding of the equality mean($\hat{Y}_{predicted}$) = mean($Y_{observed}$) insured the internal validity of the model testifying that there was agreement between the observed data and our model predictions (i.e. internal validation). In this iteration of the ViLA-Obesity model we did not perform any external validation because we were specifically interested in building a cohort that would characterize the population of Los Angeles County. To extend the model to other populations, we could adjust our intercepts to match the site-specific observed prevalence.[125]

- ***Overview of the model simulation and statistical analysis***

ViLA-Obesity represents a simulation model or suite within our ViLA platform. It integrates trends in the causes and consequences of obesity, focusing on diabetes as a key obesity consequence during the life course. Each agent is simulated from birth to middle adulthood (aged 60 to 65 years) in ten discrete time steps representing critical life stages. At each time step the agent's age is simulated using a uniform distribution bounded within the specific critical life

stages. Our ViLA cohort represents a closed cohort where everyone is followed until middle adulthood, that is, to *time=9*. (See **Table 3.2**) During the simulation, agents were born in a specific neighborhood, exhibited healthy and unhealthy behaviors (e.g. SSB, physical activity, smoking), gained/lost weight and developed type 2 diabetes with a certain probability dependent on the agent's current state. The agents were allowed to change neighborhoods in a specified way. In fact, the neighborhood identification would remain the same but the characteristics of the neighborhood where the agent would live could change once, between 18-24 years and between 40-50 years to reflect the possibility that during these life stages, agents could potentially move. We used Monte Carlo simulation within the SAS 9.4 software (Cary, NC). All data preparation and analysis were also done in SAS. (See **Figure 3.1, Figure 3.2** and **Figure 3.3**)

3.4. Results

Calibration and validation

We compared our simulations results to observed data to calibrate and validate our model. The observed means and proportions representing the population of Los Angeles County were taken whenever available from the 2009 CHIS data. We performed the calibration by age interval and plotted the simulated and observed means and proportions in **Figure 3.4**. Our simulation results broadly matched the age-specific means and proportions from CHIS 2009. However, there were some small but notable departure from the observed data for physical activity, fresh fruit and vegetable consumption, smoking and diabetes prevalence.

Trends in obesity and type 2 diabetes

Figure 3.5 depicts the overall and racial subgroup trends (incidence and prevalence) in obesity and type 2 diabetes over time in the ViLA Obesity model.

We found that the obesity age-specific incidence proportion was generally increasing from about 10% to about 30% across the individual life span with two notable peaks at age 6-12 and 30-39. Likewise, the age-specific incidence proportion of type 2 diabetes increases from less than 2% at age 18-24 to reach a peak of about 25% at age 40-50.

The prevalence of obesity was highest in childhood with about 25% of children considered obese between the age of 6 and 12 years. During adulthood, the prevalence of obesity rose to reach a maximum of 40% between the age of 60 and 65 years.

Compared to Whites, the incidence and prevalence of obesity and type 2 diabetes were generally higher among the non-White subpopulation. There were marked disparities in the prevalence of type 2 diabetes compared to that of obesity. The racial disparity gap in the prevalence of type 2 diabetes was greatest during middle adulthood but that in the prevalence of obesity was small but more uniform across ages.

Trends in drivers of health behaviors

Figure 3.6 shows the overall and racial subgroup trends in key health behaviors. The consumption of fast-food was generally high and decreasing with age. It was highest during childhood and adolescence with approximately 75 to 85% of children and adolescents consuming fast-foods more than one time per week. The consumption of sugar-sweetened beverage was also generally high and decreasing with age. It was highest during childhood and adolescence with approximately 60% to 70% of children and adolescents consuming more than one 12-oz drink of

SSB per day. Engaging in moderate-to-vigorous physical activity was generally low and decreasing with age. It was lowest during adolescence with only about 20% of adolescents engaging in moderate-to-vigorous physical activity. The consumption of fresh fruits and vegetables was fairly constant over time. It was lowest during childhood with only about 40 to 50% of children aged 6 to12 consuming more than five servings of fruit and vegetables per day. About one out of five individuals were breastfed for six months or longer during their first year of life.

Cumulative incidence and average incidence rate of obesity and type 2 diabetes in the ViLA-Obesity model

Table 3.5 presents the cumulative incidence and average incidence rate of obesity and type 2 diabetes in the ViLA Obesity model

Type 2 diabetes: The 48-year risk or cumulative incidence of type 2 diabetes in the ViLA-Obesity model was 53.4% , 95%CI (0.53.1%, 0.53.7%) and the average incidence rate of type 2 diabetes was about 13 cases per 1000 persons-years: 95%CI (12.679, 12.897) for adults aged 18 to 65 years.

Obesity: The 16-year risk or cumulative incidence of obesity was 32.1%, 95%CI (31.8%, 32.4%) and the average incidence rate of obesity was about 22 cases per 1000 persons-years, 95%CI (22.034, 22.526) for children aged 2 to17 years. The 48-year risk or cumulative incidence of obesity was 81% 95%CI (80.8%, 81.3%) and the average incidence rate of obesity was about 28 cases per 1000 persons-years, 95%CI (27.766, 28.154) for adults aged 18 to 65 years.

Table 3.1 Los Angeles County, California actual and simulated population sizes

	Actual population[(110)]	***Simulated population***
Number of census tracts	*2,346*	*235*
Population Density	*4,185.25*	*418*
*Number of census tracts * population density*	*9,818,605*	*98,230*

Note: The simulated population represents a 100th of the LAC total population rounded to the nearest ones. Population density was calculated by dividing the total population size by the number of census tracts.

Table 3.2 Life period, time-points and age-group

Life period	**Time**	**Age group**
Birth	0	0-1
Early Childhood	1	2-5
Middle Childhood	2	6-12
Adolescence	3	13-17
Young Adulthood	4	18-24
Young Adulthood	5	25-29
Young Adulthood	6	30-39
Middle Adulthood	7	40-49
Middle Adulthood	8	50-59
Middle Adulthood	9	60-65

Table 3.3 Input parameters for the distribution of individual and neighborhood time-invariant variables

Variables	***Values: Mean (SE) or %***	***Sources***	***Years***	***Distribution***	***Bound***
Neighborhood-level parameters					
Percent non-White	*0.72 (0.26)*	*ACS, 2014* [110]	*0-65*	*Normal*	*[0,1]*
Percent below federal poverty level	*0.19 (0.13)*	*ACS, 2014* [110]	*0-65*	*Normal*	*[0,1]*
Percent bachelor graduates graduate or above	*0.28 (0.21)*	*ACS, 2014* [110]	*0-65*	*Normal*	*[0,1]*
Individual-level parameters					
Sex	*Male: 49%*	*ACS, 2014* [110]	*0-65*	*Bernoulli*	*[0,1]*
Race (Whites vs Non-Whites)	*Derived from neighborhood specific percent non-White*	*ACS, 2014* [110]	*0-65*	*Bernoulli*	*[0,1]*
Income group (Below or at FPL vs. Above FPL)	*Derived from neighborhood specific percent below the federal poverty level*	*ACS, 2014* [110]	*0-65*	*Bernoulli*	*[0,1]*
Marital Status	*Married: 44%*	*ACS, 2014* [110]	*18-65*	*Bernoulli*	*[0,1]*

SE: Standard error; FPL: Federal Poverty Level; ACS: American community survey

Table 3.4 Input parameters for the distribution of individual time-varying variables

Variables	Values (varies by age group)	Sources	Years	Distribu-tion	Bound
Breastfeeding Breastfed exclusively for six months or more	0-1 year: 22%	CDC[127]	0-1	Bernoulli	[0,1]
Fast-food consumption Ate fast-food more than one times (1 to 4) in past week	2-5 years: 67% 6-12 years:76% 13-17 years: 84% 18-39 years: 76% 40-65 years: 61%	CHIS, 2009 [128]	2-65	Bernoulli	[0,1]
Moderate-to-vigorous physical activity Physically active at least one hour per day for 7 days [age 2-17 years] Moderate physical activity >=30 min/day for 5 days (including walking): [age 18-65 years]	2-5 years: 31% 6-12 years:22% 13-17 years: 13% 18-39 years: 31% 40-65 years: 24%	CHIS, 2009 [128]	2-65	Bernoulli	[0,1]
Sugar-sweetened beverage consumption Drank one or more glasses of soda or other sugary drinks yesterday	2-5 years: 67% 6-12 years:76% 13-17 years: 84% 18-39 years: 76% 40-65 years: 61%	CHIS, 2009 [128]	2-65	Bernoulli	[0,1]
Fresh fruit and vegetable consumption Ate five or more servings of fruits and vegetables	2-5 years: 62% 6-12 years:44% 13-17 years: 42% 18-39 years: 49% 40-65 years: 53%	CHIS, 2009 [128]	2-65	Bernoulli	[0,1]
Smoking Current smoker	18-39 years: 24% 40-65 years: 15%	CHIS, 2009 [128]	18-65	Bernoulli	[0,1]
Alcohol consumption Binge drinking	18-39 years: 76% 40-65 years: 61%	CHIS, 2009 [128]	18-65	Bernoulli	[0,1]
Type 2 diabetes Yes	18-39 years: 1.4% 40-65 years: 13.3%	CHIS, 2009 [128]	18-65	Bernoulli	[0,1]
Body mass index (kg/m2)	0-1:birth: 16.33 (1.49) 2-5 years: 16.41 (1.99) 6-12 years:19.18 (4.66) 13-17 years: 23.69 (5.73) 18-39 years: 27.85 (6.90) 40-65 years: 30.23 (6.90)	WHO[116] LAHANES,2011 [129] NHANES[130]	0-65	Normal	

Table 3.5 Incidence rates and cumulative incidence of obesity and type 2 diabetes in the ViLA-Obesity model (n=98,230)

	ALL			
	Obesity (2-65)	**Obesity Childhood (2-17)**	**Obesity Adulthood (18-65)**	**Type 2 diabetes adulthood (18-65)**
Total number	98230	98230	98230	98230
Events	87625	31544.000	79606	52426
Person-years (py)	3183963	1415891	2847196	4099783
Incidence rate (per 1000 py)	27.521 (27.339, 27.704)	22.279 (22.034, 22.526)	27.959 (27.766, 28.154)	12.788 (12.679, 12.897)
Cumulative incidence	0.892 (0.890, 0.894)	0.321 (0.318, 0.324)	0.810 (0.808, 0.813)	0.534 (0.531, 0.537)
	Whites			
	Obesity (2-65)	**Obesity Childhood (2-17)**	**Obesity Adulthood (18-65)**	**Type 2 diabetes adulthood (18-65)**
Total number	35862	35862	35862	35862
Events	31072	10023	28067	14162
Person-years (py)	1245482	523022	1090448	1571629
Incidence rate (per 1000 py)	24.948 (24.672, 25.227)	19.164 (18.792, 19.543)	25.739 (25.440, 26.042)	9.011 (8.864, 9.161)
Cumulative incidence	0.866 (0.863, 0.870)	0.280 (0.275, 0.284)	0.783 (0.778, 0.787)	0.395 (0.3899, 0.400)
	Nonwhites			
	Obesity (2-65)	**Obesity Childhood (2-17)**	**Obesity Adulthood (18-65)**	**Type 2 diabetes adulthood (18-65)**
Total number	62368	62368	62368	62368
Events	56553	21521	51539	38264
Person-years (py)	1938481	892869	1756748	2528154
Incidence rate (per 1000 py)	29.174 (28.934, 29.415)	24.103 (23.783, 24.427)	29.338 (29.086, 29.592)	15.135 (14.984, 15.288)
Cumulative incidence	0.907 (0.904, 0.909)	0.345 (0.341, 0.349)	0.826 (0.823, 0.893)	0.614 (0.610, 0.617)

The incidence measures were calculated for first-time diagnosis of obesity or T2DM among at-risk individuals.

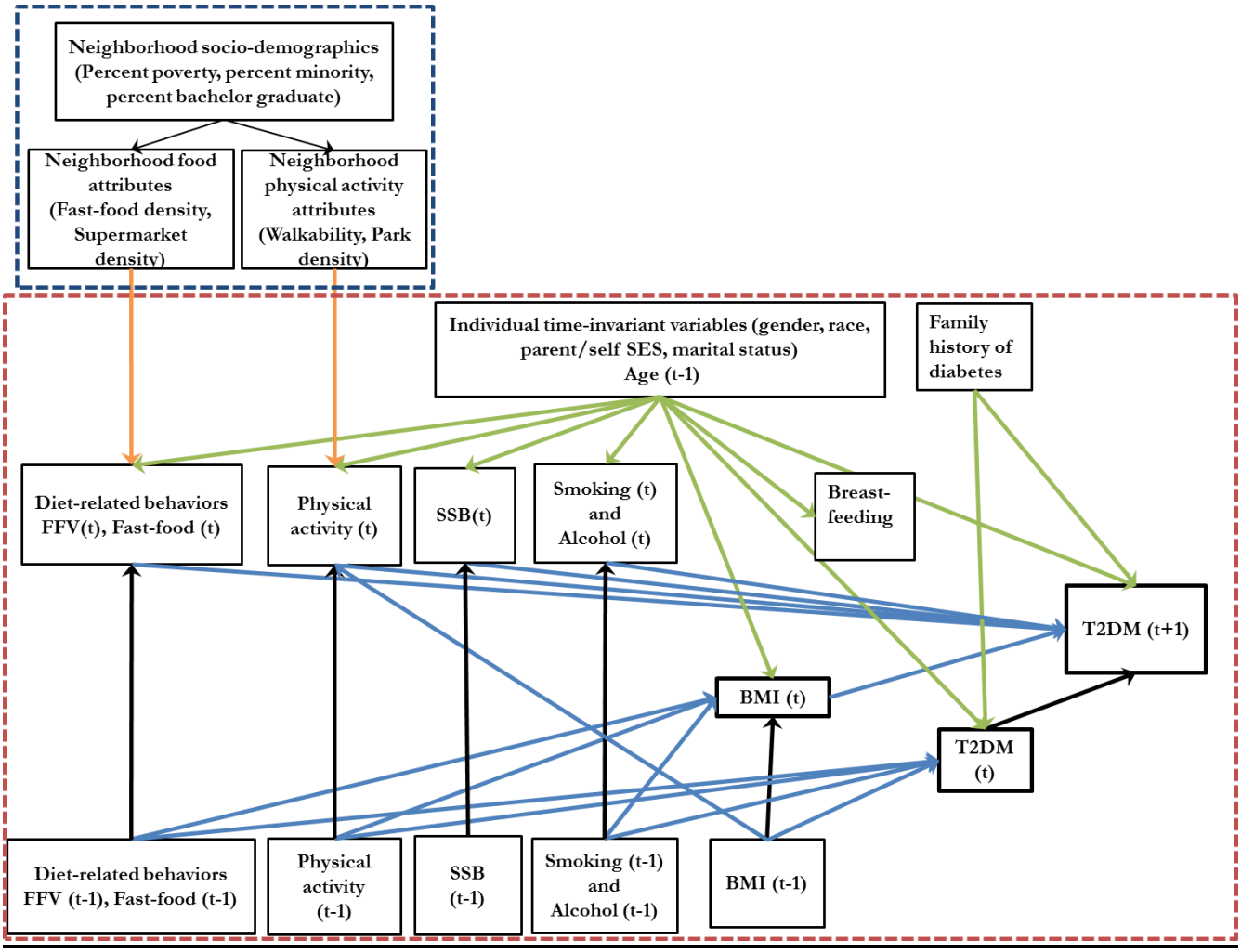


Figure 3.1 Conceptual directed acyclic diagram underlying the data-generating process. SSB: sugar-sweetened beverage consumption; BMI: body mass index; FFV: Fresh fruit and vegetable consumption; T2DM: type 2 diabetes; Ado: Adolescence. T is an index of time. The dark blue dotted square represents the neighborhood variables and the red dotted square represents the individual level variables. “Evidence level 1” parameters are represented by the blue arrows, “evidence level 2” parameters are represented by the orange arrows and “evidence level 3” parameters are represented by the green arrows. The black arrows represent calibrated regression parameters.

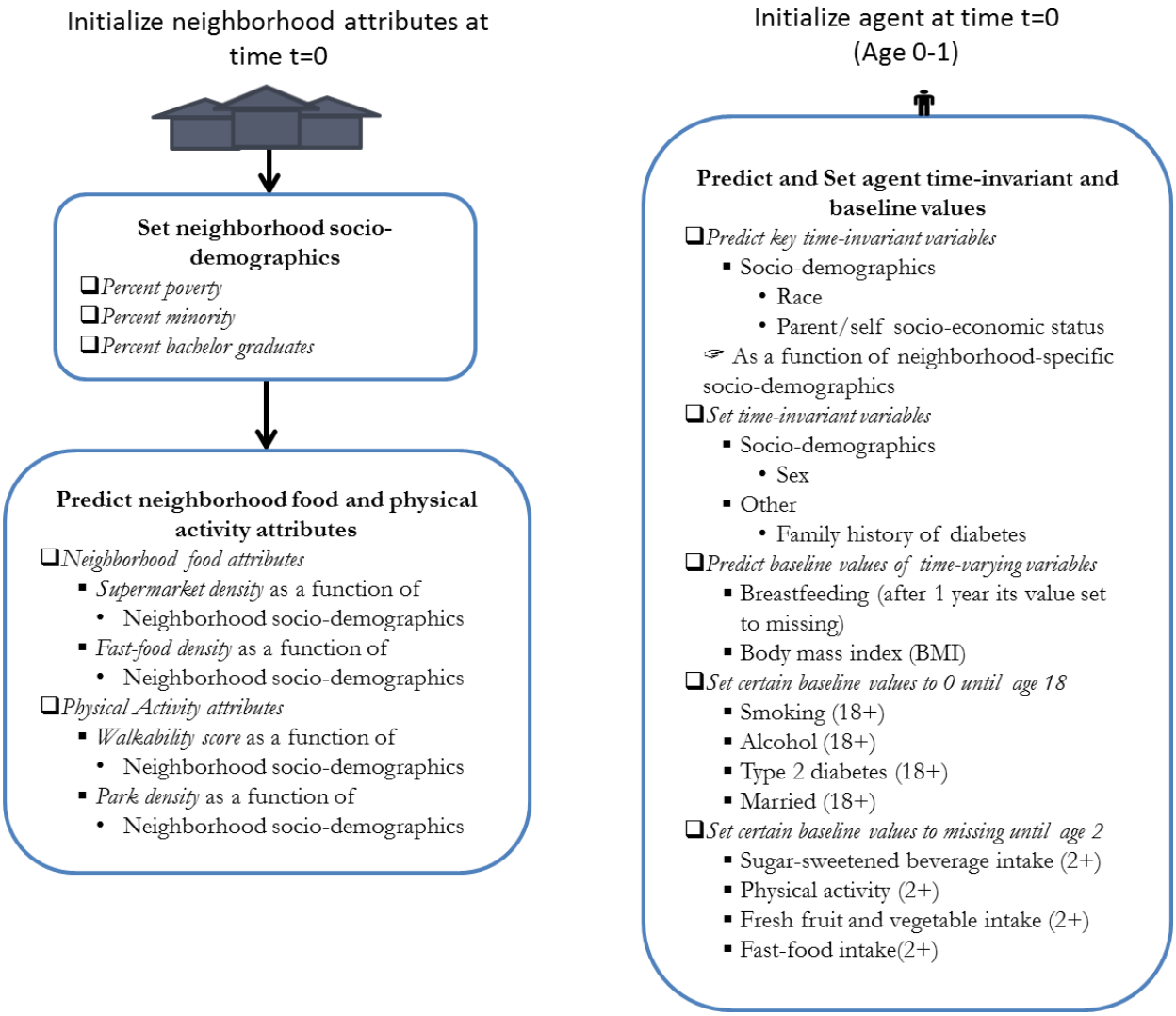


Figure 3.2 Model initialization diagram of the ViLA-Obesity model

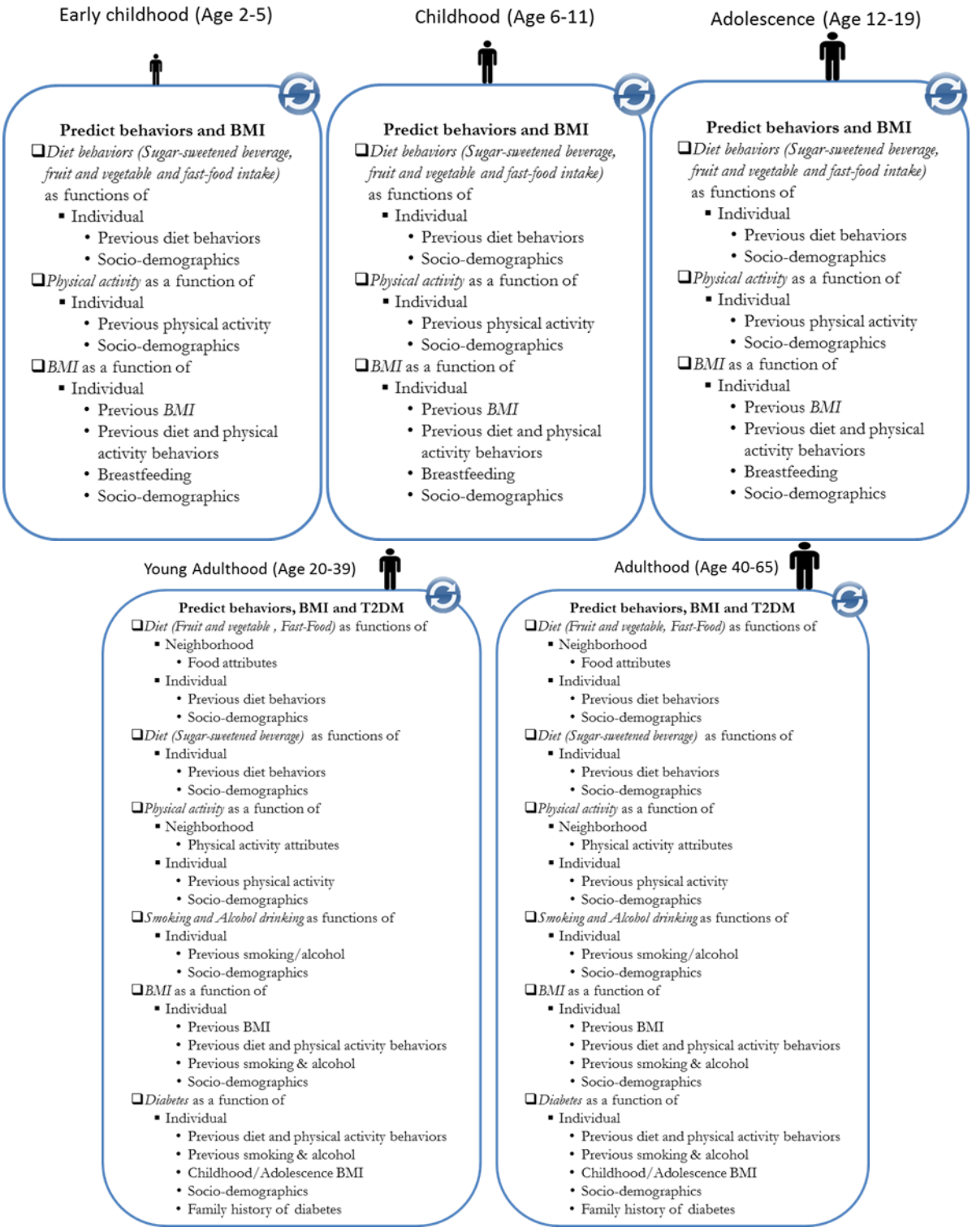


Figure 3.3 Model execution diagram of the ViLA-Obesity model, BMI: body mass index; T2DM: type 2 diabetes; SES: Socio-economic status

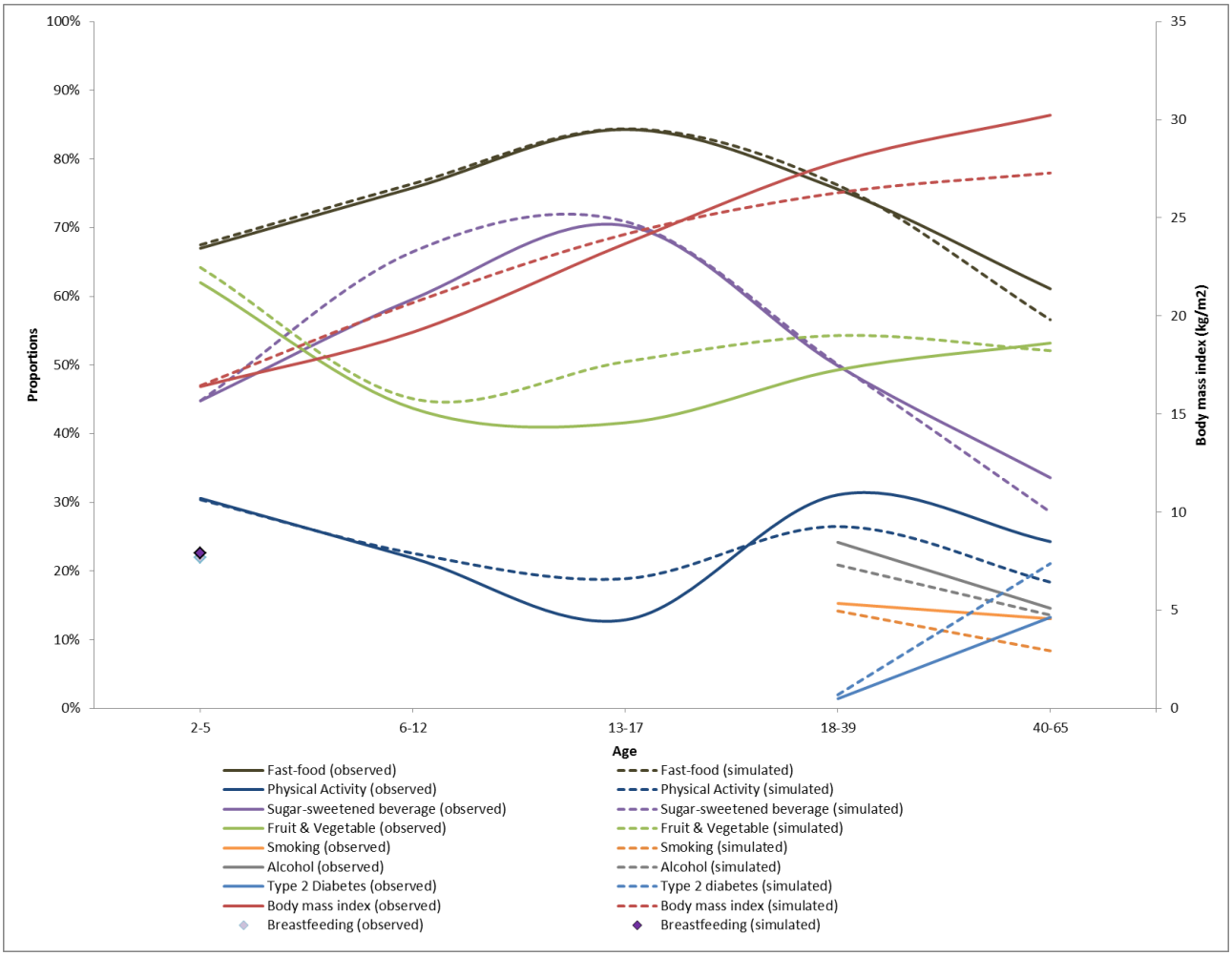


Figure 3.4 Calibration of the ViLA-Obesity model. The figure depicts observed (plain lines) and simulated data (dotted lines)

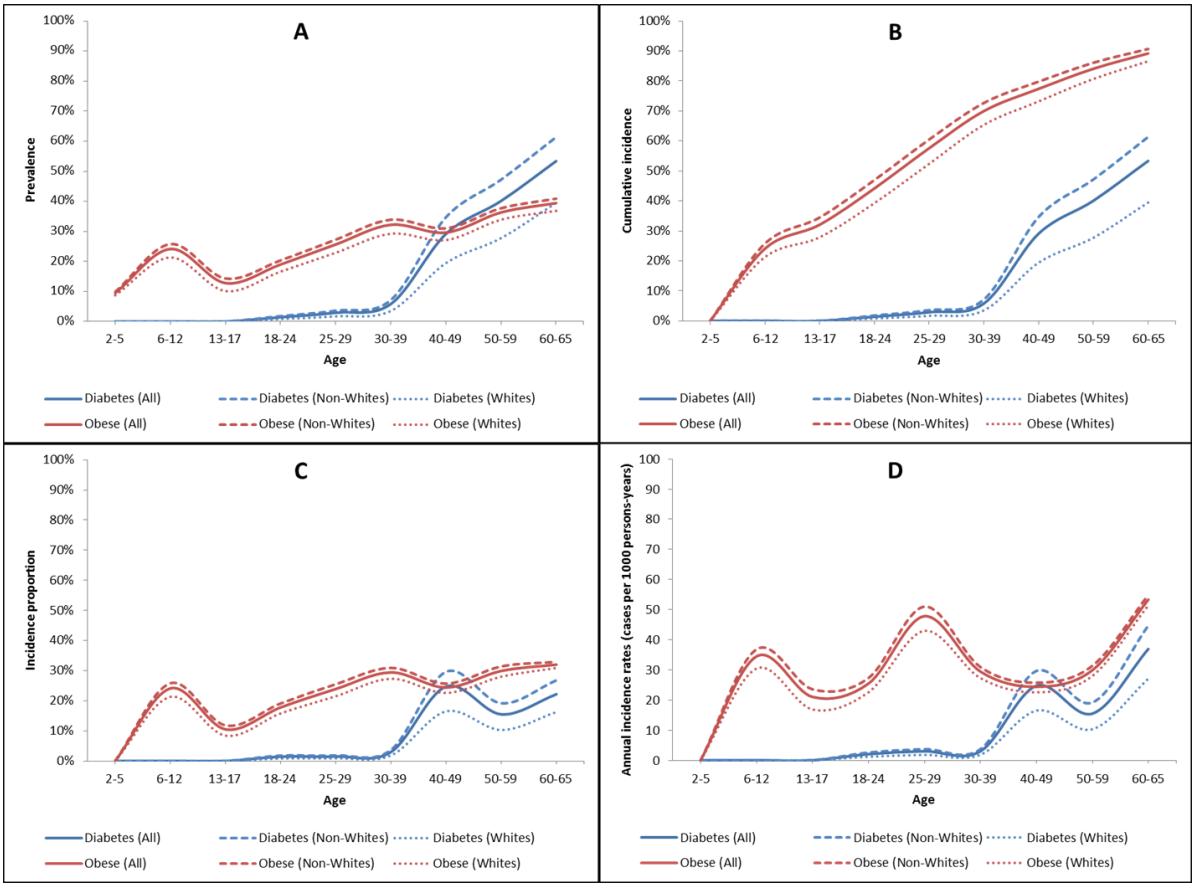


Figure 3.5 Obesity and type 2 diabetes prevalence (A), cumulative incidence (B), age-specific incidence proportion (C), and annual incidence rates (D) in the ViLA-Obesity model

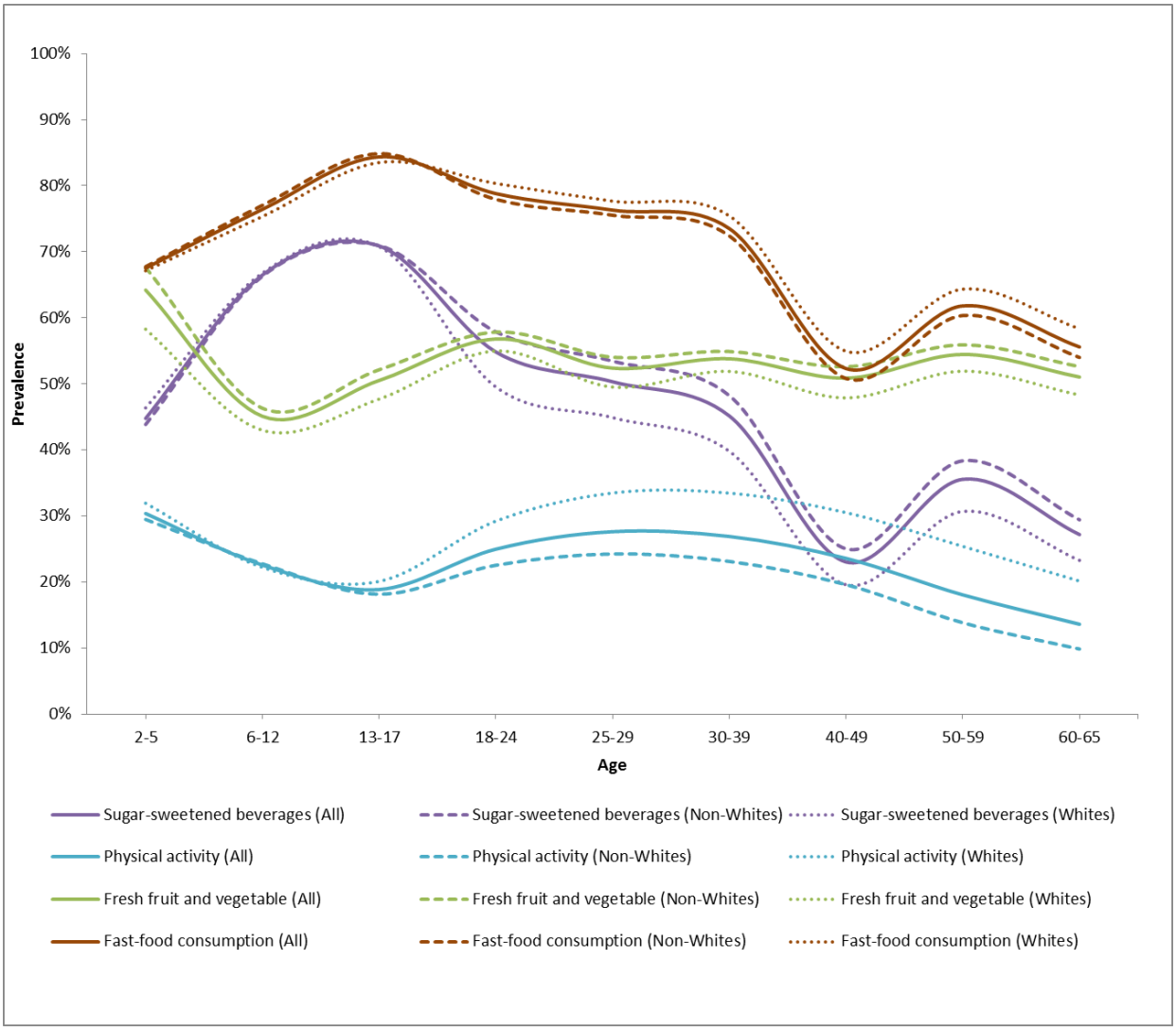


Figure 3.6 Proportion of obesity- and type 2 diabetes-related health behaviors over time in the ViLA-Obesity model

3.5. Discussion

The purpose of this study was to build an agent-based model of a cohort of children born in Los Angeles County and followed from birth into adulthood in order to study the life course development of obesity and of its effects on diabetes mellitus.

Our findings suggest that the incidence and prevalence of obesity and type 2 diabetes within the ViLA-Obesity model were generally high and increasing over time during the individual life span. The prevalence of obesity was highest during childhood and among individuals in their 30s while the prevalence of type 2 diabetes started rising among individuals in their 40s. In addition, one in three children and adolescents and four in five adults will become obese before age 65 and one in two adults will develop type 2 diabetes before age 65 in the simulated cohort. There were some racial differences in the prevalence and incidence of obesity and type 2 diabetes. The non-White subpopulation experienced higher proportions of individuals who became obese or developed type 2 diabetes at any point in time throughout the 64-year follow-up compared their White counterparts. The presence of such racial disparities in obesity and type 2 diabetes has been well documented in Los Angeles. [57,131]

Furthermore, our results also suggested that the proportion of individuals engaging in moderate-to-vigorous physical activity and consuming at least five servings of fresh fruit and vegetables was generally low while the proportion of individuals consuming fast-food and drinking sugar-sweetened beverages was generally high within the simulated cohort. There were also some racial differences among these obesity-related health behaviors. Among the Non-White subpopulation, there was a lower proportion of individuals who engaged in moderate-to-physical activity and a higher proportion of individuals who drank more than one sugar-sweetened beverage a day compared to their White counterparts. In contrast, among the White

subpopulation, there was a lower proportion of individuals who ate fresh fruit and vegetables and a higher proportion of individuals who ate fast-food more than once per week compared to their non-White counterparts.

This study provided a unique perspective of the development of obesity and type 2 diabetes among individuals who would have been followed from birth into adulthood in Los Angeles. This approach allowed us to simultaneously appreciate the aging effect on and forecast the future burden of obesity and type 2 diabetes within a birth cohort between 2009 and 2074 (i.e. 2009+65), something that has seldom been done in the literature. One consequence of this is that, unless done for calibration purposes, we should be cautious when comparing our estimates to past and projected prevalence and incidence of obesity and diabetes. The reason for this is that many trend estimates are based on cross-sectional data reflecting a given period effect and are often averaged across several age-groups and birth cohorts.[132][133] Nevertheless, these past and projected trends remain important for gaging the current and potential future state of obesity and diabetes in Los Angeles and the US. For instance, in 2011, the prevalence of obesity was 22.4% among children and 23.6% among adults[57] and the prevalence of diabetes was 9.9%[131]among adults in Los Angeles County. In the absence of projection studies in Los Angeles County, one can look to regional and national projection data to better appreciate the burden of disease attributable to obesity and type 2 diabetes. In fact, the UCLA Health forecasting tool, a simulation model that simulated individual life course among California's adult population, predicted that the obesity and type 2 diabetes prevalence will reach 9.93% and 30.8% by 2020 in their baseline scenario.[134] In addition, other projection studies based on nationally representative data found that the prevalence of impaired glucose tolerance could reach 15% by 2048 [135] and that the prevalence of obesity could reach 51.1% by the year 2030.[136] The latter

study also predicted that 80%, 90% and 100% of Americans will become obese by the year 2072, 2087 and 2102, respectively and that the non-White subpopulation may reach those levels sooner compared to Whites.[136]. Interestingly, when using the linear annual rate of increase reported in that study and the prevalence of obesity among adults in Los Angeles in 2011, we estimated that the projected prevalence of obesity in 2074 would be approximately 67%. Lastly, the predicted life-time risk of diagnosed diabetes from age 20 was estimated to be about 40% for men and women in a nationally representative sample.[137] All of these projections reflect similar alarming trends as suggested by our model and their insights warrant immediate action to reverse or slow the epidemic in the US and in Los Angeles County in particular. To that effect, the Los Angeles County Department of Public Health has made many health recommendations to prevent obesity and type 2 diabetes.[57] These recommendations have shown modest but promising impacts *in silico* through systems dynamic simulations.[138] Unlike agent-based models, systems dynamics models are less flexible and unable to track the impacts of interventions implemented at the individual level.[139] As a result, we will evaluate key health interventions implemented at the individual level in Los Angeles using the ViLA-Obesity agent-based model.

This study has several limitations. First, the calibration and validation of the ViLA-Obesity model was suboptimal in the absence of a base cohort in Los Angeles that followed individuals from birth to adulthood and studied our exposures and outcomes of interests. Nevertheless, we used age-group-specific means and proportions from publicly available data (i.e. CHIS) representing whenever available the population of Los Angeles County in 2009. This has some limitations since it does not allow one to disentangle the cohort effects from the age effects. However, in the absence of longitudinal data, using age-group specific data in a specific year appears to be a better alternative than using repeated cross-sectional data to calibrate our

model. In fact, although both methods ignore the cohort effects, using repeated cross-sectional surveys would not allow one to disentangle the age and period effect. Second, while we have incorporated relevant obesity-related environmental exposures, we did not account for the possibility of residual social network effect in this iteration of the model. We hope to explore the added insights gained from incorporating social network effects in the next iteration of the model. Third, the ViLA-Obesity model represented a simplified version of the Los Angeles County population in that the simulated cohort was closed, that is agents could not drop out, die, experience a competing risk, beget children, move in and out of the cohort.

The next phases of this project are numerous and include but are not limited to (i) assessing the role and mechanisms through which childhood adiposity affects type 2 diabetes risk in adulthood independently of adulthood adiposity; (ii) evaluating and comparing the periodic and long-term impact of proposed theoretical interventions implemented singly or in combination. Ultimately, ViLA will be progressively and continuously updated to include other disease processes (e.g. cardiovascular diseases) and will, as Archimedes, serve as a virtual laboratory that represents reality as close as possible given current knowledge.[108]

Conclusion

We developed and validated a virtual cohort representing Los Angeles County wherein we explored the development of obesity and diabetes from birth to adulthood. Our findings suggest that the incidence and prevalence of obesity and type 2 diabetes within the ViLA-Obesity model were generally high and increasing with age during the individual life span. In this virtual Los Angeles, one in three children and adolescents and four in five adults will become obese before age 17 and age 65 respectively and one in two adults will develop type 2 diabetes before age 65.

We also noted the presence of racial disparities in obesity, type 2 diabetes and obesity-related behaviors. This experiment illustrates the usefulness of agent-based models in forecasting the future burden of disease within a population over time and reiterates the need for effective interventions to curb the epidemic.

3.6. Appendix

eTable 3.1 Simplified equation structure underlying the model

$SSB_t = B(1,\ expit\ (Int_{SSB} + \beta_{SSBt-1} * ssb_{t-1} + \beta_{AGE} * age_t + \beta_{MALE} * male + \beta_{NONWHITE} * nonWhite + \beta_{LOWINC} * lowinc + \beta_{MARRIED} * married\ ))$
$FFD_t = B(1,\ expit\ (Int_{FFD} + \beta_{FFDt-1} * ffd_{t-1} + \beta_{AGE} * age_t + \beta_{MALE} * male + \beta_{NONWHITE} * nonWhite + \beta_{LOWINC} * lowinc + \beta_{MARRIED} * married + \beta_{EnvFFD} * EnvFFD\ ))$
$MVPA_t = B(1,\ expit\ (Int_{MVPA} + \beta_{MVPAt-1} * mvpa_{t-1} + \beta_{AGE} * age_t + \beta_{MALE} * male + \beta_{NONWHITE} * nonWhite + \beta_{LOWINC} * lowinc + \beta_{MARRIED} * married + \beta_{EnvPRK} * EnvPRK + \beta_{EnvWLK} * EnvWLK))$
$FFV_t = B(1,\ expit\ (Int_{FFV} + \beta_{FFVt-1} * ffv_{t-1} + \beta_{AGE} * age_t + \beta_{MALE} * male + \beta_{NONWHITE} * nonWhite + \beta_{LOWINC} * lowinc + \beta_{MARRIED} * married + \beta_{EnvSMD} * EnvSMD\ ))$
$ALC_t = B(1,\ expit\ (Int_{ALC} + \beta_{ALCt-1} * alc_{t-1} + \beta_{AGE} * age_t + \beta_{MALE} * male + \beta_{NONWHITE} * nonWhite + \beta_{LOWINC} * lowinc + \beta_{MARRIED} * married\ ))$
$SMK_t = B(1,\ expit\ (Int_{SMK} + \beta_{SMKt-1} * smk_{t-1} + \beta_{AGE} * age_t + \beta_{MALE} * male + \beta_{NONWHITE} * nonWhite + \beta_{LOWINC} * lowinc + \beta_{MARRIED} * married\ ))$
$BMI_t = N(Int_{BMI} + \beta_{BMIt-1} * bmi_{t-1} + \beta_{SSBt-1} * ssb_{t-1} + \beta_{FFDt-1} * ffd_{t-1} + \beta_{MVPAt-1} * mvpa_{t-1} + \beta_{FFVt-1} * ffv_{t-1} + \beta_{AGE} * age_t + \beta_{MALE} * male + \beta_{NONWHITE} * nonWhite + \beta_{LOWINC} * lowinc + \beta_{MARRIED} * married\ , SD\_BMIt^2)$
$T2DM_t = B(1,\ expit\ (Int_{D2M} + \beta_{BMIt-1} * bmi_{t-1} + \beta_{BMI\_Ado} * bmi_{Ado} + \beta_{SSBt-1} * ssb_{t-1} + \beta_{FFDt-1} * ffd_{t-1} + \beta_{MVPAt-1} * mvpa_{t-1} + \beta_{FFVt-1} * ffv_{t-1} + \beta_{AGE} * age_t + \beta_{MALE} * male + \beta_{NONWHITE} * nonWhite + \beta_{LOWINC} * lowinc + \beta_{MARRIED} * married + \beta_{FamD2M} * famd2m\ ))$ *For those with* $T2DM_{t-1} = 0$

Note that β represents a generate notation for regression coefficients and is expected to differ across equations and age-groups (i.e. at birth, early childhood, middle childhood, adolescence, young adulthood and middle adulthood). Expit is the inverse function of the log-odds or logit function. EBF: Exclusive breastfeeding; FFD: Fast-food consumption; MVPA: Moderate-to-vigorous physical activity; SSB: Sugar-sweetened beverage consumption; FFV: Fresh fruit and vegetable consumption; SMK: Smoking; ALC: Alcohol consumption; EnvWLK: Environmental or neighborhood walkability; EnvPRK: Environmental or neighborhood park Access; EnvSMD: Environmental or neighborhood supermarket density; EnvFFD: Environmental or neighborhood fast-food density; BMI: body mass index; T2DM: type 2 diabetes; Ado: Adolescence. T is an index of time

eTable 3.2 General outline for the data sources of parameters for effect and association measures

Relations	*Variables*	*Evidence type*
Individual-level exposures to individual-level outcomes	*Individual-level exposures* • *Breastfeeding* • *Sugar-sweetened beverage* • *Physical activity* • *Fast-food* • *Fruit and vegetable* *Individual-level outcomes* • *BMI* • *Type 2 diabetes mellitus*	*Evidence-level 1* *RCTs, Systematic Reviews, Meta-analyses, cohort studies* ➔ *From the literature*
Neighborhood-level exposures to individual-level exposures	*Neighborhood-level exposures* • *Walkability* • *Park access* • *Supermarket density* • *Fast-food density* *Individual-level exposures* • *Physical activity* • *Fast-food consumption* • *Fruit and vegetable consumption*	*Evidence-level 2* *Cross-sectional studies* ➔ *From the literature*
Neighborhood-level socio-demographics to neighborhood-level exposure relations ***Covariates-Exposures Covariates-Outcomes relations***	*Neighborhood-level socio-demographics* • *Percent non-White* • *Percent below federal poverty level* • *Percent bachelor graduates* *Neighborhood-level exposures* • *Walkability* • *Park access* • *Supermarket density* • *Fast-Food density* *Individual-level covariates* • *Sex* • *Race* • *Marital status* • *Low-income status* • *Family history diabetes* *Individual-level exposures* *Individual-level outcomes*	*Evidence-level 3* *Merged publicly available survey data (ACS, NETS, WalkScore data, NHANES)* ➔ *From our analysis*

eTable 3.3 Input parameters for the effect/association between individual-level exposures and individual-level outcomes ('Evidence level 1' parameters)

Exposure variable	Dependent variable	Point Estimate	Model covariates	Study	Notes
Exclusive breast-feeding	Body mass index	MD=-0.14 (-0.26, -0.02)	Age, gender, birth weight, BMI of the mother and educational level of the mother	(140)	
Moderate-to-vigorous physical activity	Body mass index	MD=-0.43 (-0.63, -0.23)	Age, sex	(141)	
	Type 2 diabetes	RR= 0.65 (0.59, 0.71)	N/A	(142)	
	Body mass index	MD=0.08 (0.03, 0.13)	N/A	(143)	
Sugar-sweetened beverage consumption	Type 2 diabetes	RR= 1.28 (1.12; 1.46)	Adiposity, within person variation, sociodemographic variables, clinical factors (family history of diabetes or prevalent diseases), and lifestyle factors, including diet	(144)	
Fresh fruit and vegetable consumption	Body mass index	MD=-0.13	Baseline age, BMI and change in the following lifestyle variables: smoking status, physical activity, hours of sitting or watching TV, hours of sleep, fried potatoes, juice, whole grains, refined grains, fried foods, nuts, whole-fat dairy, low-fat dairy, sugar-sweetened beverages, sweets, processed meats, non-processed meats, trans fat, alcohol, and seafood	(145)	Outcome was weight in kg but was converted to BMI by dividing weight in kg by a common US adult height (1.645 meter). Exposures were fruits and vegetables separately but was combined to obtain one exposure (fruit and vegetable consumption/ day)
	Type 2 diabetes	RR=0.96 (0.91, 1.01)	smoking, alcohol, total energy intake, BMI, physical activity, FHDM, education and other dietary factors	(146)	

SE: Standard error; MD: Mean difference; OR: Odds ratio; HR: Hazard ratio; RR: risk ratio

eTable 3.3 Input parameters for the effect/association between individual-level exposures and individual-level outcomes ('Evidence level 1' parameters) (continued)

Exposure variable	Dependent variable	Point Estimate	Model covariates	Study	Notes
Fast-food consumption	Body mass index	MD=0.66	age, sex, education, site, baseline weight height, alcohol, TV, physical activity	(147)	Outcome was weight in kg in Blacks and Whites separately but was converted to a common BMI by dividing weight in kg by a common US adult height (1.645 meter)
	Type 2 diabetes	HR/RR=1.51(1.25, 1.83)		(148)	Exposure was consumption of processed red meat
Body mass index in childhood	Type 2 diabetes	OR=1.24		(149)	The effect was expressed in in terms of odds ratio per standard deviation BMI but authors stated that the reported "[odds ratio] was approximately equivalent to a 24% increase in odds of diabetes per kg/m2 in BMI"
Body mass index	Moderate-to-vigorous physical activity	OR=0.96 (0.94, 0.98)	Smoking habits, sex, sedentary lifestyle at age 41, and changes in BMI from ages 41 to 44 and 44 to 46	(124)	Outcome was sedentary lifestyle so we took the inverse to express the effect of BMI on physical activity The OR presented is an annualized OR

SE: Standard error; MD: Mean difference; OR: Odds ratio; HR: Hazard ratio; RR: risk ratio

eTable 3.4 Input parameters for the effect/association between neighborhood-level exposures and individual-level exposures ('Evidence level 2' parameters)

Predictors	Dependents	Point estimates	Model covariates	Study	Notes
Neighborhood supermarket (per square mile)	Fruits and vegetables consumption	RR=1.33(1.05, 1.69)	age, race, sex, per capita annual income	(150)	Actual outcome: Alternative Healthy Eating Index
Neighborhood Fast-food density (#outlets/mile)	Fast-food consumption	OR=1.11(0.98, 1.26)	Age, education, per capita HH income, race, sex, site	(151)	Outcome is fast-food ≥ 1 times/week within 1 mile vs. never
Neighborhood walkability	Physical activity	OR=1.74(1.51, 2.01)	age, gender, education, BMI, days in the U.S., and habitual physical activity level in Cuba	(121)	Outcome is whether engaged in purposive walking last week Original walk score exposure has been dichotomized (i.e. walk score >=70) and odds ratio for engaging in purposeful walking re-adjusted
Access to Parks	Physical activity	OR=1.50(1.06, 2.13)	Age, gender, education, children <18 in home, SES	(152)	Outcome: ≥ 6 walking sessions/week totaling >180 minutes. Exposure: Very good access to public open spaces (i.e. = top quartile of access) vs. very poor access to public open spaces; Access to public open spaces is defined on the basis of distance, attractiveness and size

CI: Confidence interval; SE: Standard error; MD: Mean difference; OR: Odds ratio; HR: Hazard ratio; RR: risk ratio

eTable 3.5 Input parameters for the effect/association between neighborhood-level demographics and neighborhood-level exposures ('Evidence level 3' parameters)

Dependents	Model predictors and standard errors	Source	Notes
High neighborhood walkability	Intercept: log-odds(0.0171*) Percent Non-White: OR=20 Percent below FPL: OR=6.70 Percent bachelor graduates: OR=41.21	ACS, 2014[110] Walkscore.com	High neighborhood walkability was defined as having a Walk score ≥ 70 (Very walkable to walker's paradise) vs. poor walkability (i.e. walk score < 70, Car-dependent to somewhat walkable)
Park Access	Intercept: log-odds (0.5055*) Predominantly Non-White: OR=1.85 Predominantly below FPL: OR=1.32	Wolch et al. [122]	We used the contingency tables in the article to construct estimate Access to parks was defined as the percent of population living within a quarter-mile buffer
Fast-food density	Intercept: 0 Percent Non-White: MD=0.99 Percent below FPL: MD=5.86 Percent bachelor graduates: MD:1.40 Standard error: 3.49	ACS, 2014 [110] NETS[123]	
Supermarket density	Intercept: -0.40 Percent Non-White: MD=0.51 Percent below FPL: MD=3.74 Percent bachelor graduates: MD=1.12 Standard error: MD=2.38	ACS, 2014[110] NETS [123]	

** = calibrated intercept; FPL: Federal poverty level; OR: Odds ratio; MD: Mean difference; Predominantly White was defined as having percent non-White >=75%; predominantly poor was defined as having a percent below federal poverty level>=40% as done in* Wolch et al. [122]

eTable 3.6 Input parameters for the effect/association between individual-level covariates and individual-level exposures, NHANES 1999-2014 ('Evidence level 3' parameters)

			Predictors												
			Inter-cept*	Lagged * (OR)	Age (OR)	Male (OR)	Non-White (OR)	Low-Income (OR)	Low-Income (OR)	Married (OR)	BMI (OR)	EnvPRK (OR)	EnvWLK (OR)	EnvSMD (OR)	EnvFFD (OR)
Outcomes	Birth	EBF	0.231	·	·	0.98	1.00	0.9	0.9	·	·	·	·	·	·
	Early child-hood	MVPA	0.260	·	1.05	1.49	0.88	1.11	1.11	·	·	1.00	1.00	·	·
		FFD	0.646	·	1.05	1.08	1.00	1.10	1.10	·	·	·	·	·	1.00
		FFV	0.586	·	0.98	0.93	1.50	1.24	1.24	·	·	·	·	1.00	·
		SSB	0.296	·	1.34	1.21	0.90	2.26	2.26	·	·	·	·	·	·
	Middle child-hood	MVPA	0.254	0.869	0.97	1.25	1.01	1.09	1.09	·	0.96 &	1.00	1.00	·	·
		FFD	0.639	2.203	1.02	0.98	1.10	0.99	0.99	·	·	·	·	·	1.00
		FFV	0.600	0.198	1.02	1.18	1.39	1.37	1.37	·	·	·	·	1.00	·
		SSB	0.318	9.679	1.15	1.46	1.02	1.40	1.40	·	·	·	·	·	·
	Adolescence	MVPA	0.221	0.069	0.98	1.33	0.93	0.90	0.90	·	0.96 &	1.5	1.74	·	·
		FFD	0.637	4.759	1.02	0.95	1.10	1.07	1.07	·	·	·	·	·	1.00
		FFV	0.600	0.198	1.02	1.18	1.29	1.34	1.34	·	·	·	·	1.00	·
		SSB	0.358	8.004	1.15	1.52	1.02	0.37	0.37	·	·	·	·	·	·

** The intercept and Lagged variable regression coefficients have been obtained from our calibration algorithm to match the observed means and prevalence. NHANES: National health and nutrition examination survey 1999-2014; OR: Odds ratio; EBF: Exclusive breastfeeding (i.e. Exclusively breastfed ≥ 6months) ; FFD: Fast-food consumption (i.e. Ate fast-food ≥ 1 times in past week) ; MVPA: Moderate-to-vigorous physical activity (i.e.Engage in moderate-to-vigorous physical activity); SSB: Sugar-sweetened beverage consumption (i.e. Drank ≥ 1 glasses of soda or sugary drinks); FFV: Fresh fruit and vegetable consumption; SMK: Smoking (i.e. current smoking); ALC: Alcohol consumption (i.e. Binge drank alcohol the past month); EnvWLK: Environment or neighborhood walkability; EnvPRK: Environment or neighborhood park Access; EnvSMD: Environment or neighborhood supermarket density; EnvFFD: Environment or neighborhood fast-food density;*

&These odds ratios were taken from the literature ('evidence level 1') whereas the others are computed from NHANES 1999-2014[130].

eTable 3.6 Input parameters for the effect/association between individual-level covariates and individual-level exposures, NHANES 1999-2014 ('Evidence level 3' parameters) (continued)

		Predictors												
		Inter-cept*	Lagged* (OR)	Age (OR)	Male (OR)	Non-White (OR)	Low-Income (OR)	Low-Income (OR)	Married (OR)	BMI (OR)	EnvPRK (OR)	EnvWLK (OR)	EnvSMD (OR)	EnvFFD (OR)
Young adult-hood	MVPA	0.174	19.688	0.98	1.17	0.69	0.57	0.57	1.06	0.96[&]	1.50	1.74	·	·
	FFD	0.659	1.448	0.98	1.32	0.89	1.15	1.15	0.87	·	·	·	·	1.11
	FFV	0.604	0.079	0.99	1.73	1.38	1.39	1.39	1.21	·	·	·	1.33	·
	SSB	0.395	1.020	0.97	2.42	1.44	1.68	1.68	0.86	·	·	·	·	·
	ALC	0.220	0.80	0.97	1.94	0.87	1.56	1.56	0.74	·	·	·	·	·
	SMK	0.220	1.04	0.97	1.60	0.47	1.94	1.94	0.54	·	·	·	·	·
Adult-hood	MVPA	0.130	19.298	0.98	1.17	0.57	0.58	0.58	1.21	0.96[&]	1.50	1.74	·	·
	FFD	0.651	0.869	0.97	1.25	0.87	1.14	1.14	0.83	·	·	·	·	1.11
	FFV	0.570	0.098	0.99	1.72	1.41	1.36	1.36	1.18	·	·	·	1.33	·
	SSB	0.371	0.427	0.96	2.38	1.48	1.62	1.62	0.80	·	·	·	·	·
	ALC	0.070	21	0.96	2.33	21	2.10	2.10	0.62	·	·	·	·	·
	SMK	0.060	21	0.98	1.67	21	2.10	2.10	0.44	·	·	·	·	·

** The intercept and Lagged variable regression coefficients have been obtained from our calibration algorithm to match the observed means and prevalence. NHANES: National health and nutrition examination survey 1999-2014; OR: Odds ratio; EBF: Exclusive breastfeeding (i.e. Exclusively breastfed ≥ 6months) ; FFD: Fast-food consumption (i.e. Ate fast-food ≥ 1 times in past week) ; MVPA: Moderate-to-vigorous physical activity (i.e.Engage in moderate-to-vigorous physical activity); SSB: Sugar-sweetened beverage consumption (i.e. Drank ≥ 1 glasses of soda or sugary drinks); FFV: Fresh fruit and vegetable consumption; SMK: Smoking (i.e. current smoking); ALC: Alcohol consumption (i.e. Binge drank alcohol the past month); EnvWLK: Environment or neighborhood walkability; EnvPRK: Environment or neighborhood park Access; EnvSMD: Environment or neighborhood supermarket density; EnvFFD: Environment or neighborhood fast-food density;*

[&]These odds ratios were taken from the literature ('evidence level 1') whereas the others are computed from NHANES 1999-2014[(130)].

eTable 3.7 Input parameters for the effect/association between individual-level covariates and individual-level outcome, NHANES 1999-2014, ('Evidence level' 3 parameters)

		Birth	Early Childhood	Middle Childhood	Adolescence	Young Adulthood		Adulthood	
		Outcomes							
		BMI (MD)	BMI (MD)	BMI (MD)	BMI (MD)	BMI (MD)	T2DM (OR)	BMI (MD)	T2DM (OR)
Predictors	Intercept	15.74	16.24	17.067	18.89	20.559	0.00002*	22.68	0.00032*
	Lagged	·	0.006	-0.070	0.35	0.19	·	0.28	·
	BMI_Ado	·	·	·	·	·	1.24[&]	·	1.24[&]
	Age	0.4947	-0.10	0.86	0.56	0.18	1.12	0.02	1.07
	Male	0.4389	0.20	-0.19	-0.60	-0.68	0.99	-1.05	1.45
	Non-White	0.15	0.15	0.72	0.90	0.88	1.74	0.36	2.14
	Low-income	0.11	0.11	0.32	0.37	0.63	1.55	0.13	1.59
	Married	·	·	·	·	-0.06	1.21	-0.61	1.14
	BMI	·	·	·	·	·	1.11	·	1.11
	Exclusively breastfed ≥ 6months	·	-0.14[&]	·	·	·	·	·	·
	Engage in moderate-to-vigorous physical activity	·	-0.43[&]	-0.43[&]	-0.43[&]	-0.43[&]	0.65[&]	-0.43[&]	0.65[&]
	Ate fast-food ≥ 1 times in past week	·	0.66[&]	0.66[&]	0.66[&]	0.66[&]	1.51[&]	0.66[&]	1.51[&]
	Eat ≥ 5 fresh fruits and vegetables/day	·	-0.13[&]	-0.13[&]	-0.13[&]	-0.13[&]	0.96[&]	-0.13[&]	0.96[&]
	Drank ≥ 1 glasses of soda or sugary drinks	·	0.08[&]	0.08[&]	0.08[&]	0.08[&]	1.28[&]	0.08[&]	1.28[&]
	Current smoker	·	·	·	·		1.25	-2.15	1.13
	Binge drank alcohol the past month	·	·	·	·		1.50	0.62	1.26
	Has family history of type 2 diabetes	·	·	·	·	·	4.07	·	3.57
	Standard deviation	1.49	1.994	4.657	5.733	6.9	·	6.9	·
	Minimum	10.76	12.58	12.40	13.30	15.5	·	8.9	·
	Maximum	23.56	33.20	46.100	50.70	62.9	·	72.9	·

**calibrated intercept; OR: Odds ratio; MD: Mean difference*

[&]These parameters were taken from the literature ('evidence level 1') whereas the others were computed from NHANES 1999-2014.

Chapter 4. Modeling the role of childhood adiposity in the development of adult Type 2 diabetes in a 64-year follow-up study in Los Angeles: An agent-based simulation study

4.1. Abstract

Background: Compared to normal-weight children, obese children are at higher risk of becoming obese adults putting them at increased risk of developing type 2 diabetes in later adulthood. The contribution of childhood obesity to adult type 2 diabetes risk not due to adulthood adiposity and more generally, the causal pathways through which childhood obesity increases adult type 2 diabetes risk are not well understood. This study aimed to investigate the overall contribution of childhood obesity to incident diabetes levels not due to adult adiposity overall and in different racial groups.

Objective: To investigate the overall contribution of childhood obesity to incident diabetes levels not due to adult adiposity overall and in different racial/ethnic groups.

Methods: We used data from 98,230 simulated individuals from the Virtual Los Angeles Cohort study aged 6 to 49 years. Specifically, we applied the g-computation algorithm to causal mediation analysis to investigate the role of childhood obesity in the development of adult type 2 diabetes.

Results: The marginal adjusted odds ratio (aOR) for the total effect of childhood obesity on adult type 2 diabetes was 1.37 (95%CI 1.32–1.46). Much of the effect of childhood obesity on adult type 2 diabetes was mostly attributable to pathways other than through adult obesity; the aOR for the pure direct effect was 1.36 (1.31–1.41)). In all racial/ethnic subpopulations, a small percentage of the total effect of childhood obesity on adult type 2 diabetes was attributable to childhood obesity affecting adult obesity and subsequently affecting adult type 2 diabetes.

Conclusion: Across all racial/ethnic groups, childhood obesity remains a risk factor for adult type 2 diabetes independent of its effects on adult obesity. This finding reiterates the need to consider early prevention of childhood obesity as a means of primary prevention of type 2 diabetes. As demonstrated in this study, agent-based simulation models should be used as virtual laboratories for synthesizing best existing evidence and for exploring new mechanisms in obesity research.

Keywords: agent-based model, simulation, mediation, g-formula, obesity, diabetes, Los Angeles

4.2. Introduction

For decades, obesity has been recognized as a major public health problem affecting millions of Americans including the most vulnerable segment of the population, namely children, adolescents and lower-income minorities. [2] This situation is no different for Los Angeles County which has seen its rates of obesity almost double over the last two decades with non-Hispanic Whites and African-American bearing the heaviest toll.[153] In 2011, one in four adults and one in four children were considered obese in Los Angeles County.[57] This is especially alarming as childhood obesity has been shown to be a risk factor for later adult chronic conditions.[7,15] In fact, compared to normal-weight children, obese children are at higher risk of becoming obese adults [154] and adult obesity increases the risk of developing type 2 diabetes in later adulthood.[7] These findings suggest that childhood obesity may be a risk factor for adult type 2 diabetes through adult adiposity and that the increased risk in type 2 diabetes due to obesity may be due to the tracking of excess weight from childhood into adulthood.

What is less known today is whether childhood obesity affects adult adiposity independently from adulthood adiposity.[50] More generally, the causal pathways through which childhood obesity increases adult type 2 diabetes risk are not well understood. The ability to open such a black box can assist policymakers such as the Los Angeles County Department of Public Health in identifying causal pathways that if interrupted would yield the greatest decrease in type 2 diabetes at the lowest cost. Investigations of this sort entail mediation and interaction analyses. In the past, methods such as the so-called "difference method" have been used to estimate mediated (or indirect) effects in the exploration of such mechanisms but they have been unsatisfactory as they can lead to distorted results.[155,156] More recently though, novel methods

based on the g-methods of Robins[87] have allowed researchers to disentangle the path-specific effects of exposures or interventions.[157–159]

Given that there are no long-running experimental trials (and hence, no cohorts that have been followed for any significant period of time) in Los Angeles County, we previously created a virtual cohort calibrated to the population of Los Angeles to study obesity and type 2 diabetes in Los Angeles. The specific objective of the study was to investigate the contribution of childhood obesity to incident diabetes that is independent of its effect on adult adiposity, and determine if race/ethnicity modifies this contribution.

4.3. Methods

Study population and data sources

We used data from the ViLA-Obesity model, a stochastic dynamic discrete-time agent-based model for the study of obesity and type 2 diabetes and calibrated to the population of Los Angeles County. The model simulated 98,230 agents spread out in 235 simulated neighborhoods from birth to middle adulthood. Each simulated individual was born in 2009 in a specific neighborhood of Los Angeles County and could exhibit healthy and unhealthy behaviors (e.g. physical activity, fast-food consumption…).At each time step, the model updated the individuals' behaviors, changed their body mass indexes and generated a probability of developing type 2 diabetes as a function of the agent's current state.

Measures and variables

- *Exposure: Childhood obesity between age 6 to 12*

The exposure of interest was childhood obesity in middle childhood between the age of 6 and 12. Childhood obesity was defined using the WHO guidelines on the basis of the body mass index (BMI) Z-scores calculated using CDC-provided SAS codes. [115] We used Z-scores instead of percentiles since Z-scores are comparable across ages and sex and are better for longitudinal assessment.[114] A child with a BMI Z-score (BMIz) less than -2 was classified as underweight; a BMIz greater or equal to -2 but less than 1 was classified as normal-weight; a BMIz greater or equal to 1 but less than + 2 was classified as overweight and a BMIz greater or equal to 2 was classified as obese.[116].

- *Mediators: Adult obesity between age 30 to 39 and physical activity between age 25 and 39*

The primary mediator of interest was adult obesity between the age of 30 and 39 (binary variable). Using the WHO guidelines, an individual with a BMI less than 18.5 was classified as underweight; a BMI greater or equal to 18.5 but less than 25 was classified as normal-weight; a BMI greater or equal to 25 but less than 30 was classified as overweight and a BMI greater or equal to 30 was classified as obese. [117]

The secondary mediator of interest was the adult physical activity level between age 25 and 39 (binary variable).

- *Outcome: Adult type 2 diabetes between age 40 to 49*

The outcome of interest was the development of incident adult type 2 diabetes between the age of 40 and 49 (binary variable).

- *Covariates and intermediate health behaviors*

The following variables were considered in this study: individuals' socio-demographics (age, sex, socio-economic status, and marital status, race), individuals' behaviors (sugar-sweetened beverage consumption, physical activity, fruit and vegetable consumption, fast-food consumption, smoking, alcohol consumption), neighborhood walkability and neighborhood access to parks. All these variables were binary with the exception of age which was continuous.

Statistical analyses

- *Causal Graph*

In this study, we developed a directed acyclic diagram [72] to represent our assumptions about the underlying pathways from childhood obesity to adult type 2 diabetes in our simulation model. The relationships between covariates, exposure, mediators and outcomes were depicted in the causal diagram (see Figure 4.1).

- *G-computation algorithm*

We used the g-computation algorithm of Robins (applied to the parametric g-formula) to decompose the effect of childhood obesity on adult type 2 diabetes. G-computation is a generalization of the standardization method for time-varying exposures and confounders.[87] It is in general more flexible than the other g-methods (inverse probability of treatment weighting, g-estimation) and is particularly appealing in the context of complex data structure where confounding variables are affected by prior exposures.[87,160] However, it requires correct model specification when modeling all covariates and may be more sensitive to violations of assumptions.[160]

- *G-computation assumptions*

To conduct our causal mediation analysis, it was assumed that there was conditional exchangeability (i.e. no uncontrolled confounding assumption), positivity[94], consistency,[95] no interference (i.e. stable unit treatment value assumption or SUTVA),[161] and no other sources of bias (i.e. no selection bias, no measurement error and no model misspecification). The assumption of consistency means that for every individual whose exposure status is $A=a$, his potential outcome Y_a under the intervention $doA=a$ is equal to his observed outcome [i.e. $P(Y_a|A=a) = P(Y|A=a)$]. The assumption of positivity means that for every level of a covariate L, the

probability of observing the exposure *A* given the covariate *L* is not zero, i.e. there is enough variability and there are no zero cells when one stratifies the exposure by each level of the covariate *L* (i.e. for every level of *L*, $P(A=a|L=l) > 0$). The SUTVA assumption is sometimes referred to as the independence or the no-interference assumption, which means that the potential outcome Y_{ia} of individual *i* should not depend on the mechanism by which the individual *i* receives treatment *A* nor depend on whether individual *j* receives treatment *A* or *B*. In the context of mediation analysis, the no-uncontrolled confounding assumption consists of four parts[162,163]: (i) no uncontrolled confounding between exposure and outcome, (ii) no uncontrolled confounding between mediator and outcome, (iii) no uncontrolled confounding between exposure and mediator and lastly (iv) no exposure-induced mediator-outcome confounder. In our study, the latter assumption (iv) is violated since childhood adiposity was allowed to affect subsequent physical activity levels which in turn can affect subsequent obesity risk. Fortunately, recent work has proposed solutions to circumvent this problem.[163] We briefly described the two estimation approaches used to decompose the effect of interest.

❖ *G-computation estimation and effect decomposition*

In order to circumvent the problem of exposure-induced mediator-outcome confounder (fourth assumption), we applied two analytical approaches described in [163] to readily compute natural effects and other types of effects. Let OBE_2, $MVPA_5$, $MVPA_6$, OBE_6, $T2DM_7$ represent childhood adiposity at age 6-12, adult level of physical activity at age 25-29, adult level of physical activity at age 30-39, adult obesity at age 30-39 and adult type 2 diabetes at age 40-49, respectively. The subscripts represent time indices (see **Table 4.1** for details). Let *Y, A, L, V, M* and *C* denote random variables that respectively take the value *y, a, l, v, m* and *c*. *Y* is the

outcome, *A* the exposure, *M* the mediator of interest, *L* a mediator that is an exposure-induced confounder of the relationship between *M* and *Y* and *V* a direct child (descendant) of *L* and *A*. For any variable *W*, W_a (or similarly $W_{A=a}$, W^a or $W^{A=a}$) is the potential outcome of *W* had *A* been set to *a*. For instance, $Y_{A=a}$, $L^{A=a}$, $M^{A=a}$, $V_{A=a}$ is the potential outcome of *Y*, *L*, *M or V* had *A* been set a. We will sometimes use them in combination such as $Y_{A=a,\, L^{A=a},\, V^{A=a,L}{}_{A=a,\, M^{A=a,\, V}{}_{A=a}}}$ to mean the potential outcome value of *Y* had *A* been set to a, *L* to $L^{A=a}$, *V* to $V^{A=a,L}{}_{A=a}$ and *M* to $M^{A=a,\, V}{}_{A=a}$.

In the first approach "joint mediator approach", we considered the set $M = \{MVPA_5, MVPA_6, OBE_6\}$ jointly (i.e. simultaneously) as the mediator of interest. In other words, from childhood obesity to adult diabetes there were essentially two pathways: (i) one direct and (ii) one indirect that combines pathways through adult obesity (OBE_6) and pathways through adult physical activity ($MVPA_5$, $MVPA_6$).

In the second approach "path-specific approach" we considered adult obesity (OBE_6) as the actual mediator of interest. Put another way, from childhood obesity to adult diabetes there were essentially three pathways: (i) pathways involving neither adult obesity nor adult level of physical activity (i.e. $OBE_2 \rightarrow T2DM_7$) (ii) effects not involving adult level of physical activity (i.e. $OBE_2 \rightarrow OBE_6 \rightarrow T2DM_7$) and (iii) effects involving only adult level of physical activity (i.e. Combination of $OBE_2 \rightarrow MVPA_5 \rightarrow MVPA_6 \rightarrow T2DM_7$, $OBE_2 \rightarrow MVPA_5 \rightarrow OBE_6 \rightarrow T2DM_7$ *and* $OBE_2 \rightarrow MVPA_6 \rightarrow T2DM_7$) summarized as $OBE_2 \rightarrow MVPA_{adu} \rightarrow T2DM_7$. In the second approach, note that we did not estimate actual natural effects but rather different path-specific effects.

Let us now define all the quantities we estimated in this study. More extensive definitions and expressions can be found in Wang et Arah [(164)] and Vanderweele et al[(163)].

✓ *The expressions for the natural decomposition are given as follow:*

The total effect (TE) measured the overall extent to which childhood obesity causes adult type 2 diabetes. It was given by the following expression:

E_{TE}= E [$T2DM7_{OBE2=1}$ – $T2DM7_{OBE2=0}$]

The pure direct effect (PDE) measured the extent to which childhood obesity causes adult type 2 diabetes through pathways other than through the joint mediator set $M = \{MVPA_5, MVPA_6, OBE_6\}$ and was given by the following expression:

E_{PDE} = E [$T2DM7_{OBE2=1,\,M}^{OBE2=0}$ – $T2DM7_{OBE2=0,\,M}^{OBE2=0}$]

The total direct effect (TDE) measured the extent to which childhood obesity causes adult type 2 diabetes through pathways other than through the joint mediator set $M = \{MVPA_5, MVPA_6, OBE_6\}$ allowing the joint mediator set to simultaneously boost up or tune down such effect at the same time. It was given by the following expression:

E_{TDE} = E [$T2DM7_{OBE2=1,\,M}^{OBE2=1}$ – $T2DM7_{OBE2=0,\,M}^{OBE2=1}$]

The pure indirect effect (PIE) measured the extent to which childhood obesity causes adult type 2 diabetes through the joint mediator set $\{MVPA_5, MVPA_6, OBE_6\}$ only, not accounting for the possible interaction between childhood obesity and the joint mediator set $M = \{MVPA_5, MVPA_6, OBE_6\}$. It was given by the following expression:

E_{PIE} = E [$T2DM7_{OBE2=0,\,M}^{OBE2=1}$ – $T2DM7_{OBE2=0,\,M}^{OBE2=0}$]

The total indirect effect (TIE) measured the extent to which childhood obesity causes adult type 2 diabetes through the joint mediator set $\{MVPA_5, MVPA_6, OBE_6\}$ only, but accounting for the possible interaction between childhood obesity and the joint mediator set $M = \{MVPA_5, MVPA_6, OBE_6\}$. It was given by the following expression:

E_{TIE} = E [$T2DM7_{OBE2=1,\,M}^{OBE2=1}$ – $T2DM7_{OBE2=1,\,M}^{OBE2=0}$]

The controlled direct effect (CDE) measured the extent to which childhood obesity causes adult type 2 diabetes when fixing the joint mediator set at specific value for everyone in the population. There are three types of CDEs: (i) the CDE_{ref} (CDE at the reference level) or CDE when fixing the joint mediator set to the reference level of 0; (ii) the CDE_{idx} (CDE at the index level) or CDE when fixing the joint mediator set to the index level of 1 and (iii) the CDE_{sto} (stochastic CDE) or CDE when allowing the joint mediator set to attain a certain controlled distribution in the population. These quantities were given by the following expressions:

$$E_{CDEref} = E\,[T2DM7_{OBE2=0,\,M=0} - T2DM7_{OBE2=1,\,M=0}]$$

$$E_{CDEidx} = E\,[T2DM7_{OBE2=0,\,M=1} - T2DM7_{OBE2=1,\,M=1}]$$

$$E_{CDEsto} = E\,[T2DM7_{OBE2=0,\,M=m} - T2DM7_{OBE2=1,\,M=m}]$$

- ✓ *The expressions for the path-specific effects were also given as follow:*

The effect involving neither adult obesity nor adult level of physical activity ($OBE_2 \rightarrow T2DM_7$) was expressed as follows:

$$E_{OBE2\rightarrow T2DM7} = E\,[T2DM7_{OBE2=1,\,MVPA5^{OBE2=0},\,MVPA6^{OBE2=0,\,MVPA5}_{OBE2=0,\,OBE6^{OBE2=0}}} - T2DM7_{OBE2=0,\,MVPA5^{OBE2=0},\,MVPA6^{OBE2=0,\,MVPA5}_{OBE2=0,\,OBE6^{OBE2=0}}}]$$

The effect not involving adult level of physical activity ($OBE_2 \rightarrow OBE_6 \rightarrow T2DM_7$) was expressed as follows:

$$E_{OBE2\rightarrow OBE6\rightarrow T2DM7} = E\,[T2DM7_{OBE2=1,\,MVPA5^{OBE2=0},\,MVPA6^{OBE2=0,\,MVPA5}_{OBE2=0,\,OBE6^{OBE2=1,\,MVPA5^{OBE2=0}}}} - T2DM7_{OBE2=1,\,MVPA5^{OBE2=0},\,MVPA6^{OBE2=0,\,MVPA5}_{OBE2=0,\,OBE6^{OBE2=0,\,MVPA5^{OBE2=0}}}}]$$

The effect involving only adult level of physical activity ($OBE_2 \rightarrow MVPA_{adu} \rightarrow T2DM_7$) was expressed as follows:

$$E_{OBE2 \rightarrow MVPA \rightarrow T2DM7} = E\left[T2DM7_{OBE2=1,\ MVPA5}^{\ OBE2=1}{}_{,\ MVPA6}^{\ OBE2=1,\ MVPA5}{}_{OBE2=1,\ OBE6}^{\ OBE2=1,\ MVPA5}{}_{OBE2=1} - T2DM7_{OBE2=1,\ MVPA5}^{\ OBE2=0}{}_{,\ MVPA6}^{\ OBE2=0,\ MVPA5}{}_{OBE2=0,\ OBE6}^{\ OBE2=0,\ MVPA5}{}_{OBE2=0}\right]$$

We completed all data preparation, parametric modelling, simulations and analysis in SAS version 9.4 (SAS Institute, Inc., Cary, NC)

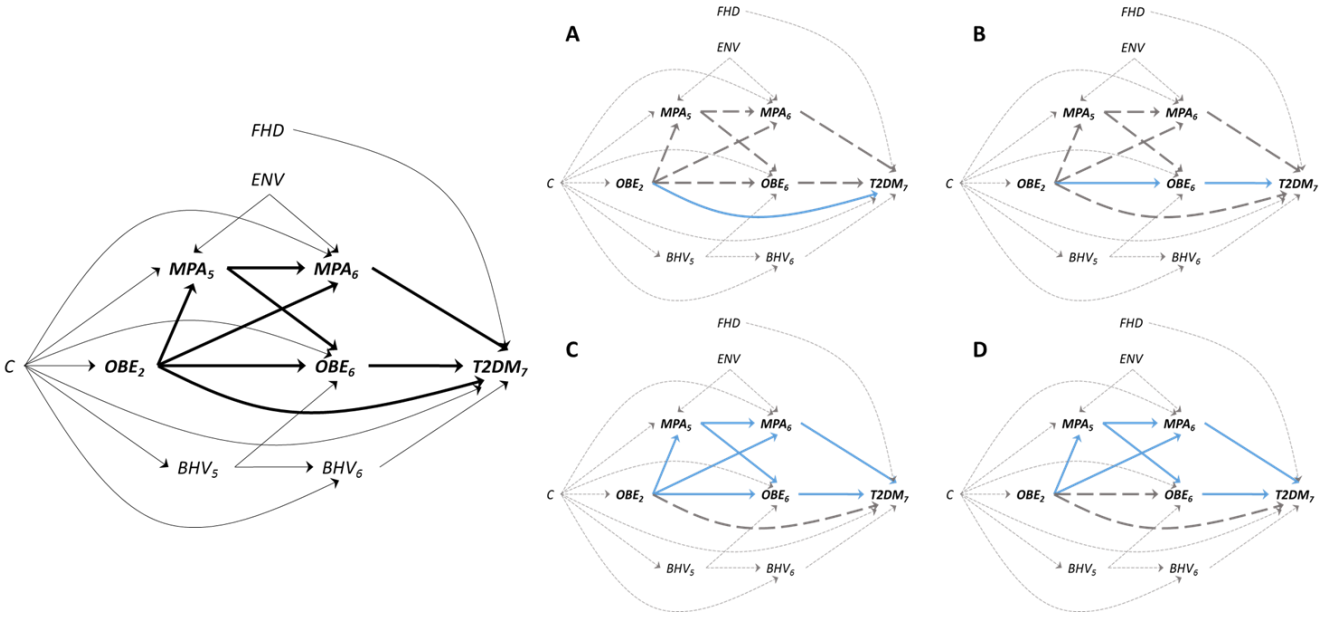


Figure 4.1 Simplified directed acyclic graph (DAG) of the assumptions about the data-generating processes between childhood obesity and type 2 diabetes 2: Middle childhood (6-12); 5: Young adulthood (25-29); 6: Young adulthood (30-39); 7: Middle adulthood (40-49); C: socio-demographics (age, sex, socio-economic status, marital status); BHV: time-varying behaviors (sugar-sweetened beverage consumption, fast-food consumption, fresh fruit and vegetable consumption, , smoking, alcohol drinking); OBE: Obesity; T2DM: Type 2 diabetes; ENV: (Neighborhood Access to Parks, Neighborhood walkability); FHD: Family history of type 2 diabetes. The bold lines depict the pathways from childhood obesity to adult type 2 diabetes. DAG A: Direct natural effect (PDE, TDE), B: Effect $OBE_2 \rightarrow OBE_6 \rightarrow T2DM_7$, C: Indirect natural effect (TIE, PIE), D: Effect $OBE_2 \rightarrow MVPA_{adu} \rightarrow T2DM_7$

4.4. Results

Table 4.2 describes the baseline and follow-up characteristics of the simulated cohort. Two thirds of our population was non-White and about one fourth had an income below or at the federal poverty level. Consumption of fast-food was found in 75% of children and in about 50% of adults in their 40s. About one fourth and two thirds of individuals were obese in childhood and adulthood respectively. One in four individuals had type 2 diabetes in their 40s.

Table 4.3 presents the decomposition of the effects of childhood obesity on adult type 2 diabetes estimated using g-computation. The marginal adjusted odds ratio (aOR) for the total effect of childhood obesity on adult type 2 diabetes was 1.37 (95%CI 1.32–1.46). The results were similar using either method. Under both approaches, much of the effect of childhood obesity on adult type 2 diabetes was mostly attributable to pathways other than through adult obesity (e.g. pure direct effect aOR: 1.36 (1.31–1.41)). Only 3% of the total effect of childhood obesity on adult type 2 diabetes was attributable to childhood obesity affecting adult obesity and subsequently affecting adult diabetes. (**Figure 4.2**)

Table 4.1 Life period, time-points and age groups in ViLA

Time	Age group	Life period
0	0-1	Birth
1	2-5	Early Childhood
2	6-12	Middle Childhood
3	13-17	Adolescence
4	18-24	Young Adulthood
5	25-29	Young Adulthood
6	30-39	Young Adulthood
7	40-49	Middle Adulthood
8	50-59	Middle Adulthood
9	60-65	Middle Adulthood

Table 4.2 Baseline and follow-up characteristics of simulated individuals in the ViLA-Obesity model (n=98,230)

	Childhood (6-12)	**Adulthood (30-39)**	**Adulthood (40-49)**
Age in years (Mean, SD)	9 (1.78)	34.52 (2.63)	44.48 (2.63)
Male (%)	49%	49%	49%
Low-income (i.e. below or at FPL) (%)	22%	22%	22%
Married (%)	0%	44%	44%
Non-White (%)	63%	63%	63%
Has family history of type 2 diabetes (%)	8%	8%	8%
Ate fast-food ≥ 1 times in past week (%)	76%	74%	52%
Physically active at least one hour per day (%)	23%	27%	24%
Drank ≥ 1 glasses of SSB (%)	66%	45%	23%
Eat ≥ 5 fresh fruits and vegetables (%)	45%	54%	51%
Current smoker (%)	0%	12%	9%
Binge drank alcohol the past month (%)	0%	17%	13%
High neighborhood walkability (%)	28%	27%	27%
High neighborhood access to Parks (%)	54%	56%	55%
Body mass index (kg/m^2 Mean, SD)	(20.68) 4.39	27.24 (6.26)	26.24 (6.97)
Obese (%)	24%	32%	30%
Has type 2 diabetes (%)	0%	3%	25%

All categorical variables are binary. FPL: Federal Poverty Level, SD: Standard deviation, SSB: sugar-sweetened beverage consumption

Table 4.3 Decomposition of the effect of childhood obesity on adult type 2 diabetes in the ViLA-Obesity model using g-computation in a marginal structural model

Method	OR[a] (95% CI)
Joint mediator approach	
($MVPA_5$, $MVPA_6$, OBE_6} as the joint mediator set)	
Pure direct effect (PDE)	1.36 (1.31 – 1.41)
Total indirect effect (TIE)	1.01 (1.00 – 1.02)
Pure indirect effect (PIE)	1.01 (1.00 – 1.02)
Total direct effect (TDE)	1.36 (1.31 – 1.41)
Stochastic (marginal) controlled direct effect (CDE_{sto})	1.37 (1.37 – 1.38)
Controlled direct effect at reference level (CDE_{ref})	1.39 (1.33 – 1.46)
Controlled direct effect at index level (CDE_{idx})	1.38 (1.31 – 1.44)
Total Effect	1.37 (1.32 – 1.46)
Path-specific approach	
(OBE_6 as the actual mediator)	
Effect involving neither adult obesity nor PA (OBE_2→ $T2DM_7$) (PSDE)	1.36 (1.31 – 1.41)
Effect not involving PA (OBE_2→ OBE_6→ $T2DM_7$) (PSIE-A)	1.00 (0.99 – 1.01)
Effect involving only PA (OBE_2→ $MVPA_{adu}$ → $T2DM_7$) (PSIE-B)	1.01 (1.01 – 1.01)

[a]*marginal odds ratio*

PA is short for adult level of physical activity; CI: confidence interval, PSDE: Path-specific direct effect, PSIE: Path-specific indirect effect

Table 4.4 Sensivity analysis for decomposition of the effect of childhood obesity on adult type 2 diabetes in the ViLA-Obesity model

	Pure direct effect (PDE) (Odds ratios)				Total indirect effect (TIE) (Odds ratios)			
	Mediator: Adult obesity at				**Mediator: Adult obesity at**			
	25-29	**25-29**	**30-39**	**40-49**	**50-59**	**30-39**	**40-49**	**50-59**
Exposure: Childhood obesity at 2-5	1.05	1.05	1.04	1.02	1.01	1.00	1.01	1.00
Exposure: Childhood obesity at 6-12	1.43	1.43	1.36*	1.29	1.01	1.01*	1.01	1.01
Exposure: Childhood obesity at 13-17	6.39	6.39	5.26	5.39	1.02	1.01	1.00	1.00
	30-39	**30-39**	**40-49**	**50-59**	**60-65**	**40-49**	**50-59**	**60-65**
	Outcome: Adult T2DM at				**Outcome: Adult T2DM at**			

** Estimates presented in* **Table 4.3**

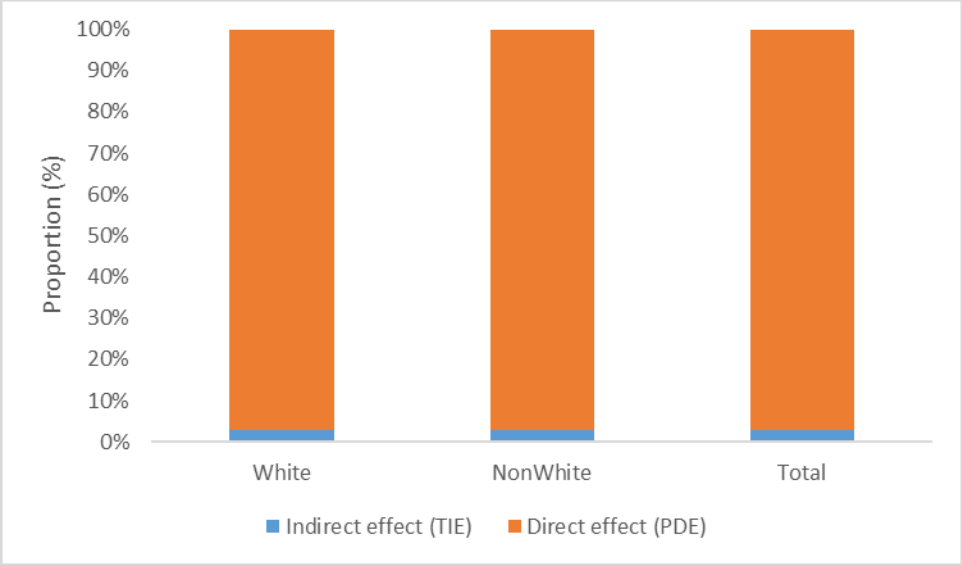


Figure 4.2 Proportion of the effect of childhood adiposity on adult type 2 diabetes that is mediated through adult adiposity by race/ethnicity in ViLA. TIE: Total indirect effect and PDE: pure direct effect

4.5. Discussion

The purpose of this study was to investigate the overall contribution of childhood obesity and racial/ethnic differences in the contribution of childhood obesity to incident adult type 2 diabetes. Using the g-computation algorithm [87] within the virtual cohort of Los Angelinos, we examined and quantified the pathways through which childhood obesity affects type 2 diabetes. Our findings suggest that much of the effect attributable to childhood obesity in the development of incident type 2 diabetes was due to pathways other than through adult obesity (so-called 'direct effect'). A corollary of this seen in the findings is that childhood obesity affects the risk of incident type 2 diabetes independently of adult adiposity. In fact, the effect of childhood obesity through adult obesity and adult level of physical activity appears to be minimal in this study. Additionally, we did not find the presence of racial/ethnic disparities in the effect of childhood obesity on type 2 diabetes.

Our findings support the conclusion that childhood obesity increases the risk of incident adult type 2 diabetes independently of adult obesity. In other words, there are other mechanisms from childhood obesity that do not involve adult adiposity that result in the development of incident adult type 2 diabetes. This has also been seen in many follow-up studies.[165,166] These studies and the present study, re-emphasize the need to start diabetes prevention during childhood—a critical period of development, in order to stop the causal chain of reaction that unravels as soon as excess weight is established in childhood. This is warranted since type 2 diabetes is a condition whose later consequences can be disabling and fatal [7,167], and that type 2 diabetes can remain undiagnosed in a third of diabetic patient.[22]

The finding of a minimal to insignificant effect of childhood obesity on adult type 2 diabetes through adult obesity was contrary to our expectations. In fact, we expected to see a

higher effect that would have been mediated through adult obesity because of the fact that excess weight can track from childhood to adulthood [99] and that adult obesity is a risk factor for type 2 diabetes [7]. However, there are many plausible explanations for this phenomenon. First, in a hypothetical chain of causation, the first cause or most distal cause will tend to have a much attenuated effect (because of the longer pathway to the outcome) than a more proximal cause which is closer to the outcome. Second, many studies have also found that adjusting for adult current BMI did not alter substantially the effect of childhood obesity on adult type 2 diabetes.[165] This suggests perhaps that there is small portion of the effect of childhood obesity on adult type 2 diabetes that would be mediated through adult obesity. Nevertheless, these findings do not dispute the fact that childhood obesity is a risk factor for adult obesity or that adult obesity is a risk factor for incident adult type 2 diabetes but rather stipulate that the effect of childhood obesity on adult type 2 diabetes mediated through adult obesity is minimal. In fact, a systematic review reported that only 31% of future adult diabetes could be attributable to childhood obesity and concluded that the "majority of adult obesity-related morbidity occurs in adults who were of healthy weight in childhood".[149] The idea is that, as far as the role of obesity in the development of incident type 2 diabetes, it is adult obesity that has a much bigger role to play than childhood obesity. This has been somewhat suggested in our sensitivity analysis where the effect of childhood obesity on type 2 diabetes not due to adult adiposity is much greater when the individual is obese in adolescence as compared to when the individual is obese in the early or middle childhood (See table **Table 4.4**).

There are several biological explanations that can explain how childhood obesity can affect incident adult type 2 diabetes. In fact, there exist potential biological subclinical mechanisms whereby childhood obesity can alter biological parameters that will ultimately cause

diabetes in the adulthood without necessarily increasing body weight per se during adulthood. Studies have shown that obese children are at higher risk for dyslipidemias, high blood pressure and impaired fasting glycaemia, all of which are risk factors for prediabetes and insulin resistance [28,168]. What is more, is that a high proportion of individuals with impaired fasting glycaemia will progress to type 2 diabetes.[169]

The findings of the present study have important implications for type 2 diabetes prevention. In Los Angeles county, the local department of public health has been leading major efforts to curb the obesity epidemic in the county.[57,138] This study will advance understanding of the mechanisms through which childhood obesity increases the risk of type 2 diabetes in adulthood; increase awareness of the need to recognize childhood obesity prevention as a primary means of reducing risk of adult type 2 diabetes; and allow the identification of the most feasible intervention that could potentially yield the greatest decrease in type 2 diabetes rates in the county.

The main limitation of this study is the use of a virtual cohort of Los Angelinos instead of a real cohort of individuals. This can be problematic especially if the virtual cohort does not reflect reality or is not able to reproduce expected results. This issue is related to that of calibration and validation of the cohort itself. Fortunately though, the ViLA-Obesity model has been validated against many sources of data representing the population of Los Angeles County. As with any model, there can still remain areas of shortcomings not yet apparent. However, we are somewhat reassured to see that our findings are mainly in line with the literature. Another limitation related to the first is that in the ViLA-Obesity model, we do not allow for new individuals to enter the cohort once it started or for current individuals to be lost-to-follow-up, die before the end of follow-up or experience competing risks that can prevent them from

experiencing diabetes in the adulthood. In essence, our model assumes that the simulated population is closed even though this may not be true in the real population. Nevertheless, many empirical studies are also analyzed under the assumption of closed population even if not explicitly stated.

Conclusion

Across all racial/ethnic groups, childhood obesity remains a risk factor of adult type 2 diabetes independent of its effects on adult obesity. This finding reiterates the need to consider early prevention of childhood obesity as a means of primary prevention of type 2 diabetes. As demonstrated in this study, agent-based simulation models should be used as virtual laboratories for synthesizing best existing evidence and for exploring new mechanisms and heterogeneity in obesity research.

Chapter 5. Evaluating the effectiveness of key health interventions on obesity and diabetes throughout the life course in the Virtual Los Angeles Cohort

5.1. Abstract

Background: For decades, obesity has been a major public health problem in the US and has been one of the most predominant players in the increase of the incidence of type 2 diabetes. There is a growing interest in which interventions or combinations of interventions are likely responsible for the recent decline in childhood obesity, can yield the greatest impact for the least effort, and in how and when to implement such interventions to reduce racial/ethnic disparities in obesity and diabetes.

Objective: The overarching goal of this study was to evaluate the effectiveness of key health interventions on obesity and diabetes throughout the life course in the virtual Los Angeles Cohort.

Methods: This study used data from a virtual Los Angeles cohort of 98,230 simulated individuals aged 2 to 65 years. We analyzed the data using the g-computation algorithm to evaluate the following interventions: (i) breastfeeding for six months or longer, (ii) reducing sugar-sweetened beverage consumption, (iii) increasing access to parks and (iv) designing a pedestrian-friendly community.

Results: The 48-year risk of type 2 diabetes under the sugar-sweetened beverage, the breastfeeding, the neighborhood walkability and the neighborhood park access interventions was 0.51 (95% CI 0.51 to 0.52), 0.54 (95%CI 0.53 to 0.54), 0.53 (95%CI 0.53 to 0.53) and 0.53 (95%CI 0.53 to 0.53) respectively. The 64-year risk of obesity under the breastfeeding, the neighborhood walkability and the neighborhood park access interventions were similar and equal

to 0.89 (95%CI 0.89 to 0.89). Combining all four intervention yielded a modest decrease in type 2 diabetes (population risk ratio (RR)=0.94 (95%CI 0.93 to 0.95)).

Conclusion: To be effective, most interventions have to be implemented in combinations with one another and virtually at every critical life stages throughout the life span. This study illustrates the usefulness of agent-based simulation models for evaluating the effectiveness of key health interventions on complex health issues such as obesity and diabetes throughout the life course.

Keywords: agent-based model, simulation, obesity, diabetes, life-course, g-formula, cohort, Los Angeles

5.2. Introduction

For decades, obesity has been a major public health problem in the US [96,170] and has been one of the most predominant players in the increase of the incidence of type 2 diabetes.[20,171] To remedy this, researchers and policymakers have made significant advances in obesity and diabetes prevention by identifying key risk factors that increase an individual's risk of becoming obese or developing diabetes in the adulthood as well as key healthful behaviors that can potentially reduce or prevent the occurrence of these conditions. [172,173] Although the rates have begun to level off among children,[174] they still remain persistently high, especially in many disadvantaged groups despite major ongoing efforts. There is a growing interest in knowing which interventions or combinations of interventions are likely responsible for the recent decline in childhood obesity, can yield the greatest impact for the least effort, and in how and when to implement such interventions to reduce racial/ethnic disparities in obesity and diabetes.

In light of this, researchers have since undertaken many incremental steps to evaluate the impact of many important health interventions. Firstly, randomized trials, as the gold standard for establishing causality, have contributed vastly to assessing the effectiveness of certain therapies or interventions in obesity prevention[175] or diabetes prevention for instance.[176] However, these randomized trials have offered limited evidence because they are not always generalizable to the population of interest [177] (due to the selective inclusion of participants in the study) and are typically costly and often cannot follow subjects for an extended period of time. Secondly, to address the shortcomings of randomized trials, researchers have also used existing observational studies for their attractive edge (i.e. longer follow-up, less restrictive

eligibility criteria and lower cost) to evaluate hypothetical interventions in given populations using complex methods such as the g-formula.[87,88,178] Unfortunately, the results of such endeavor although carefully computed are subject to the misspecification of the model and the presence of uncontrolled confounding to list a few.[88] Finally, some researchers have taken a step further by synthesizing existing knowledge into simulation models, thereby creating a virtual laboratory where the data generating mechanisms are known and where hypothetical interventions can be tested *in silico*.[108] Such models (typically referred to as "microsimulation" models) are very promising but fall short in that they tend to only focus on the individual level and do not include aspect of the built-environment, a key player in the obesity epidemic. In the same way, other simulation models such as systems dynamics models that only incorporate environment or aggregate-level entities[69] tend to overlook the specificity present at the individual level.

To fill the gap in assessing health intervention impacts on obesity and diabetes while addressing all the aforementioned shortcomings, we propose to use an agent-based simulation model. We chose to do so by focusing on Los Angeles County because it is one the most populous and most ethnically diverse counties in the US[109] and because they have been major efforts implemented in the county to curb the epidemic.[57] To achieve the same goal, other researchers have taken the lead in evaluating the impact of hypothetical and implemented interventions in California [134] and Los Angeles County[138] but used micro-simulation models and systems dynamics models, respectively. In contrast, following the example of Orr et al [179] and Day et al[81], we developed and used an agent-based model of a cohort of individuals representing Los Angeles County in order to study the evolution of obesity and diabetes and evaluate the impact of hypothetical and implemented interventions.

The purpose of this study was to evaluate the effectiveness of key health interventions on obesity and diabetes throughout the life course in the virtual Los Angeles Cohort (ViLA). Specifically, we assessed the overall impact and racial/ethnic disparities in the impact of (i) breastfeeding for six months or longer, (ii) reducing sugar-sweetened beverage consumption, (iii) increasing access to parks and recreations and (iv) designing pedestrian-friendly community on the incidence of obesity in childhood through adulthood and of diabetes in the adulthood.

5.3. **Methods**

- *Study population and sources of data*

We used data from the ViLA-Obesity model, a stochastic dynamic discrete-time agent-based model developed for the study of obesity and type 2 diabetes and calibrated to the population of Los Angeles County for the study of obesity and type 2 diabetes. The model simulated 98,230 agents spread out in 235 simulated neighborhoods from birth to middle adulthood. Each simulated individual was born in 2009 in a specific neighborhood of Los Angeles County and could exhibit healthy and unhealthy behaviors (e.g. physical activity, fast-food consumption…). Simulated agents were allowed to change neighborhoods at birth (0-1year), young adulthood (18-24 years) and during middle adulthood (40-49 years) with the same predictive probability. At each time step, the model updated the individuals' behaviors, changed their body mass indexes and generated a probability of developing type 2 diabetes as a function of the agent's current state.(See **eTable 5.1** for detail about the time-steps)

- *Measures and variables*
  - *Hypothetical Interventions*

The Los Angeles County Department of Public Health (LAC/DPH) with the Center for Disease Control and Prevention (CDC) implemented several interventions to curb the obesity epidemic. Among them, the "Community Putting Prevention to Work (CPPW) with the RENEW project (Renew Environments for Nutrition, Exercise, and Wellness) are of noteworthy importance. In brief, the RENEW project implemented from 2010 to 2012 "sought to implement policy, systems, and environmental changes to improve nutrition, increase physical activity, and reduce obesity, especially in disadvantaged communities".[58] Therefore, we proposed to evaluate the long-term effects of two individual-level dietary interventions (i.e. breastfeeding promotion, and reduction of sugar-sweetened beverages) and two environmental physical activity-related interventions (i.e. increasing access to parks and recreations and designing pedestrian friendly communities) on obesity and diabetes incidence in the ViLA cohort. The four interventions evaluated in this study were based on binary exposures. There were four primary interventions:

The breastfeeding intervention was implemented in the first year of life and consisted in altering the breastfeeding exposure status of simulated individuals to become "breastfed exclusively for at least six months" (if not already so) (eTable 5.2).

The sugar-sweetened beverage (SSB) intervention (i.e. eliminate the SSB consumption) was implemented throughout the life course at eight possible time points and consisted in altering the SSB consumption exposure status of simulated individuals to become "drink zero glass of soda or other sugary drinks" (if not already so) (eTable 5.2).

The neighborhood park access intervention (i.e. to increase physical activity opportunities by increasing access to parks) was implemented at three possible time points: birth, young adulthood and middle adulthood and consisted in altering the neighborhood park access exposure status to become "high neighborhood park access" (if not already so). Neighborhood park access was defined as the percent of population living within a quarter-mile buffer and was based on California aggregated data obtained from [122]. (eTable 5.2).

The neighborhood walkability intervention (i.e. to design pedestrian friendly communities by increasing the community walkability score) was also implemented at three possible time points: birth, young adulthood or middle adulthood and consisted in altering (if not already so) the neighborhood walkability exposure status to become "high neighborhood walkability". This variable was based on the neighborhood Walk Score®, a validated commercial walkability measurement tool that measures neighborhood walkability and pedestrian friendliness.[118–120] Briefly, it was based on the distance from a specific address to various amenities such that amenities within a 5-minute walk (.25 miles) were given maximum points whereas no points were given after a 30-minute walk. We defined a neighborhood to be walkable (i.e. high neighborhood walkability) if the Walk Score® in that neighborhood was at or above 70 because such scores were considered very walkable to walker's paradise (i.e. most to all errands could be accomplished on foot).[118] (eTable 5.2)

- *Three secondary interventions evaluated*

The physical activity intervention (i.e.to increase the level of physical activity) was implemented throughout the life course at eight possible time points and consisted in altering the

physical activity level exposure status of simulated individuals to meet the age-appropriate recommended physical level (if not already so).

The fast-food intervention (i.e. to eliminate fast-food consumption) was implemented throughout the life course at eight possible time points and consisted in altering the fast-food consumption exposure status of simulated individuals to become "do not consume fast-food" (if not already so).

The fresh fruit and vegetable intervention (i.e. to increase fresh fruit and vegetable consumption) was implemented throughout the life course at eight possible time points and consisted in altering the fruit and vegetable consumption exposure status of simulated individuals to become "Eat five or more servings of fruit and vegetable per day" (if not already so).

- *Implementation of interventions*

All interventions (both primary and secondary) interventions were evaluated singly and in combination with one another. When evaluated in combination, two sets of combined interventions were explored. The first set of combined interventions which included the four primary interventions (i.e. referred to as "combined interventions) was composed of two individual-level dietary interventions (i.e. breastfeeding promotion, and reduction of sugar-sweetened beverages) and two environmental physical activity-related interventions (i.e. increasing access to parks and recreations and designing pedestrian friendly communities). The second set of combined interventions which included two primary and three secondary interventions (i.e. referred to as "combined intervention plus") was composed of four dietary interventions (i.e. breastfeeding promotion, reduction of sugar-sweetened beverages, reduction of fast-food consumption and increase of fresh fruit and vegetable consumption) and one physical

activity intervention (i.e. increase of the level of physical activity). (See **eTable 5.2** in appendix for detail about the interventions). We implemented the interventions throughout the individual life course, during childhood, young adulthood and middle adulthood. We specifically defined and projected the cumulative incidence of obesity and type 2 diabetes under a natural course (i.e. no intervention, status quo), an optimistic scenario (i.e. idealistic scenario) and a pessimistic scenario (i.e. "worst-off" scenario).

As depicted in **Figure 5.1**, consider an individual who during the course of his/her life used to drink one or more sodas or other sugary drinks (SSB) per day at age 13-17 (i.e. SSB=1 at time t=3), at age 25-29 (i.e. SSB=1 at time t=5) and at age 40-49 (i.e. SSB=1 at time t=7) but did not drink any soda or sugary drinks at other time points at age 2-5 (i.e. SSB=0 at time t=1), at age 6-12 (i.e. SSB=0 at time t=2), at age 18-24 (i.e. SSB=0 at time t=4), at age 30-39 (i.e. SSB=0 at time t=6), at age 50-59 (i.e. SSB=0 at time=8). In the natural course or status quo simulation (i.e. no intervention), the individual SSB status remained unchanged throughout follow-up. In the optimistic scenario, individuals were simulated to be exposed to the intervention (i.e. assigned the desired level of the exposure) at all possible time points (i.e. SSB=0 at all time-points). In other words, the SSB exposure status of the individual at time=3, 5 and 7 would become "did not drink any soda or other sugary drinks per day" and would remain unchanged at the other time points at time t=1, 2, 4, 6 and 8 (since at these time-points the individual had already the desired level of the exposure). It was the opposite in the pessimistic scenario. The SSB exposure status of individuals were simulated and altered to become "drink one or more sodas or other sugary drinks" a time t=1, 2, 4, 6 and 8 and unchanged at time 3, 5 and 7 (i.e. SSB=1 at all time-points).

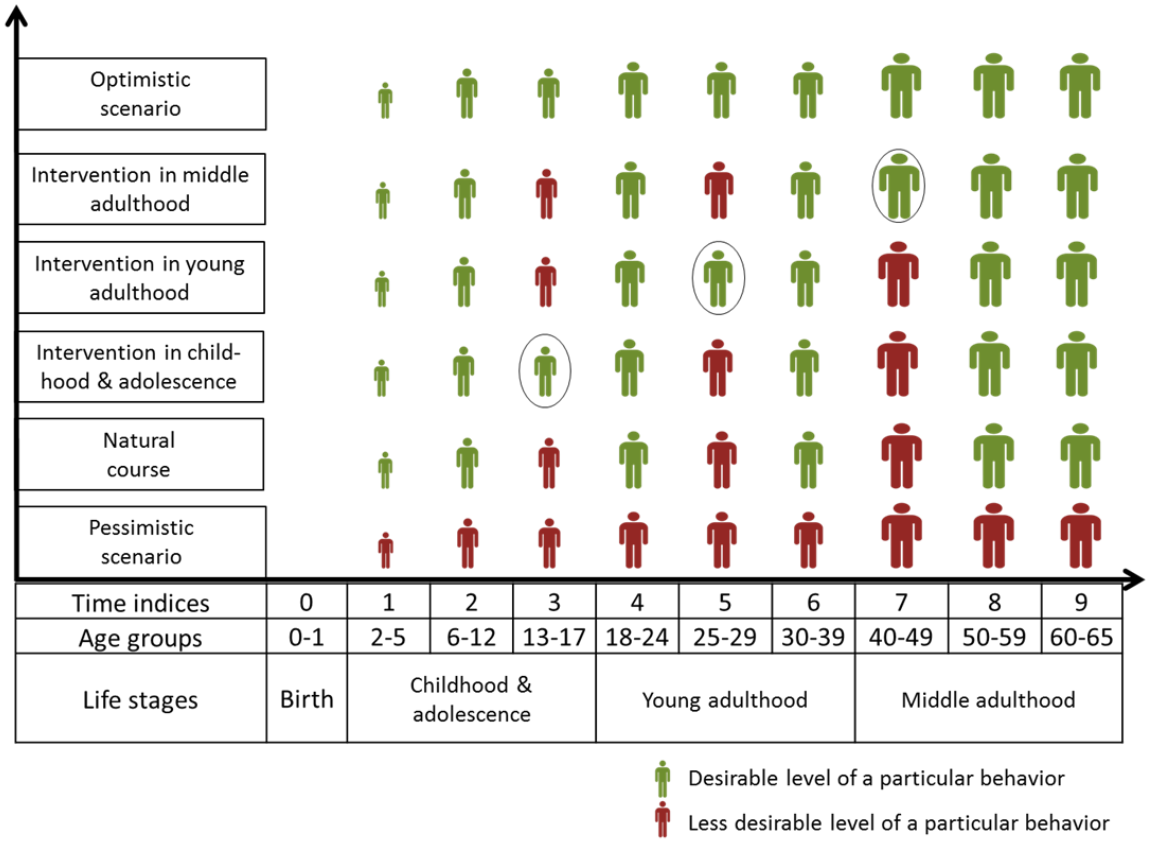


Figure 5.1 Hypothetical intervention regimens implemented throughout an individual life span. Throughout the life-course (i.e. 8 discrete time-steps from age 2 to 65) interventions were implemented in childhood, in young adulthood, in middle adulthood, and at all relevant time-points (i.e. optimistic or idealistic scenario) and compared to the natural course (i.e. status quo). For reference, a pessimistic scenario is also implemented (i.e. worse-off scenario).

❖ *Outcomes: Incident obesity and type 2 diabetes*

Obesity. Childhood obesity (e.g. 2-17 years) was defined using the WHO guidelines on the basis of body mass index (BMI) Z-scores calculated using CDC-provided SAS codes [115]. We used Z-scores instead of percentiles since Z-scores are comparable across ages and sex and are better for longitudinal assessment.[114] A child with a BMI Z-score (BMIz) greater or equal to

2 was classified as obese.[116]. Adult obesity (e.g. 18-65) was also defined using WHO guidelines. An individual with a BMI greater or equal to 30 was classified as obese. [117]. To calculate measures of incidences, we considered the first time an individual was diagnosed as being obese (i.e. "first occurrence" of obesity among at-risk individuals—that is individuals who were not obese in the previous time-step).

Type 2 diabetes. Incident type 2 diabetes between ages 18 and 65 was the outcome of interest.

- *Covariates*

The following covariates were considered in this study: age (continuous), sex (binary), race (binary), socio-economic status (binary), and marital status (binary) and family history of type 2 diabetes (binary)

Statistical analyses

In this study, we developed a directed acyclic diagram [72] to represent our assumptions about the underlying data generating mechanisms of obesity and diabetes in our simulation model. The relationships between covariates, exposures, mediators and outcomes are depicted in the causal diagram (see **Figure 5.2**). We used the g-computation algorithm of Robins (applied to the parametric g-formula), a generalization of the standardization method for time-varying exposures and confounders.[87] We implemented various hypothetical interventions and predicted the potential cumulative incidences of obesity and diabetes within the simulated cohort. All analyses were conducted in SAS 9.4 (SAS Institute, Inc., Cary, NC).

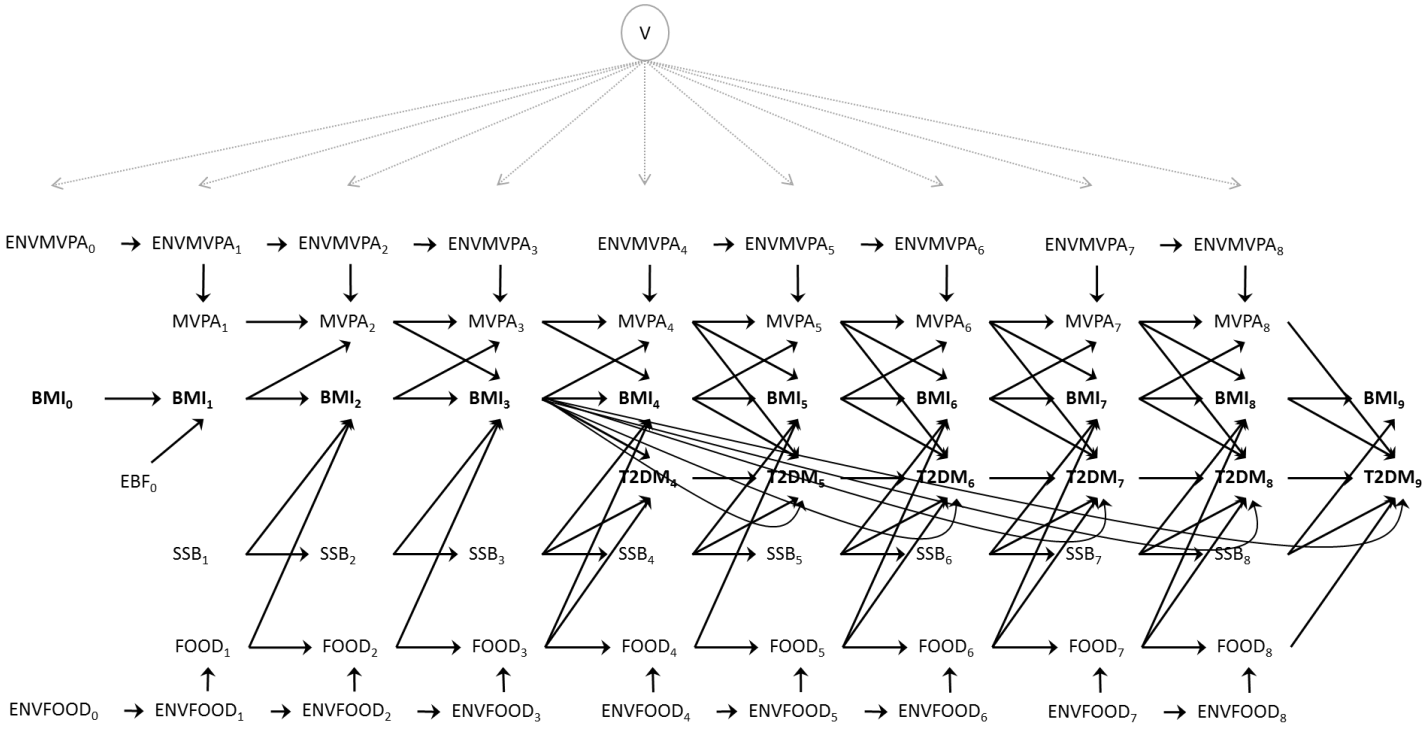


Figure 5.2 Simplified causal diagram of the underlying data generating process. V is a set of time-invariant covariates that affect all variables in the diagram. For clarity, we did not attempt to draw lines from V to all single variables in the diagram. At the individual level, V represents age, sex, race, marital status, socio-economic status. At the neighborhood level, V represents the percent non-White, percent of individuals who have a graduate degree, the percent of families below the federal poverty level. The latter affects ENVMVPA (neighborhood physical activity opportunities: park access and neighborhood walkability) and ENVFOOD (neighborhood food environment: supermarket density, fast-food density). BMI: body mass index, T2DM: type 2 diabetes, EBF: exclusive breastfeeding, SSB: sugar sweetened beverage consumption, MVPA: Moderate-to-vigorous physical activity, FOOD: fast-food and fruit and vegetable consumption. The indices represent the 10 discrete time steps form birth (t= 0) to middle adulthood (t=9).

5.4. Results

Table 5.1 describes the baseline and follow-up characteristics of the ViLA-Obesity model. Two thirds of the population was non-White and about one fourth had an income below or at the federal poverty level. One in six children and one in three adults were considered obese. Among adults aged 18 to 65, about one tenth of the individuals had type 2 diabetes.

Table 5.2 describes the simulated end-of follow-up cumulative incidence of obesity and type 2 diabetes under various hypothetical interventions. The 64-year risk of obesity (from 2 to 65 years) and the 48-year risk of type 2 diabetes (from 18 to 65 years) at the end of follow-up under the no intervention scenario (i.e. natural course) were 0.89 (95%CI 0.89 to 0.89) and 0.55 (95% CI 0.53 to 0.54), respectively. Under the SSB intervention, the 64-year risk of obesity did not differ from that of the natural course, but the 48-year risk of type 2 diabetes appeared to be lower than that of the natural course 0.51 (95% CI 0.51 to 0.52). The 48-year risk of type 2 diabetes under the breastfeeding, the neighborhood walkability and the neighborhood park access interventions was 0.54 (95%CI 0.53 to 0.54), 0.53 (95%CI 0.53 to 0.53) and 0.53 (95%CI 0.53 to 0.53) respectively. The 64-year risk of obesity under the breastfeeding, the neighborhood walkability and the neighborhood park access interventions was 0.89 (95%CI 0.89 to 0.89), 0.89 (95%CI 0.89 to 0.89) and 0.89 (95%CI 0.89 to 0.89), respectively.

Figure 5.3 and **Figure 5.4** present the cumulative incidence of incident obesity and type 2 diabetes over time under a combination of interventions. The intervention included the optimistic scenario of the sugar-sweetened beverage, the breastfeeding, the neighborhood walkability and the neighborhood park access interventions. In particular, in **Figure 5.3**, the curves of the cumulative incidence of obesity under the natural course and the optimistic scenario of the combined interventions were almost overlapping for children, adults and for the

population as a whole. The cumulative incidence under the combination of interventions was consistently high among the non-White segment of the population throughout the life span. For diabetes, the optimistic scenario of the combined interventions yielded a cumulative incidence that was lower than that of the natural course over the life span.

Figure 5.5 and **Figure 5.6** present the population impact of various interventions on obesity and type 2 diabetes. The three single most effective interventions on type 2 diabetes prevention in this study were the fast-food intervention (population risk ratio RR=0.82 (95%CI 0.82 to 0.83)), followed by the physical activity intervention RR=0.84 (95%CI 0.84 to 0.85) and the sugar-sweetened beverage intervention RR=0.95 (95%CI 0.94 to 0.96). Eliminating fast-food consumption tended to have some effect on obesity prevention RR=0.97 (95%CI 0.96 to 0.97). The "combined interventions plus" yielded the greatest effect for both type 2 diabetes (RR=0.63(95%CI 0.63 to 0.64)) and obesity (RR=0.94(95%CI 0.93 to 0.94)).

Table 5.1 Baseline and follow-up characteristics of simulated individuals in ViLA (n=98,230)

	Childhood (2-17)	**Adulthood (18-65)**
Age in years (Mean, SD)	9.17 (4.89)	40.67 (14.84)
Male (%)	49%	49%
Low-income (i.e. below or at FPL) (%)	22%	22%
Married (%)	--	44%
Non-White (%)	63%	63%
Has family history of type 2 diabetes (%)	08%	08%
Breastfeeding for six months or longer* (%)	23%	--
Drank ≥ 1 glasses of SSB/ day (%)	61%	40%
Ate ≥ 5 servings of fruit and vegetable/day (%)	53%	53%
Ate fast-food more than ≥ 1 time in past week (%)	76%	66%
Engage in moderate to vigorous physical activity (%)	24%	22%
High neighborhood walkability (%)	27%	27%
High neighborhood access to Parks (%)	54%	56%
Body mass index (Mean, SD)	20.43 (5.25)	26.79 (6.63)
Obese (%)	15%	30%
Has type 2 diabetes (%)	--	10%

*FPL: Federal Poverty Level, SD: Standard deviation, SSB: sugar-sweetened beverage consumption, *Applicable only to children between 0 and 1.*

Table 5.2 Simulated cumulative incidence of obesity and type 2 diabetes under hypothetical interventions (n=98,230)

Interventions	**64-year risk of obesity**	**48-year risk of type 2 diabetes**
00-Natural Course (no intervention)	0.892 (0.890 to 0.894)	0.537 (0.534 to 0.540)
01-Eliminate sugar-sweetened beverage consumption	0.889 (0.887 to 0.891)	0.512 (0.509 to 0.515)
02-Exclusively breastfeed for ≥ 6 months	0.891 (0.889 to 0.893)	0.537 (0.534 to 0.540)
03-Increase neighborhood walkability	0.890 (0.888 to 0.892)	0.529 (0.525 to 0.532)
04-Increase neighborhood access to parks	0.890 (0.888 to 0.892)	0.529 (0.525 to 0.532)
05-Engage in moderate to vigorous physical activity	0.873 (0.871 to 0.875)	0.453 (0.450 to 0.456)
06-Consume ≥ 5 fresh fruit and vegetable/day	0.888 (0.886 to 0.890)	0.531 (0.528 to 0.534)
07-Eliminate fast-food consumption	0.862 (0.860 to 0.864)	0.442 (0.439 to 0.445)
08-Combined interventions (All)	0.887 (0.885 to 0.889)	0.503 (0.500 to 0.507)
09-Combined interventions Plus (All)	0.837 (0.834 to 0.839)	0.339 (0.336 to 0.342)
10-Combined interventions (Childhood)	0.890 (0.888 to 0.892)	0.529 (0.526 to 0.532)
11-Combined interventions Plus (Childhood)	0.871 (0.869 to 0.873)	0.480 (0.476 to 0.483)
12-Combined interventions (Young adulthood)	0.889 (0.887 to 0.891)	0.519 (0.516 to 0.522)
13-Combined interventions Plus (Young adulthood)	0.873 (0.871 to 0.875)	0.465 (0.462 to 0.469)
14-Combined interventions (Adulthood)	0.890 (0.888 to 0.892)	0.525 (0.522 to 0.528)
15-Combined interventions Plus (Adulthood)	0.880 (0.878 to 0.882)	0.452 (0.449 to 0.455)

Combined interventions are the interventions that include the primary interventions: the sugar-sweetened beverage consumption, breastfeeding, neighborhood walkability and the neighborhood access to park interventions. The "combined interventions Plus" includes most primary interventions and the secondary interventions: the sugar-sweetened beverage consumption, breastfeeding, physical activity, fruit and vegetable and fast-food interventions

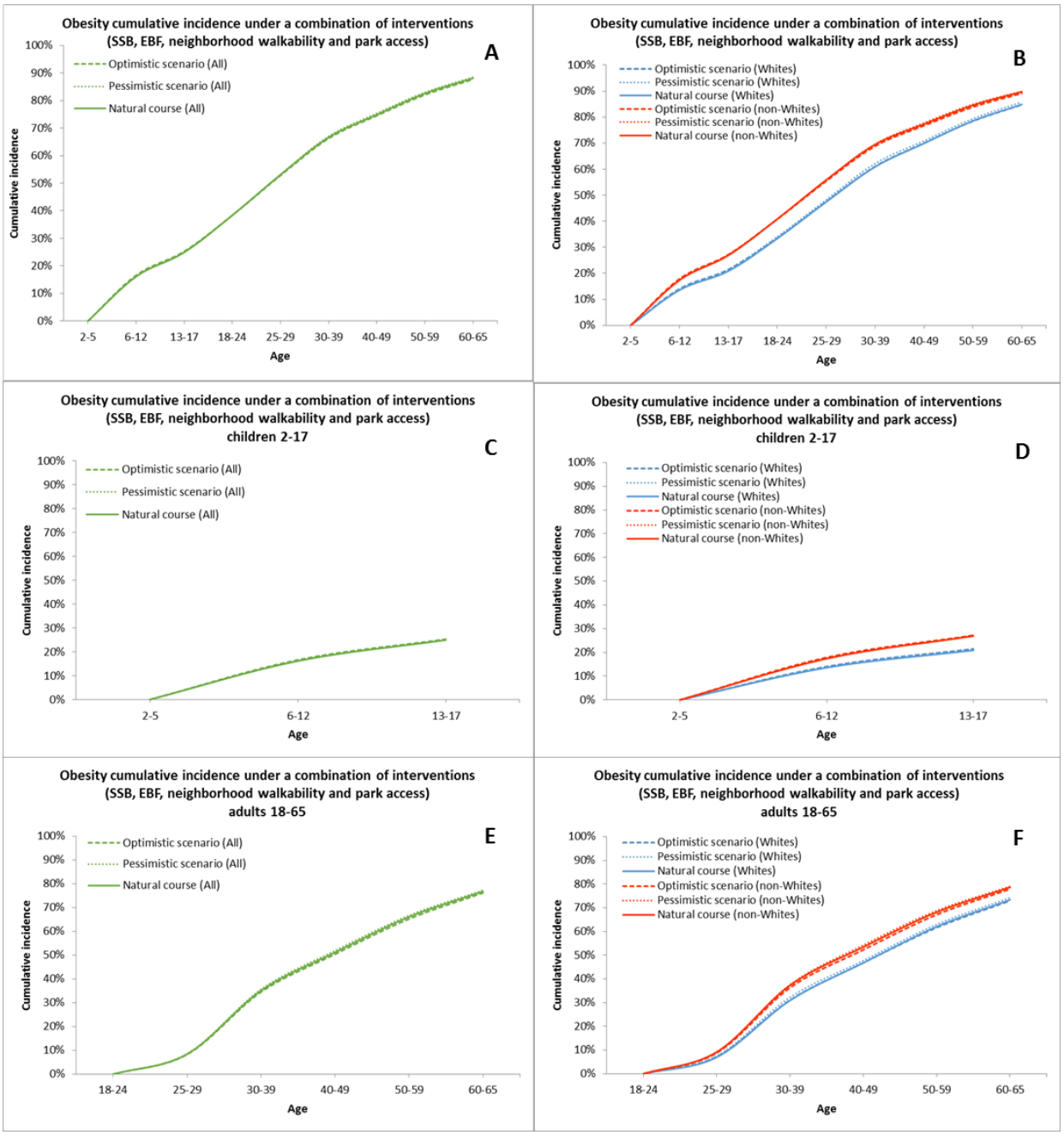


Figure 5.3 Cumulative incidence of obesity under a combination of interventions including the sugar-sweetened beverage (SSB), breastfeeding (EBF), neighborhood walkability and the neighborhood access to park interventions. (A) Obesity cumulative incidence in the total population; (B) Obesity cumulative incidence by race and ethnicity; (C) Obesity cumulative incidence among children 2-17; (D) Obesity cumulative incidence among children 2-17 by race and ethnicity; (E) obesity cumulative incidence in the among adults 18-65; (F) Obesity cumulative incidence among adults 18-65 by race and ethnicity.

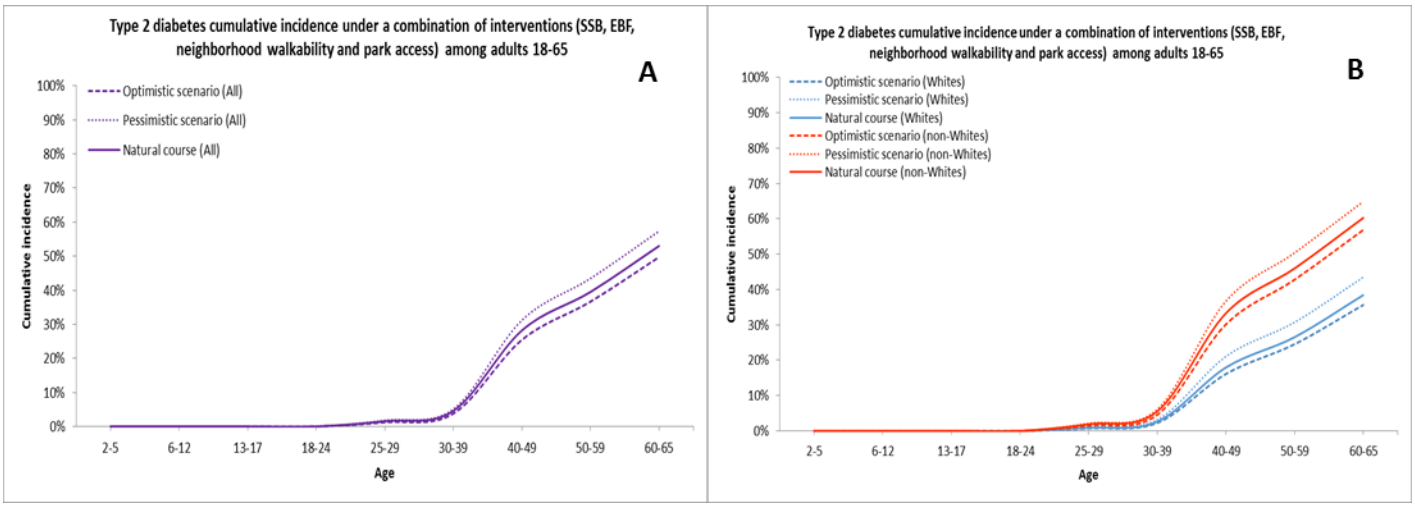


Figure 5.4 Cumulative incidence of type 2 diabetes under a combination of interventions including the sugar-sweetened beverage (SSB), breastfeeding (EBF), neighborhood walkability and the neighborhood access to park interventions. (A) type 2 diabetes cumulative incidence among adults 18-65 ; (B) type 2 diabetes cumulative incidence among adults 18-65 by race and ethnicity.

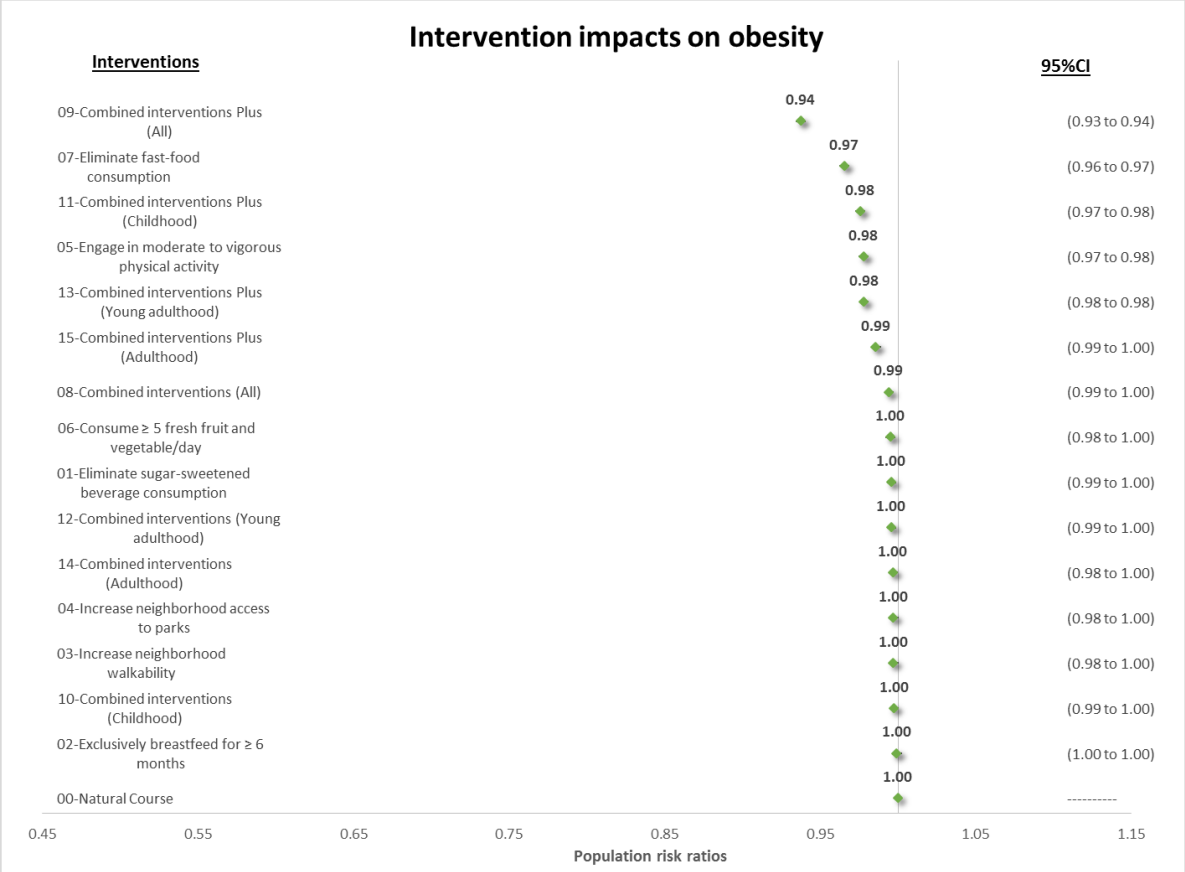


Figure 5.5 Population impact (in terms of risk ratios) of various interventions on the cumulative incidence of type 2 diabetes. Combined interventions are the interventions that include the primary interventions: the sugar-sweetened beverage consumption, breastfeeding, neighborhood walkability and the neighborhood access to park interventions. The "combined interventions Plus" includes most primary interventions and the secondary interventions: the sugar-sweetened beverage consumption, breastfeeding, physical activity, fruit and vegetable and fast-food interventions.

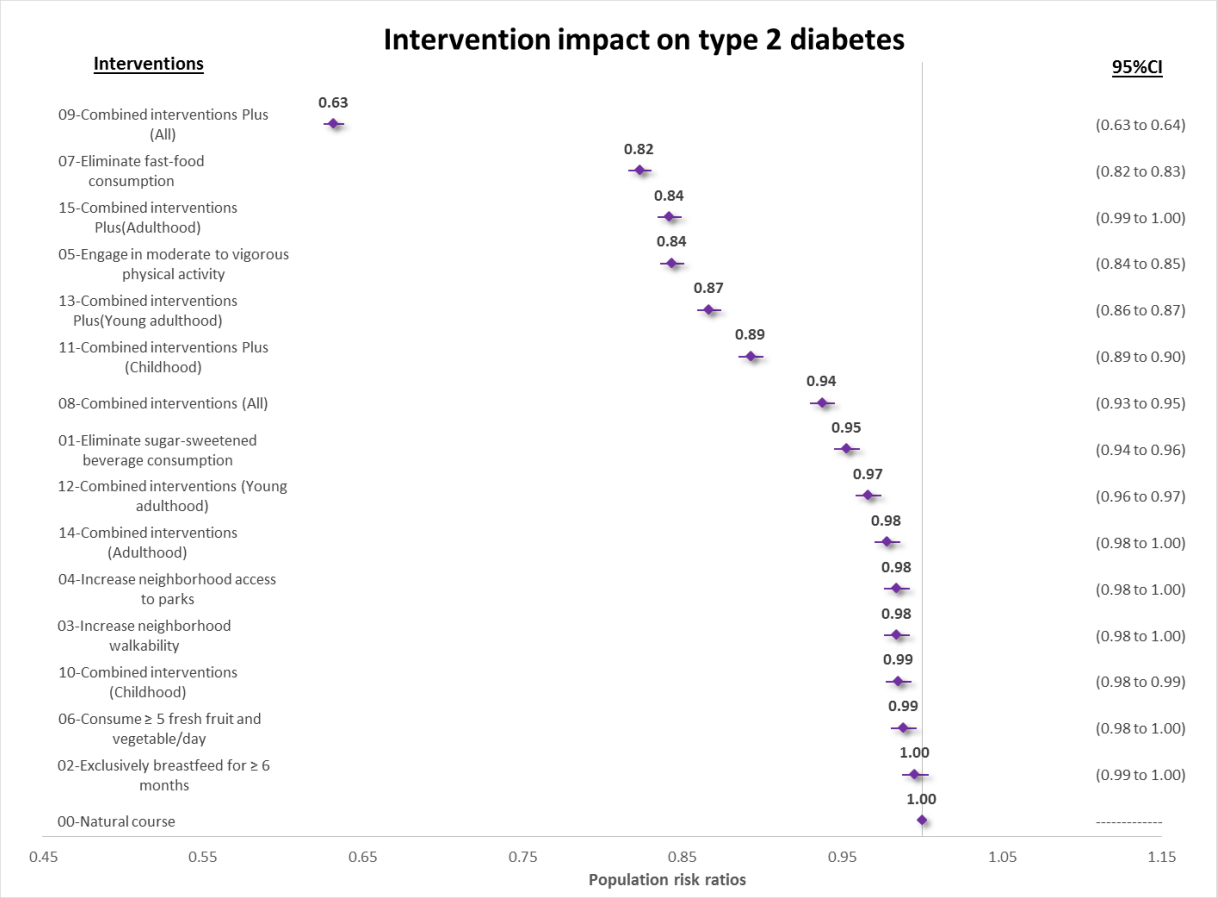


Figure 5.6 Population impact (in terms of risk ratios) of various interventions on the cumulative incidence of obesity. Combined interventions are the interventions that include the primary interventions: the sugar-sweetened beverage consumption, breastfeeding, neighborhood walkability and the neighborhood access to park interventions. The "combined interventions Plus" includes most primary interventions and the secondary interventions: the sugar-sweetened beverage consumption, breastfeeding, physical activity, fruit and vegetable and fast-food interventions.

5.1. Discussion

The purpose of this study was to evaluate the effectiveness of key health interventions on obesity and diabetes throughout the life course in the virtual Los Angeles Cohort.

Our findings suggest that among the primary interventions, breastfeeding for six months or longer, increasing neighborhood walkability or neighborhood access to parks were not effective in reducing the cumulative incidence of obesity and type 2 diabetes. In contrast, eliminating sugar-sweetened beverage consumption seemed to be somewhat effective at reducing the risk of type 2 diabetes but not the risk of obesity. Likewise, among the secondary interventions, engaging in moderate-to-vigorous physical activity and eliminating fast-food consumption appeared to be effective in reducing the excess risk in obesity and diabetes incidence. However, eating at least five servings of fresh fruits and vegetables did not have an impact in the population as a whole or in the long run. In addition, combining interventions with one another throughout the life span showed the greatest impact, especially when such combination included the sugar-sweetened beverage, the physical activity or the fast-food interventions. Furthermore, for a given effective intervention, the impact seems greater in reducing diabetes risk than in reducing obesity risk. This is probably due to the fact that these health behaviors and obesity both affect diabetes risk. Lastly, to have an impact, most interventions needed to be implemented at all possible time points (i.e. optimistic scenario). Interestingly, when considering our "combined intervention plus" (i.e. included all individual level interventions), we noticed that interventions implemented childhood were more effective in reducing obesity risk than intervention implemented in young adulthood or middle adulthood.

Conversely, interventions implemented in middle adulthood were more effective in reducing diabetes risk than interventions implemented earlier in young adulthood and childhood.

These results highlight many important insights worth mentioning. First, not all interventions are created equal. Some interventions are more effective than others. For instance, in this study, eliminating sugar-sweetened beverage, eliminating fast-food consumption and engaging in physical activity appeared more effective than the other interventions. Second, some periods appear to be more critical than others in preventing obesity or diabetes. Third, to be effective, most interventions have to be implemented continuously virtually at every stage of life and have to be implemented together, something that can be hard to achieve in real life. Fourth, the modest impacts of the interventions evaluated here testify to the persistence of obesity and diabetes and to the difficulty to curb these epidemics. This might explain why there has been only a slight leveling off of childhood obesity after many years of prevention efforts. Lastly, the findings in this study seems to support the idea that intervening on the population as a whole might prevent more burden of disease than targeting only people who are at high risk of the disease.[180] Nevertheless, when implemented as such, these interventions can help reduce the disparities in obesity and diabetes.

Our findings support the notion that most interventions will yield modest effects in the long run and in the population as a whole, especially when implemented singly. Although to date, the long-term effect of sugar-sweetened beverage consumption on obesity and type 2 diabetes [181] as well as the short term effect of reducing sugar-sweetened beverage consumption [182,183] are well established, few studies have been able to investigate the long-term health effect of sugar-sweetened beverage reduction on obesity. One study, however, involving three prospective studies showed that replacing one serving per day of sugar-sweetened beverage by

one cup a day of water was with 0.49 kg less weight gain over each 4-year period.[184] Interestingly, this study does not show an actual decrease in weight as a result of sugar-sweetened beverage reduction but rather a lower weight gain as a result of the intervention. Others have argued that such weight stabilization can be considered beneficial especially in a context where weight may be generally rising.[185] Likewise, it is conceivable that such effect may be present short after the exposure and that the effect gets attenuated over time resulting in a modest overall effect.

In this study, breastfeeding for six months or longer had no overall effect over the life-course of an individual. Although there are some evidence as to the benefits of breastfeeding in childhood obesity prevention [186,187], its role has been largely debated in the literature [188,189]. Nevertheless, it may be that there exists a short-term effect that dissipates over time. In fact, post-hoc analyses in this study (results not shown: See **eFigure 5.6** in appendix) show that there is in fact an effect in early childhood but this effect fades away over time resulting in no effect overall.

Some aspects of the built-environment have been shown to be preventive of obesity.[190] In our study, a high neighborhood walkability and/or high neighborhood access to parks was associated with a null effect on obesity and type 2 diabetes. This null effect, however, reinforces the notion that more upstream interventions may affect population health slowly if at all.[191] In fact, as seen in this study, there was a greater impact from more downstream individual-level physical activity interventions as compared to when one targets the neighborhood-level physical activity interventions. Orr et al used an agent-based model to show how neighborhood-focused upstream policies may reduce disparities in BMI and that such policy may take time to affect the population health.[192]

Our modest findings are also in line with recent simulation studies that evaluated the potential health impacts of implementing policy/systems and environmental (PSE) interventions in Los Angeles County. Their simulations show that the PSE changes if sustained have the potential to reduce the burden of obesity in the county.[138] Studies like these and the present can help the Los Angeles County department of public health direct their efforts to interventions that will yield the greatest impact for the lowest cost. In addition, our study findings can help the Los Angeles County Department of Public Health better understand why there was a slight leveling off of childhood obesity and why it may take years to see the effect of major prevention efforts. Furthermore, the present study can assist the public health department in their understanding of (i) which interventions have better potential in preventing or reducing the burden of obesity or diabetes in the county and (ii) how to implement interventions to see greater success in prevention.

This study is not without limitations. First, our findings are subject to our simulation model and are reliable insofar as the ViLA cohort represents the reality of Los Angeles. Although being continuously updated, this cohort has been validated where possible against external sources of data representative of Los Angeles County. Additionally, this cohort is inherently a close population admitting no new individuals after the start of follow-up and not allowing any process such as death to remove individuals from the population. The latter assumption though not always made explicit is a common assumption virtually made a most other studies. Lastly, though complex in nature, our agent-based model remains a simplification and abstraction of the real world and as such may not capture other important aspects that can influence how one becomes obese or develop type 2 diabetes.

Conclusion

The slow decline in obesity and diabetes rates may have been due to the modest effects of health interventions in the population at large and in the long run. This simulation study supports the notion that for maximum effectiveness, most interventions have to be implemented in combination with one another and virtually at every critical life stages throughout the life span. This study also illustrates the usefulness of agent-based simulation models for evaluating the effectiveness of key health interventions on complex health issues such as obesity and diabetes throughout the life course.

5.2. Appendix

eTable 5.1 Life period, time-points and age groups in ViLA

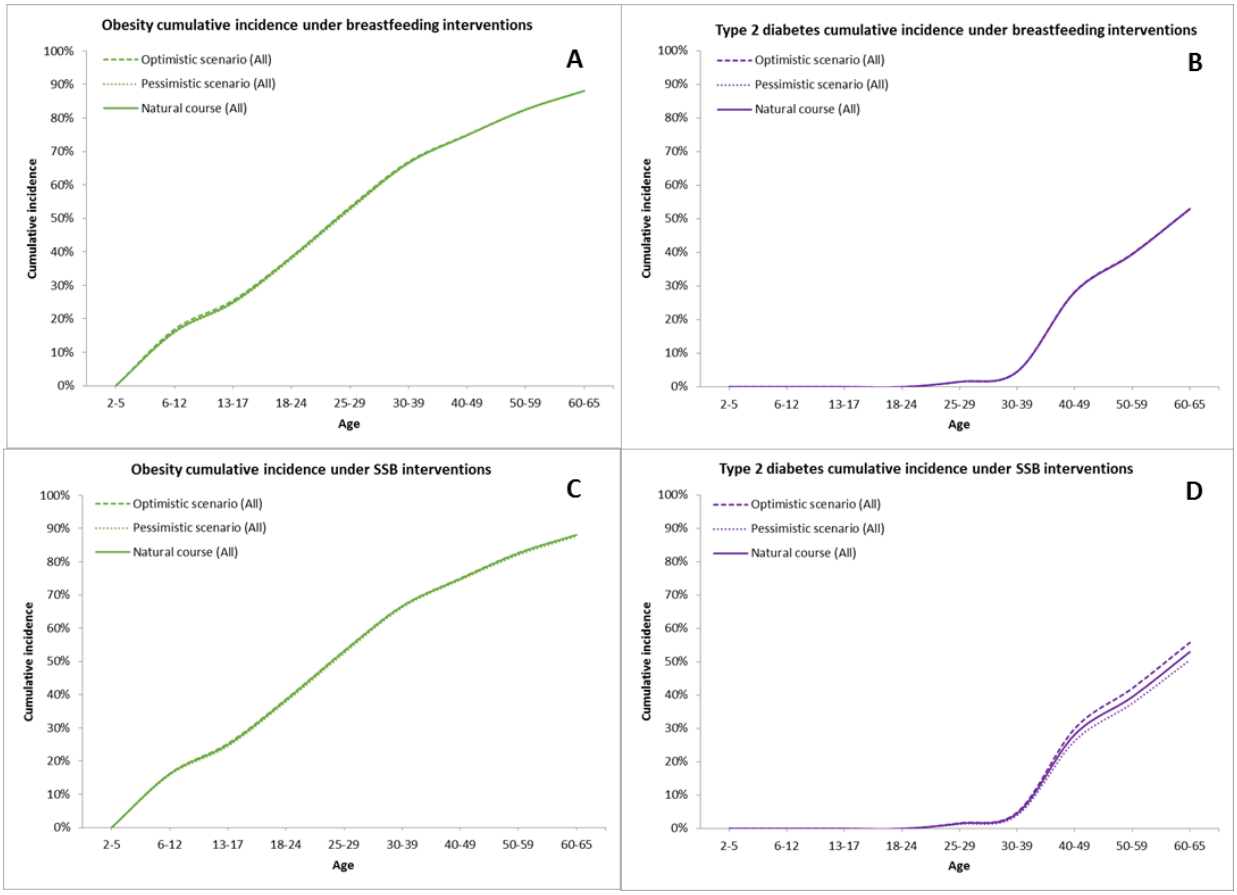
Time	Age (years)	Life period
0	0-1	Birth
1	2-5	Early Childhood
2	6-12	Middle Childhood
3	13-17	Adolescence
4	18-24	Young Adulthood
5	25-29	Young Adulthood
6	30-39	Young Adulthood
7	40-49	Middle Adulthood
8	50-59	Middle Adulthood
9	60-65	Middle Adulthood

eTable 5.2 Characteristics of interventions that will be evaluated in ViLA

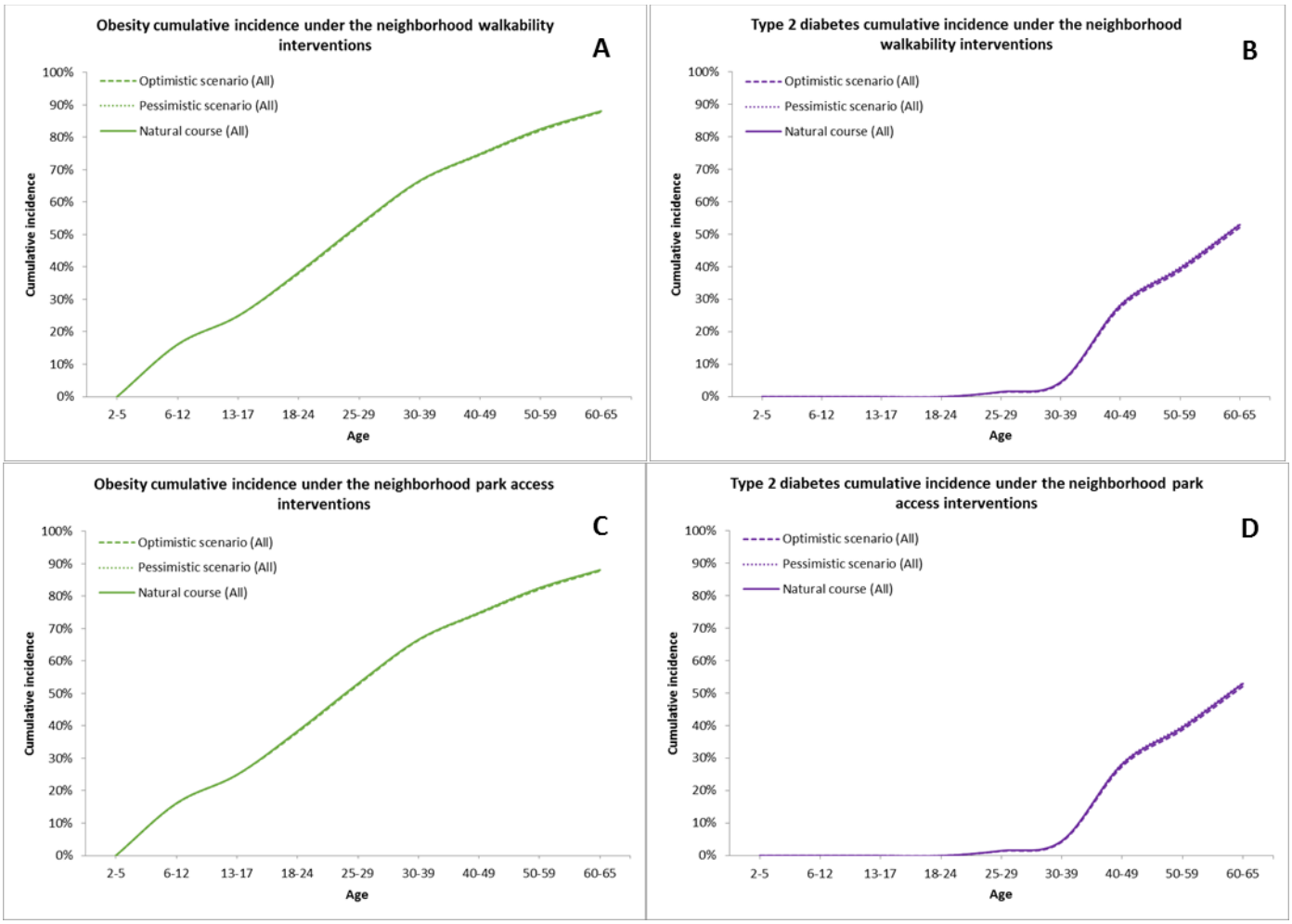
Interventions	***Targeting diet or physical activity***	***Level***	***Time points (i.e. when implemented)***	***Sources of data and parts of the RENEW project goals***
Breastfeed exclusively for at least six months	*Diet*	*Individual (behavioral)*	*0-1 (time = 0)*	*Center for Disease Control and Prevention* [46]*(CDC)* *RENEW: "…Helping to adopt and implement breastfeeding policies in County hospitals and departments and other private employers…"*[22]
Eliminate sugar-sweetened beverage consumption	*Diet*	*Individual (behavioral)*	*2-5 (time = 1)* *6-12 (time = 2)* *13-17 (time = 3)* *18-24 (time = 4)* *25-29 (time = 5)* *30-39 (time = 6)* *40-49 (time = 7)* *50-59 (time = 8)*	*California Health Interview Survey (CHIS)*[23] *RENEW: "…Growing healthier students through more nutritious school meals, including more whole grains and fresh fruits and vegetables, and reducing fat, sugar, salt and calories…"*[22]
Increase physical activity opportunities by increasing access to parks	*Physical activity*	*Environmental*	*0-17 (time = 1,2,3)* *18-39 (time = 4,5,6)* *40-65 (time = 7,8)*	*(Wolch et al., 2005)* *RENEW: "…Providing safe, open spaces for recreation through joint-use policies…"*[22]
Design pedestrian friendly communities by increasing the community walkability score	*Physical activity*	*Environmental*	*0-17 (time = 1,2,3)* *18-39 (time =4,5,6)* *40-65 (time = 7,8)*	*American Community Survey (ACS), Census Walkscore.com*[47] *RENEW: "…Creating more opportunities for walking and biking in communities by supporting the development of expanded bike networks and more pedestrian-friendly community design…"*[22]

eTable 5.2 Characteristics of interventions that will be evaluated in ViLA (continued)

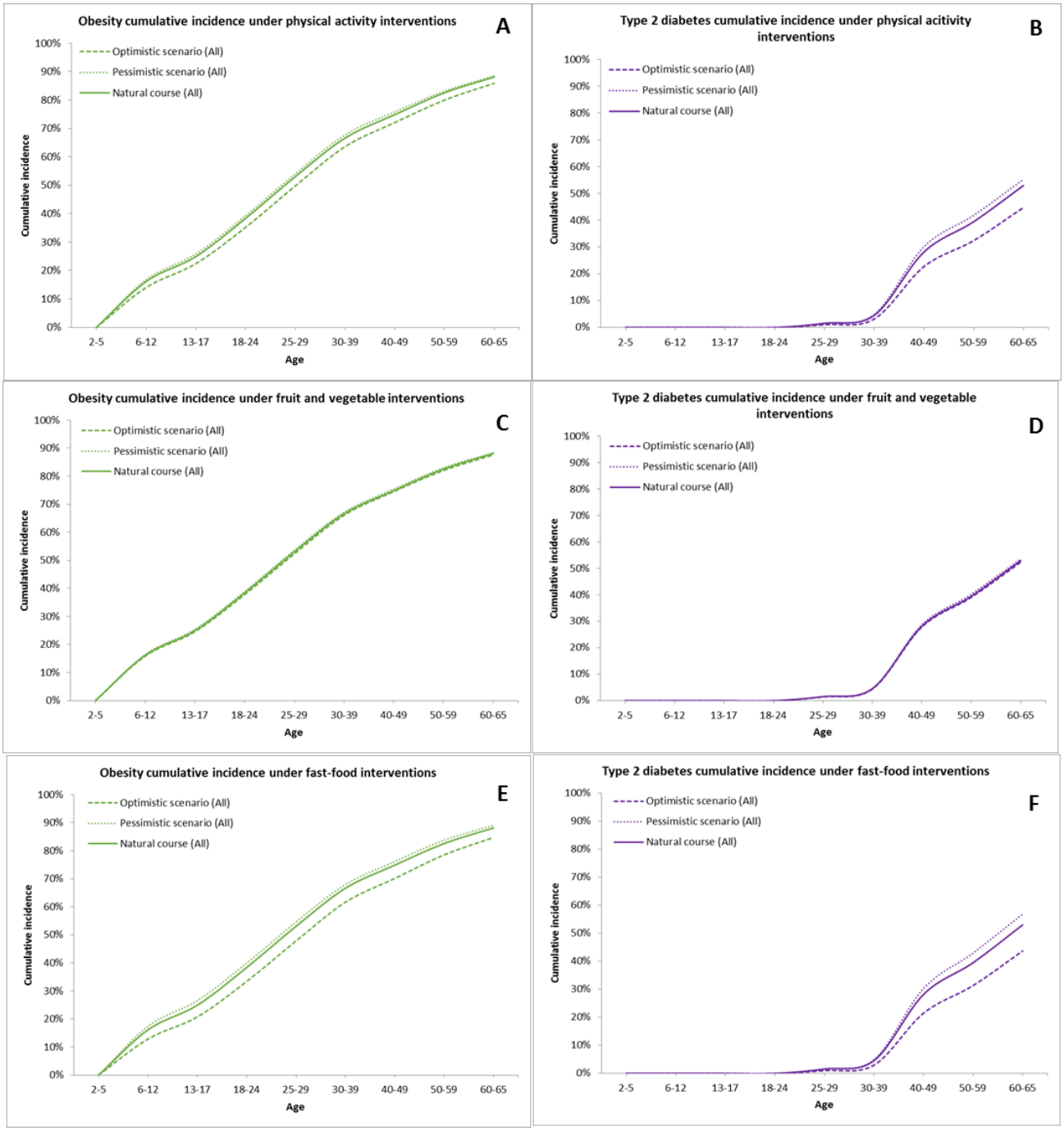
Interventions	***Targeting diet or physical activity***	***Level***	***Time points (i.e. when implemented)***	***Sources of data and parts of the RENEW project goals***
Increase fresh fruit and vegetable consumption	*Diet*	*Individual (behavioral)*	*2-5 (time = 1)* *6-12 (time = 2)* *13-17 (time = 3)* *18-24 (time = 4)* *25-29 (time = 5)* *30-39 (time = 6)* *40-49 (time = 7)* *50-59 (time = 8)*	*California Health Interview Survey (CHIS)*[23] *RENEW: "…Growing healthier students through more nutritious school meals, including more whole grains and fresh fruits and vegetables, and reducing fat, sugar, salt and calories…"*[22]
Eliminate fast-food consumption	*Diet*	*Individual (behavioral)*	*2-5 (time = 1)* *6-12 (time = 2)* *13-17 (time = 3)* *18-24 (time = 4)* *25-29 (time = 5)* *30-39 (time = 6)* *40-49 (time = 7)* *50-59 (time = 8)*	*California Health Interview Survey (CHIS)*[23] *RENEW: "…Growing healthier students through more nutritious school meals, including more whole grains and fresh fruits and vegetables, and reducing fat, sugar, salt and calories…"*[22]
Increase the level of physical activity	*Physical activity*	*Individual (behavioral)*	*2-5 (time = 1)* *6-12 (time = 2)* *13-17 (time = 3)* *18-24 (time = 4)* *25-29 (time = 5)* *30-39 (time = 6)* *40-49 (time = 7)* *50-59 (time = 8)*	*California Health Interview Survey (CHIS)*[23] *RENEW: "…Providing safe, open spaces for recreation through joint-use policies…"*[22]



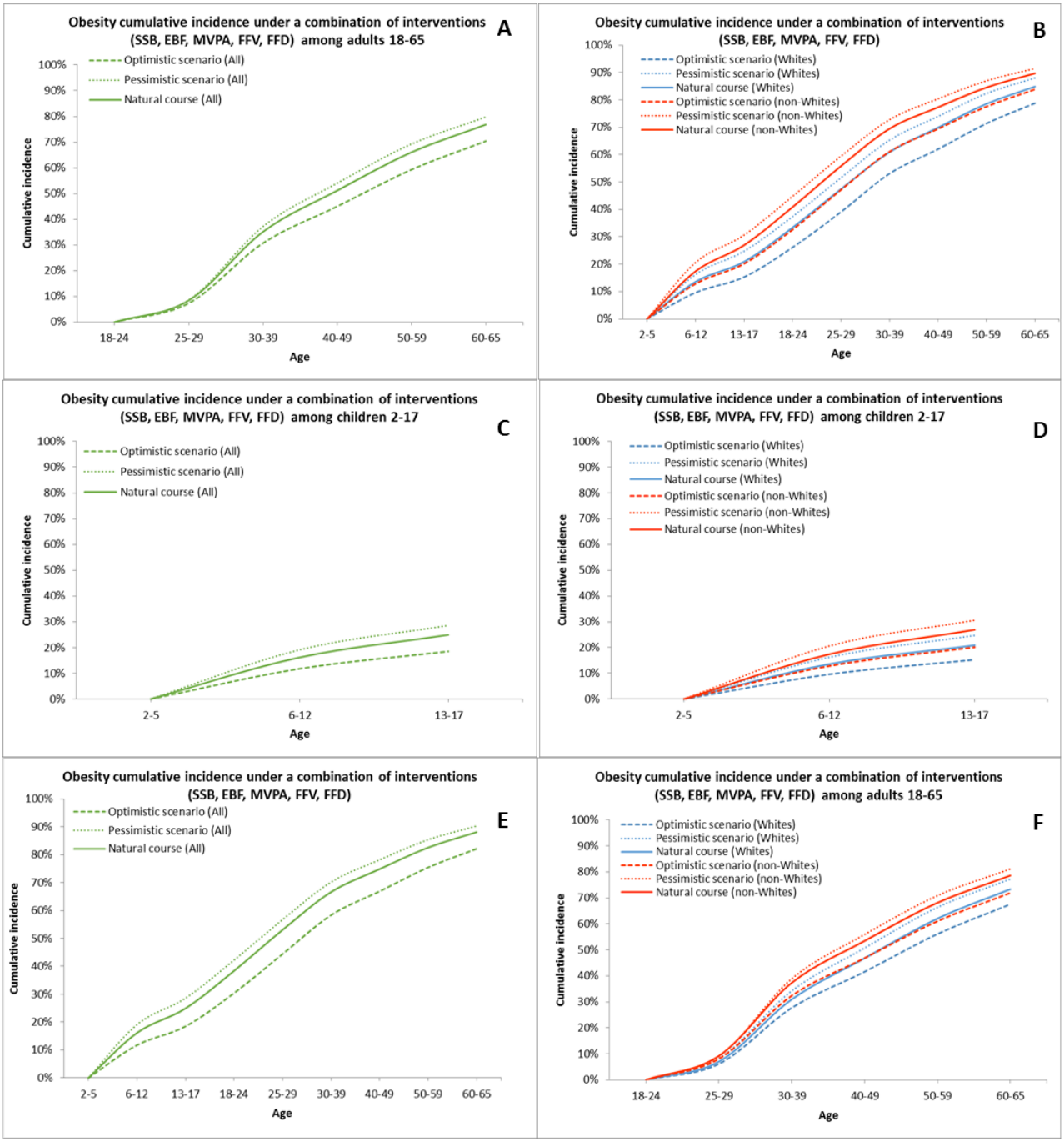
eFigure 5.1 Cumulative incidence of obesity and type 2 diabetes under the dietary interventions. (A) obesity under the breastfeeding intervention; (B) type 2 diabetes under the breastfeeding intervention; (C) obesity under the sugar-sweetened beverage (SSB) intervention; (D) type 2 diabetes under the SSB intervention.



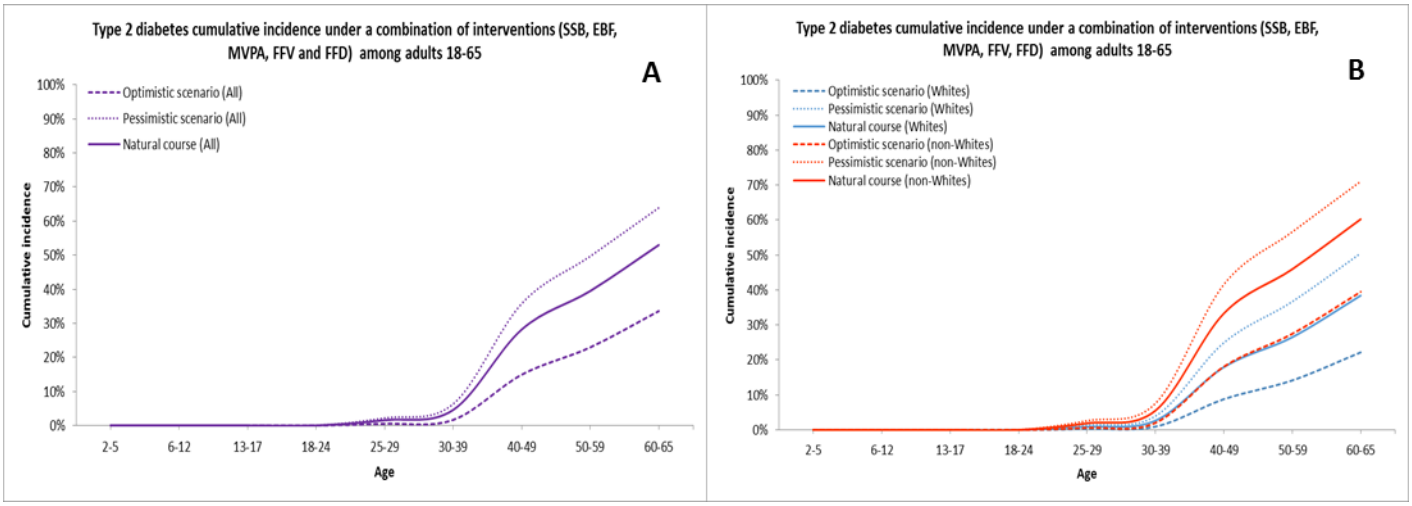
eFigure 5.2 Cumulative incidence of obesity and type 2 diabetes under the neighborhood physical activity interventions. (A) obesity under the neighborhood walkability interventions; (B) type 2 diabetes under the neighborhood walkability intervention; (C) obesity under the neighborhood park access interventions; (D) type 2 diabetes under the neighborhood park access interventions.



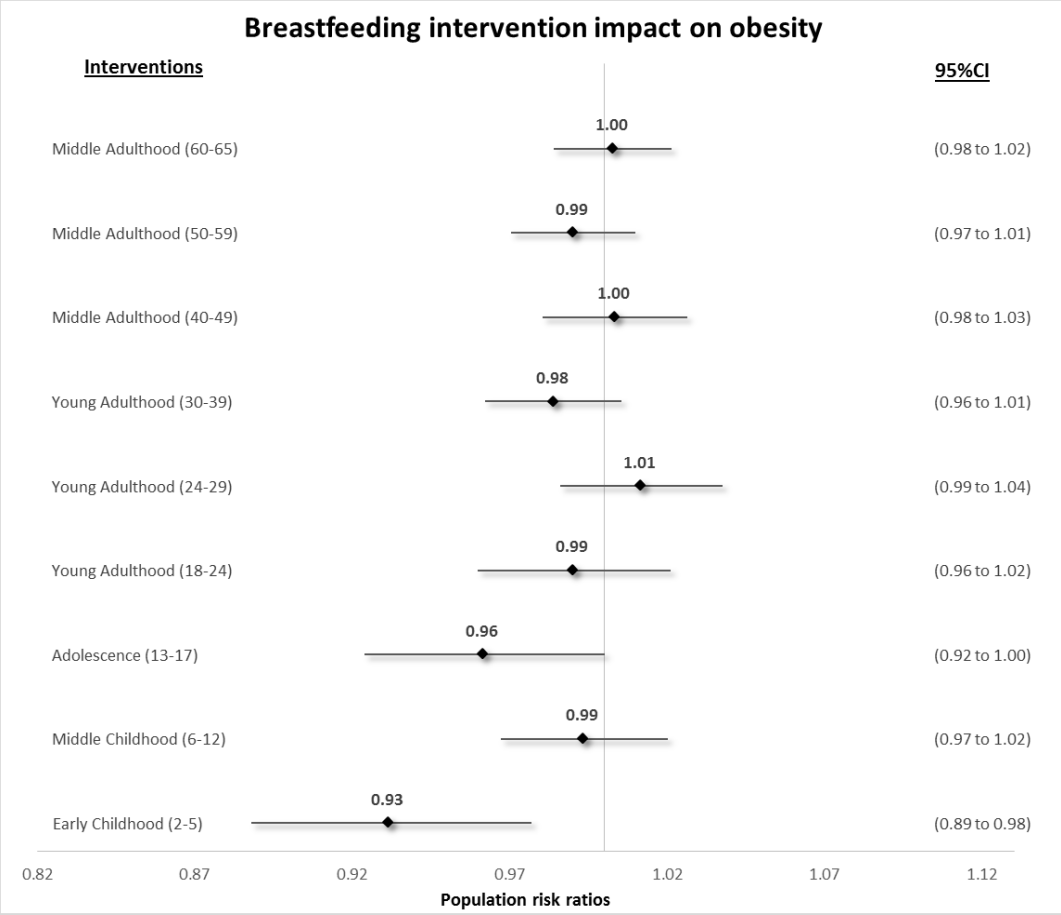
eFigure 5.3 Cumulative incidence of obesity and type 2 diabetes under the secondary interventions (physical activity, fruit and vegetable and fast-food interventions). (A) obesity under the physical activity interventions; (B) type 2 diabetes under the physical activity interventions; (C) obesity under the fruit and vegetable interventions; (D) type 2 diabetes under the fruit and vegetable interventions; (E) obesity under the fast-food interventions; (F) type 2 diabetes under the fast-food interventions



eFigure 5.4 Cumulative incidence of obesity under a combination of interventions including the sugar-sweetened beverage (SSB), breastfeeding (EBF), physical activity (MVPA), fruit and vegetable (FFV) and fast-food (FFD) interventions. (A) Obesity cumulative incidence in the total population; (B) Obesity cumulative incidence by race and ethnicity; (C) Obesity cumulative incidence among children 2-17; (D) Obesity cumulative incidence among children 2-17 by race and ethnicity; (E) obesity cumulative incidence in the among adults 18-65; (F) Obesity cumulative incidence among adults 18-65 by race and ethnicity.



eFigure 5.5 Cumulative incidence of type 2 diabetes under a combination of interventions including the sugar-sweetened beverage (SSB), breastfeeding (EBF), physical activity (MVPA), fruit and vegetable (FFV) and fast-food (FFD) interventions. (A) type 2 diabetes cumulative incidence among adults 18-65 ; (B) type 2 diabetes cumulative incidence among adults 18-65 by race and ethnicity.



eFigure 5.6 Population impact (in terms of risk ratios) of the breastfeeding intervention on the cumulative incidence of obesity over life stages.

Chapter 6. Projecting the impact of early life interventions on adiposity in children living in low income households

6.1. Abstract

Background: It is difficult to evaluate the effectiveness of interventions aimed at reducing early childhood obesity using randomized trials.

Objective: To illustrate how observational data can be analyzed using causal inference methods to estimate the potential impact of behavioral "interventions" on early childhood adiposity.

Methods: We used longitudinal data from 1054 children 1-5 years old enrolled in the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) and followed from 2008 to 2010 for a mean duration of 23 months. The data came from a random sample of WIC families living in Los Angeles County in 2008. We used the parametric g-formula to estimate the impact of various hypothetical behavioral interventions.

Results: Adjusted mean weight-for-height Z score at the end of follow-up was 0.73 (95% CI 0.65, 0.81) under no intervention, and 0.63 (95% CI 0.38, 0.87) for all interventions given jointly. Exclusive breastfeeding for six months or longer was the most effective intervention [population mean difference = -0.11 (95% CI -0.22, 0.01)]. Other interventions had little or no effect.

Conclusions: Compared with interventions promoting healthy eating and physical activity behaviors, breastfeeding was more effective in reducing obesity risk in children aged 1-5 years. When carefully applied, causal inference methods may offer viable alternatives to randomized trials in etiologic and evaluation research.

6.2. Introduction

Childhood obesity is a major public health problem affecting millions of young Americans.[3] In the United States, one in three children is obese or overweight.[3] While prevalence rates have begun to stabilize, they continue to be high and are consistently higher among African-Americans and Hispanics.[3] Children who are obese are likely to be obese as adults since excess weight tracks through the life-course, from early childhood to adulthood.[154] This puts children who are obese at higher risk of developing various non-communicable diseases later in life.[28]

When attempting to reduce childhood obesity rates, public health professionals and policy makers need to answer questions such as, "What would be the population impact of a particular health intervention on childhood obesity if every child was exposed to it [e.g. if every child stopped consuming sugar-sweetened beverages (SSB)]?" and "Which interventions or combinations of interventions would yield the greatest long-term impact on childhood obesity?" While a number of prospective observational studies have identified potential protective (e.g. exclusive breastfeeding) and harmful (e.g. SBB consumption) risk factors for childhood obesity, [193] randomized trials (RCTs) have offered limited evidence about the long-term impact of reducing harmful exposures and increasing beneficial exposures either singly or in combination with each other.[175] In addition, results from RCTs are not always generalizable to the population that would be receiving the interventions, partly because of the selective enrollment of participants into the trials.[177] Further, for practical reasons including cost and loss to follow-up, RCTs are rarely able to follow participants for the long term.

One approach to addressing these methodological limitations is to apply causal inference methods to existing observational data to quantify the potential impact of hypothetical

interventions under plausible assumptions. This approach has been used by Taubman *et al*. and Danei *et al*. in their assessments of the impact of hypothetical interventions aimed at reducing risk factors for coronary heart diseases [88] and diabetes in adult populations,[89] respectively.

Hence, the goal of this study was to illustrate the usefulness of modern causal inference methods for evaluating interventions and providing relevant information for policy decision-making. The specific objective was to quantify the potential impact of various hypothetical and plausible behavioral interventions early in life on adiposity in a multi-ethnic cohort of children aged 1-5 years living in low-income households enrolled in the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) in Los Angeles County .

6.3. Methods

Study population and sources of data

WIC provides food assistance and nutrition education to pregnant and postpartum women and children up to age five living in low-income households in the United States. In Los Angeles County, Public Health Foundation Enterprises WIC, the largest local agency WIC program in the country, maintains an administrative dataset which contains socio-demographic and anthropometric data on every child enrolled in WIC in Los Angeles County since 2003.[194] WIC staff use a standardized protocol to measure height and weight; these measurements have been shown to have high accuracy.[195] In addition, a survey of a random sample of about 5,000 WIC families living in Los Angeles County is conducted every three years to collect behavioral data so as to address the specific needs of communities living in poverty. This WIC survey is conducted in English or Spanish through a computer-assisted telephone interviewing system. Almost half of the eligible WIC participants could not be reached by phone after many attempts

(up to 16), giving a response rate of 51%. We linked survey data obtained between April 8 and July 22, 2008 to WIC administrative data to prospectively follow a cohort of 1054 children aged 1-5 years living in low-income households in Los Angeles County from 2008 to 2010. To ensure that the anthropometric measurements were obtained at a time relevant to the survey period and more specifically at an age when it was developmentally plausible for the child to engage in a specific behavior of interest (e.g. consume fast food or be physically active at the playground), we included in the sample only children who: (i) were at least 12 months old at the time of the first relevant anthropometric measurement, (ii) had three subsequent measurements, and (iii) had a baseline (first) measurement that was taken within six months of the survey. Further excluded from the sample were children with a time interval between measurements of less than three months (n = 1) (See **Figure 6.1**).

The protocol for de-identifying the WIC data for research use was approved by the Ethical and Independent Review Services' Institutional Review Board. The University of California, Los Angeles (UCLA) Institutional Review Board approved the overall study protocol.

Study variables

Weight-for-height Z score (WHZ)

The outcome variable of interest was child's weight-for-height Z score (WHZ) calculated from height and weight measurements obtained by trained WIC staff during recertification visits. WHZ is a commonly used indicator for assessing adiposity in growing children as it is independent of height.[196] WHZ was estimated from age- and gender-specific CDC growth reference values.[197] All children in the sample had three WHZ estimates obtained from heights

and weights which were mostly measured between 2008 and 2010. The third WHZ (i.e. WHZ_3) was the outcome variable of interest while the second WHZ (i.e. WHZ_2) was considered an intermediary or mediating variable. We excluded records of children with improbable WHZ (<-4 or >5) (n= 3) as suggested by CDC (**Figure 6.2**).

Risk factors and hypothetical interventions

The survey collected data on a number of obesity-related risk factors including duration of exclusive breastfeeding, television watching, fruit and vegetable consumption, playing at the playground every day, SSB consumption, and fast-food consumption (see **eTable 6.1** in appendix for survey questions). Risk factor variables were categorized to avoid sparse data issues and/or to highlight recommended levels. Hypothetical interventions were designed on the basis of these risk factors by asking the question, "What would the population mean WHZ be if every child was exposed to the most beneficial level of a particular risk factor?" In other words, this study aimed to predict the mean WHZ of a population of children exhibiting optimal (recommended or desirable) levels of the behaviors of interest. For example, in the present sample, 23% of the children were exclusively breastfed at the recommended level of six months or more (i.e. at the desirable level), while the remaining 77% were breastfed for lesser amounts of time or not at all (i.e. less desirable levels). This study aimed to predict the population mean WHZ when 100% of the population exclusively breastfeeds for six months or more, that is, the 77% of children who were initially exclusively breastfed less than six months would now be exclusively breastfed for six months or more. The following interventions were similarly evaluated: watching television for no more than one hour/day; eating at least five fruits and vegetables a day; playing at the playground every day; eliminating SSB consumption; and

eliminating fast-food consumption. We selected the desired levels of behaviors based on (i) national and international recommendations (e.g. from World Health Organization, American Academy of Pediatrics) regarding optimal child growth, (ii) plausible anticipated risk reduction documented in published literature, and (iii) available response categories used in the survey. A detailed description of the recommendations is available in the appendix (See **eTable 6.2** in appendix).

Covariates

We used a directed acyclic graph[74] to depict the hypothesized data-generating mechanism and causal structure of the processes under study (see **Figure 6.3**). In our first analytic model, we adjusted for child's baseline WHZ, and sociodemographic variables, namely, age at first relevant measurement, gender, race/ethnicity, birth weight, maternal language preference, maternal education, family size, family monthly income, maternal age and follow-up period (n=799). In Model 2, we further adjusted for maternal Body Mass Index (BMI), a potential confounder of the relationship between certain risk factors such as breastfeeding duration and childhood adiposity.[198] This analysis involved a smaller sample (n= 553) which excluded records with missing or improbable maternal BMI (BMI< 14 or BMI > 48). See **eFigure 6.2** in the appendix for a flow diagram showing sample sizes at various stages of participant inclusion.

Statistical analyses

We used the g-computation algorithm (applied to the parametric g-formula), a generalization of the standardization method for time-varying exposures and confounders,[91] to predict the potential mean WHZ under various hypothetical scenarios. We first fit linear

regression models of the outcome WHZ_3 and mediator WHZ_2 on behavioral risk factors adjusting for the selected covariates. We then used the regression coefficients obtained from these models to predict the potential outcomes and mediators under the different hypothetical interventions. We obtained the marginal mean differences (i.e. intervention impact) by taking the difference between the predicted potential mean WHZ under the various scenarios (in which the exposure distributions were altered so that 100% of the population would be exposed to the desired level of the risk factor) and the WHZ under no intervention (i.e. status quo) (in which the exposure distributions remained the same as in the original sample). Standard errors and 95% confidence intervals were obtained via bootstrapping. These steps are also described in the literature.[88,89] Analyses were conducted in SAS version 9.4 (SAS Institute, Inc., Cary, NC). It was assumed that there was: (i) no uncontrolled confounding after adjusting for the selected covariates, (ii) positivity, (iii) consistency, and (iv) no other source of bias.

We conducted sensitivity analyses to assess the robustness of our findings (1) under different sample restriction scenarios, and (2) when missing values and extreme values of WHZ and maternal BMI were imputed (see appendix **eTable 6.4**, **eTable 6.6**, **eTable 6.7**, **eTable 6.8** and **eTable 6.9**).

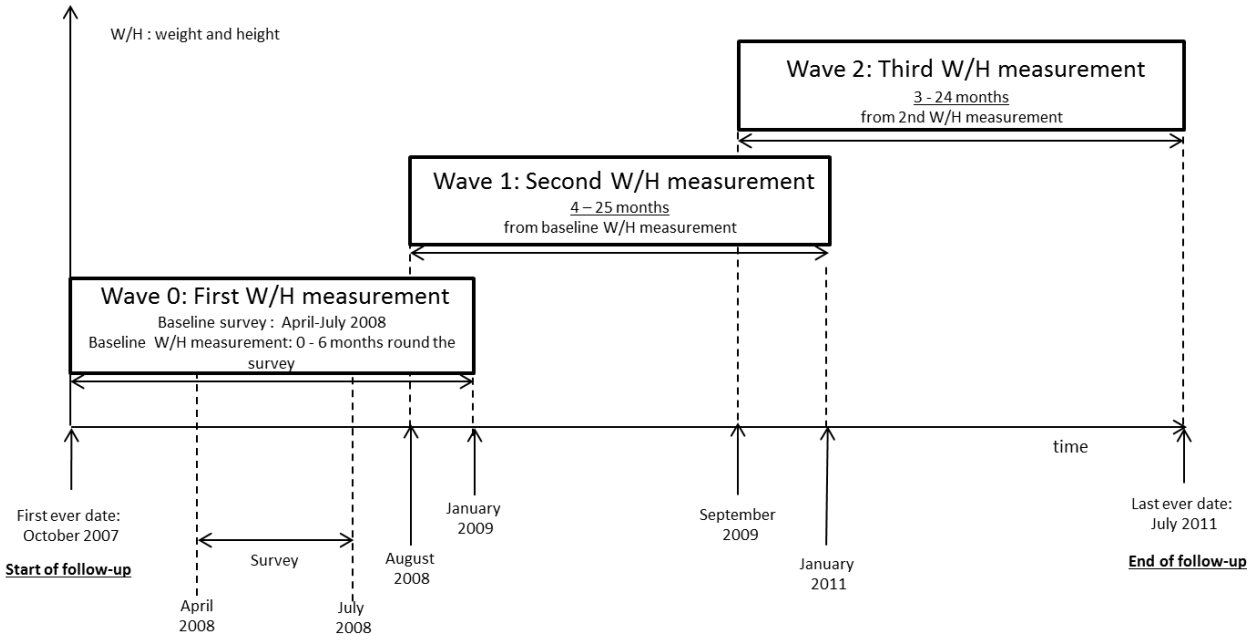


Figure 6.1 Cohort flow diagram outlining the timing of subsequent measurements by wave

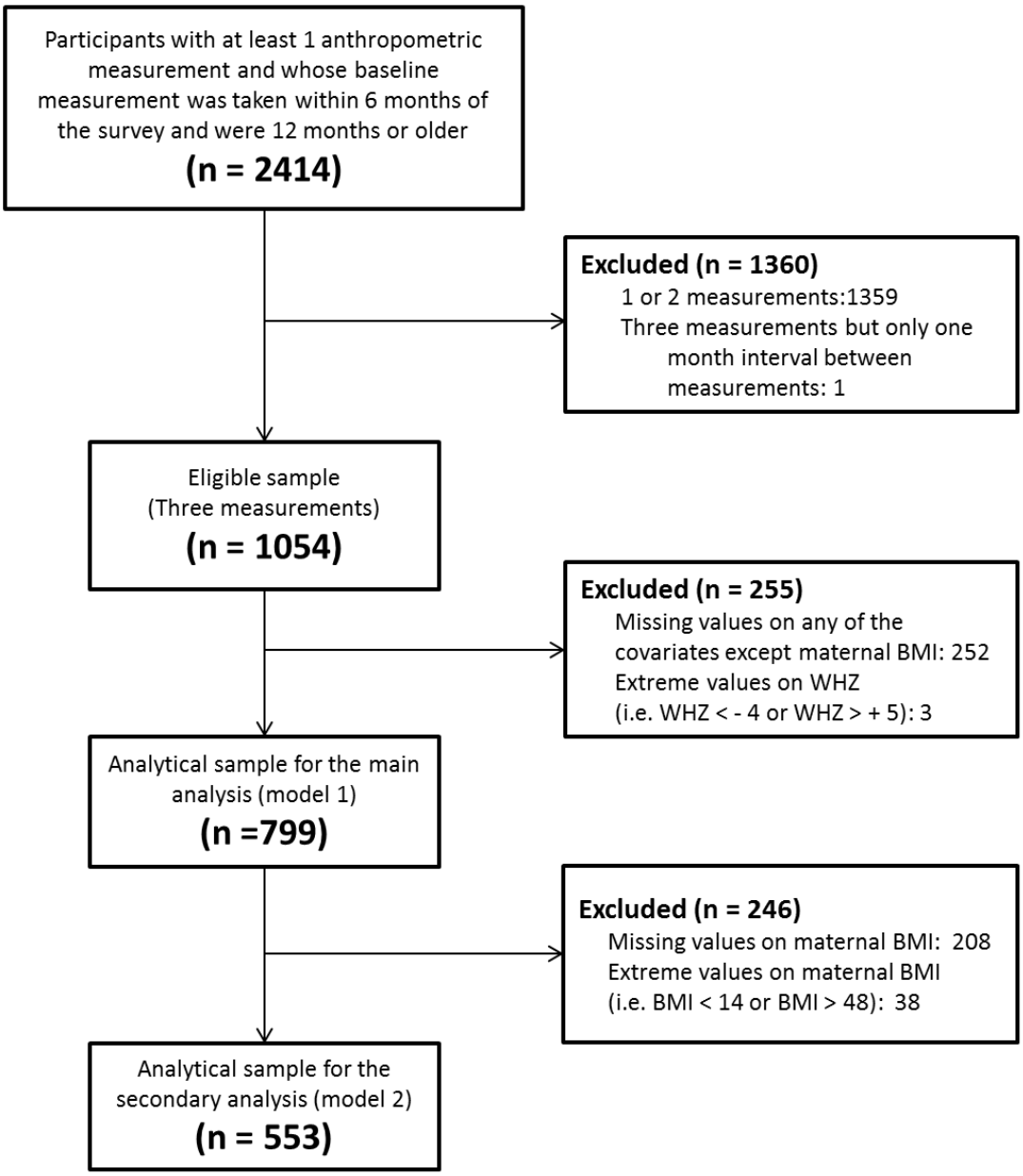


Figure 6.2 Study flow diagram showing the inclusion of participants in the final cohort WHZ: Weight-for-Height Z score. The authors included 1054 children in the cohort who met the inclusion criteria. Further restrictions due to missing values on assessed covariates yielded two analytic samples of smaller size.

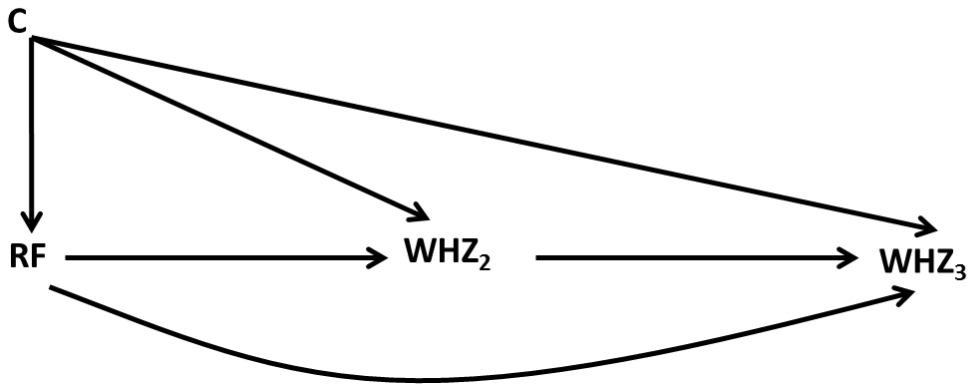


Figure 6.3 Hypothetical causal structure depicted using directed acyclic graph. RF, risk factors assessed in 2008 (breastfeeding duration, sugar-sweetened beverage consumption, fruits and vegetables, watching TV, use of parks) as exposures. C, Baseline covariates (baseline WHZ_1 , birth weight, gender, race/ethnicity, family income, family education, family size, family language preference, maternal age, maternal body mass index) acting as potential confounders. WHZ_2 , second WHZ measurement in 2009 (mediator). WHZ_3, third WHZ measurement WHZ in 2010 (outcome)

6.4. Results

Among the 1,054 children aged 1-5 years who had three measurements, 799 (76%) had complete data on all variables except maternal BMI and were included in the first analytic sample for the main analysis. Due to missing maternal values, analyses including maternal BMI as a covariate were conducted on the reduced analytic sample (n = 553 or 52% of the eligible sample) (See **Figure 6.2**). **Table 6.1** shows characteristics of the 799 children included in our main analysis. These children had a mean (SD) age of 23 (7) months at baseline; 65% were Hispanic. The median monthly family income was $1,545. Cohort members were followed, on average, for 23 months.

At baseline, one in five children was exclusively breastfed for six months or more; one-third watched television one hour or less (0 to 1h) per day; and more than half of the children consumed five or more servings of fruits and vegetables per day. Only 10% played in parks and playgrounds every day; two-thirds reported not consuming SSB; and 15% reported never consuming fast food (**eTable 6.3**).

For most interventions considered in this study, we needed to expose more than three quarters of the population in order for the entire population to be exposed to the desirable level of the behavioral factors (e.g. exclusive breastfeeding for six months or longer) (**Table 6.2**). The mean WHZ at the end of the follow-up was 0.73 (95%CI 0.65 to 0.81) under no intervention and 0.63 (0.38 to 0.87) when all interventions were imposed (Table 2). The most effective single intervention in this study was exclusive breastfeeding for six months or longer (population mean difference = -0.11, (95% CI -0.22 to 0.01) (**Figure 6.4**). The population mean difference for the other interventions were as follows: watching TV for no more than one hour a day: 0.00 (95% CI -0.10 to 0.09); eating at least five fruits and vegetables a day: 0.02 (95%CI -0.02, 0.06); eliminating SSB consumption: 0.01 (95%CI -0.03 to 0.06) and playing at the playground everyday: 0.01 (95%CI -0.14 to 0.18). Further adjusting for maternal BMI did not change the results in any substantive way (**Figure 6.4** and **Table 6.2**). Results from sensitivity analyses showed patterns similar to those reported in the main analyses (see appendix **eTable 6.6**, **eTable 6.7**, **eTable 6.8** and **eTable 6.9**).

Table 6.1 Baseline characteristics of WIC participants in the analytic sample, in 2008 (N = 799)

Baseline characteristics	Frequency (%)	Mean (SD)
Child's age in months		23 (7)
Follow-up period in months		23 (3)
Interval between measurements		11 (2)
Family size		4 (1)
Child's gender		
Male	398 (50)	
Female	401 (50)	
Child's ethnicity		
White	208 (26)	
Black	24 (3)	
Hispanic	519 (65)	
Asians and Others	48 (6)	
Child's birthweight in kg		4 (1)
Baseline WHZ		0.80 (1.15)
Maternal age		30 (7)
Maternal BMI in kg/m2 (n=553)		28 (5)
Family monthly income in $ US		1545 (766)
Family education		
High school or higher	282 (35)	
Less than high school	517 (65)	
Family language preference		
English	220 (28)	
Spanish	579 (72)	

SD: Standard deviation

Table 6.2 Mean WHZ score under hypothetical lifestyle interventions

	Interventions	Mean WHZ at the end of follow-up[a] (n = 799)	Mean WHZ at the end of follow-up[b] (n = 553)	Average percent intervened on (%)[c]
(0)	No intervention, natural course	0.73 (0.65 to 0.81)	0.71 (0.62 to 0.80)	0
(1)	Breastfeed exclusively for at least 6 months	0.62 (0.49 to 0.77)	0.63 (0.48 to 0.78)	77
(2)	Watch TV for no more than one hour a day	0.72 (0.61 to 0.84)	0.74 (0.60 to 0.87)	65
(3)	Eat at least five fruits and vegetables/day	0.75 (0.66 to 0.84)	0.71 (0.60 to 0.82)	38
(4)	Eliminate SSB consumption	0.74 (0.65 to 0.83)	0.73 (0.63 to 0.84)	35
(5)	Play at the playground everyday	0.74 (0.57 to 0.92)	0.67 (0.47 to 0.87)	90
(6)	Eliminate fast-food consumption	0.69 (0.52 to 0.86)	0.63 (0.40 to 0.87)	85
(7)	Low-risk lifestyle intervention (1- 3 combined)	0.64 (0.48 to 0.79)	0.65 (0.47 to 0.84)	94
(8)	Low-risk lifestyle intervention (1 + 4 + 6 combined)	0.60 (0.39 to 0.81)	0.56 (0.32 to 0.81)	97
(9)	All interventions (1 - 6 combined)	0.63 (0.38 to 0.87)	0.55 (0.25 to 0.85)	100

[a]Model 1 adjusted for baseline WHZ1, baseline age, gender, race/ethnicity, birthweight, maternal language preference, maternal educational level, maternal age, family size, family monthly income, follow-up time.

[b]Model 2 adjusted for model 1 covariates and maternal BMI

Note that the model for WHZ_3 further included interaction terms between race and WHZ_2 and between WHZ_1 and WHZ_2. No interaction was included in the WHZ_2 model

[c]Using the first analytic sample (n=799)

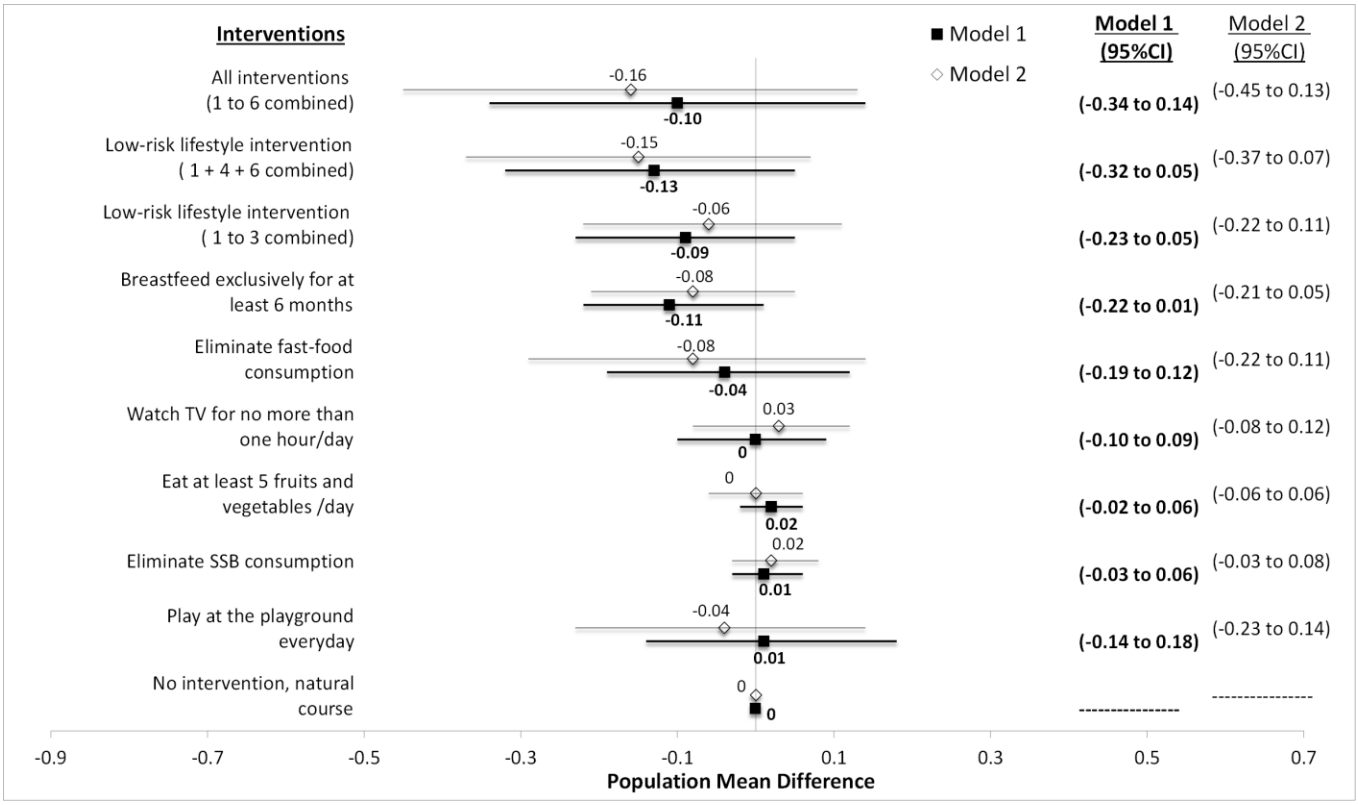


Figure 6.4 Forest-plot of the population impacts of hypothetical lifestyle interventions, WIC cohort, 2008-2010. TV, television viewing; Model 1 adjusted for baseline WHZ1, baseline age, gender, race/ethnicity, birthweight, maternal language preference, maternal educational level, maternal age, family size, family monthly income, follow-up time and Model 2 adjusted for model 1 covariates and maternal BMI.

6.5. Discussion

The purpose of this study was to evaluate the potential impact of hypothetical early behavioral interventions on childhood adiposity in a multi-ethnic and cohort of children aged 1-5 years living in low-income households. Using causal inference methods, we predicted WHZ at the end of follow-up under various hypothetical interventions and contrasted it to the status quo (no intervention) in order to estimate its potential population impact. Our findings suggest that a hypothetical intervention promoting exclusive breastfeeding for six months or longer, alone or in combination with other early behavioral interventions, may reduce a child's subsequent WHZ. The other early behavioral interventions evaluated singly in this study did not appear to have as much impact on a child's adiposity trajectory (through age 5 years) as breastfeeding alone did.

Breastfeeding is known to have many benefits. However, its role in obesity prevention is less established.[188,189] Systematic reviews and meta-analyses of observational have concluded that breastfeeding is associated with lower risk of childhood obesity,[198] and our findings are consistent with this conclusion. However, a RCT of a breastfeeding promotion intervention did not find intervention effects on adiposity measures.[199] While this was an impressive effort involving 31 hospitals and clinics and over 15,000 infants, the study took place in Belarus where obesity prevalence is relatively low. Furthermore, the analysis was based on intention-to-treat.

Our findings derived using causal inference methods, support the conclusion that breastfeeding may protect against obesity development in early childhood. Several possible biological mechanisms can explain this protective effect. First, breast milk provides a moderate amount of calories and protein as compared to formula feeding;[200] higher early intakes of protein have been shown to be associated with later adiposity.[201] Second, breast milk is also rich in factors such as leptin, which regulate satiety and subsequent growth and development.[202]

Third, it has been suggested that breastfed children may adapt better to new foods compared to formula-fed children.[203]All these mechanisms may also explain why longer duration of breastfeeding, as recommended by the WHO,[204] may help reduce the risk of developing obesity.

In this study, contrary to our expectations, watching television for no more than one hour/day, eating at least five fruits and vegetables a day, playing at the playground every day, eliminating SSB consumption and eliminating fast-food consumption, evaluated singly did not have as much impact on the child's adiposity trajectory through age 5 as exclusive breastfeeding. This is somewhat surprising as interventions developed to mitigate most of these risk factors have been observed to lower obesity risk among exposed children.[175] One reason why we failed to find an effect could be that our study focused on much younger children than those investigated in most other studies; young children are less likely to engage in vigorous physical activity and eat fast-food than older children. Another reason is that the effects of interventions are more easily detected when there is considerable variation in the behaviors studied. Alternatively, our findings may have merely reflected beneficial effects on growth that could not be detected so early in life. This has also been seen in some RCTs where interventions on parents to promote healthy behavior among children seemed to have little or no effect on childhood obesity risk.[205] Regardless, these behavioral interventions are still warranted for their potential long-term benefits on overall health and well-being.

This study has several limitations. First, we did not have measurements on certain prenatal factors such as smoking during pregnancy and gestational diabetes which are often considered potential confounders of the association between breastfeeding (and other behavioral factors) and childhood obesity.[198,206]. Nonetheless, our current covariate adjustment may have minimized this residual confounding since some measured covariates such as maternal education

and age are also predictive of these unmeasured factors.[207,208] Second, we did not adjust for the child's energy intake from solid foods and for parental feeding practices. Third, as can be expected of observational studies, our findings could have been affected by reporting bias and social desirability. However, the magnitude of such bias, if present, would likely be small in this study since a multi-item indirect questioning approach rather than a binary response approach was used to gather relevant information[209] [for example, "How old was the child, the first time (he/she) ate anything besides breast milk?" rather than "Did you breastfeed?"]. Fourth, because eligible participants had to have three consecutive measurements, they were more likely to be younger children since WIC serves children up to only age 5 years. Also, the sample consisted of a high percentage of Spanish-speaking Hispanics who are more likely to stay in WIC longer.[111] Therefore, our results are more generalizable to younger Hispanic children with Spanish-speaking mothers.

The strengths of this study include its longitudinal nature, the relatively large and ethnically diverse sample, the use of causal inference methods, the assessment of multiple behavioral interventions, the use of measured validated heights and weights, the use of WHZ as an adiposity indicator, and various sensitivity analyses. To our knowledge, this is the first attempt to use the parametric g-formula[91] to infer population-level effects of breastfeeding on obesity using individual-level effect estimates. However, it is important to note that while the findings of this study contribute to our collective effort to better understand the role of breastfeeding in obesity development during childhood, they simply provide an estimate of the impact of a breastfeeding intervention in the hypothetical scenario when women exclusively breastfeed for 6 months.

Randomized trials are not always feasible or are difficult to implement, and while they are considered the "gold standard" research design for evaluating community health interventions, they are limited in their applications in real life. This study illustrates the use of the g-computation formula, a more practical and cost effective alternative for examining the controversial role of breastfeeding in reducing childhood obesity risk. Our findings suggest that efforts to promote exclusive breastfeeding in combination with other lifestyle interventions may prove to be an effective strategy for preventing obesity later in life among minority populations and those living in poverty. It is hoped that this study will stimulate further foray into the use of modern causal reasoning and simulation methods for addressing crucial policy questions relevant to obesity and its public health consequences.[139]

6.6. Appendix

eTable 6.1 List of all relevant early life nutrition/questions asked on the 2008 survey

Survey questions	Response options
Breastfeeding duration	
How old was your child the first time he or she ate anything besides breast milk? This includes formula, baby food, cow milk, sugar water, or anything else you fed your infant.	Less than1 week 1 week but less than 1 months 1 months but less than 3 months 3 months but less than 6 months At 6 months 6 months More than 6 months (volunteered) Don't know Refused
Sugar-sweetened beverage consumption	
On an average day, about how many sodas, such as Coke or Mountain Dew, or sweetened drinks, such as Gatorade, Red Bull, or Sunny Delight, does your child drink? (Do not include diet sodas or sugar-free drinks. Please count a 12-oz can, bottle, or glass as 1 drink.) (IF NECESSARY, SAY: Just your best estimate.)	_____drinks per day None/never Don't know Refused
Fruits and vegetables consumption	
On an average day, about how many servings of fruits does NAME eat? (IF NECESSARY, SAY: Just your best estimate.)	_____ fruits per day None/never Don't know Refused

eTable 6.1 List of all relevant early life nutrition/feeding questions asked on the 2008 survey, continued

Survey questions	Response options
On an average day, about how many servings of vegetables does NAME eat? (IF NECESSARY, SAY: Just your best estimate.)	_____ vegetables per day None/never Don't know Refused
Fast-food consumption How often does NAME eat any food including meals and snacks from a fast food restaurant, like McDonald's, Taco Bell, Burger King, Kentucky Fried Chicken, or another similar place? (READ CATEGORIES)	4+ times per week 1-3 times per week less than once a week but at least once a month less than once a month or- never Don't know Refused
Frequency of television viewing On an average day, how many hours does NAME watch television? Only include time when (he) (she) is sitting and watching TV.	_____ Hours Less than 1 hour Don't know Refused
Frequency of physical activity How many days in a typical week do you take NAME to a nearby park or playground to play – every day, 3 to 6 days,	Every day 3-6 days 1-2 days Never Don't know Refused

eTable 6.2 Prevention recommendations and rationale for choosing desired level of the interventions evaluated in this study

	Risk factor/behaviors	Prevention recommendations	Source	Intervention evaluated in our study and special notes
1	**Breastfeeding duration**	Encouraging exclusive breastfeeding to 6 months of age and maintenance of breastfeeding after introduction of solid food to 12 months of age and beyond	The American Academy of Pediatrics, committee on nutrition[210] Expert committee[211] *World Health Organization*[204]	Breastfeed exclusively for at least 6 months
2	**TV viewing**	Limiting television and other screen time (no TV viewing for children before 2 years and thereafter no more than 2 hours of TV viewing per day) with no television in the room where the child sleeps (CE)	*American Pediatrics Academy-committee on Public Education*[212] Expert committee[211]	Watch TV for no more than one hour a day
3	**Fruit and vegetable consumption**	Encourage children to eat five or more servings of fruits and vegetables each day. Families may subsequently increase to 9 servings per day, as recommended by the USDA according to age, ranging from 2 cups per day for 2-year-old children to 4.5 cups per day for 17- and 18-year-old youths; (ME)	Expert committee[211]	Eat at least five fruits and vegetables a day Note that *Wang et al* in their review reported that "*There was a threshold around five servings of fruit and vegetables a day, after which the risk of all-cause mortality did not reduce further*" suggesting that eating five fruits and vegetables could offer potential health benefits [213]
4	**Sugar sweetened beverage consumption**	Minimize or eliminate sugar-sweetened beverages (ME)	Expert committee[211] The American Academy of Pediatrics, committee on nutrition[210]	Eliminate sugar-sweetened beverage consumption

eTable 6. 2 Prevention recommendations and rationale for choosing desired level of the interventions evaluated in this study (continued)

	Risk factor/behaviors	Prevention recommendations	Source	Intervention evaluated in our study and special notes
5	**Physical activity**	Promoting moderate to vigorous physical activity for at least 60 minutes each day and promoting active play and lifestyle	Expert committee[211] The American Academy of Pediatrics, committee on nutrition[210]	Play at the playground everyday *Note that the question assessing physical activity behavior (Table S1) did not ask for the duration and/or intensity of physical activity but rather for frequency of playing in the playground (i.e. Every day, 3-6 days, 1 to 2 days, Never). Therefore we used the most frequent physical activity pattern as the desired level to represent the healthiest option.*
6	**Fast-food consumption**	Limiting consumption of energy-dense foods as well as limiting eating out at restaurants, particularly fast food restaurants (CE)	Expert committee[211] The American Academy of Pediatrics, committee on nutrition[210]	Eliminate fast-food consumption *Note that the question assessing fast-food consumption (Table S1) did not ask for the amount of fast-food consumed but rather for frequency of eating at a fast food restaurant (i.e. values ranging from Never to ≥ 4 times a month). Therefore we used the least frequent pattern (i.e. never) of eating at a fast-food restaurant to represent the healthiest option*

USDA—US Department of Agriculture; CE—consistent evidence; ME—mixed evidence
Note: In some cases, the obesity prevention recommendations do not specifically suggest a desired level of the behavior. In such cases we provided special notes to further justify the chosen desired level in this study.

eTable 6.3 Behavioral risk factors of WIC participants in the analytic sample at baseline (N = 799)

Risk factors	Frequency (%)
Exclusive breastfeeding duration	
Not breastfed to < 1 week	291 (36)
1 week to < 3 months	109 (14)
3 months to < 6 months	218 (27)
6 months or more	181 (23)
Television viewing	
≥ 3 hours/day	83 (10)
2 hours/day	192 (24)
1hour/day	245 (31)
< 1hour/day	279 (35)
Fruit and vegetable consumption	
1 to 2 serving/day	59 (7)
3 to 4 serving/day	241 (30)
≥ 5 servings/day	499 (62)
Use of park and playgrounds	
Never	51 (6)
1-2 days	467 (58)
3-6 days	199 (25)
Every day	82 (10)
Sugar-sweetened beverage consumption	
≥ 3 servings /day	43 (5)
2 serving /day	65 (8)
1 serving/day	175 (22)
0 serving /day	516 (65)
Fast food consumption	
1-4 times/week	364 (46)
< 1 time/week and ≥ 1 time/month	270 (34)
< 1/month	48 (6)
Never	117 (15)

Sensitivity analyses

We conducted sensitivity analyses to assess the robustness of our findings. First, we considered two other potential scenarios wherein we further restricted the timing of subsequent measurements. In one of these two scenarios, the subsequent measurement must have been taken at least 6 months and at most 18 months after the prior measurement (n= 996). In the other scenario, the subsequent measurement must have been taken at least 9 months and at most 15 months after the prior measurement (n= 710). For the main analysis, we did not restrict the timing of the subsequent measurement other than require that the measurements had to be at least three months apart (n =1054). In all three scenarios, to address missingness, we imputed missing values and extreme values (i.e. maternal BMI). We imputed 10 datasets for each scenario using the full conditional specification (FCS) option of the SAS PROC MI procedure.

eTable 6.4 Variables used in the imputation model

	Variables used in imputation model	**No-restriction (n=1054) (%)**	**6-18 months (n = 996) (%)**	**9-15 months (n = 710) (%)**
Imputed variables	Maternal BMI in kg/m2	27	28	29
	Breast feeding duration	16	15	15
	Child's birthweight in Kg	6	6	6
	Fruits and vegetables consumption	3	3	3
	Fast-food consumption	1	1	1
	TV watching	1	1	1
	Physical activity	1	1	1
	Sugar-sweetened-beverage consumption	1	1	1
	Race	0*	0*	0*
	Family Language preference	0*	0*	0*
Complete variables	WHZ_1	0	0	0
	WHZ_2	0	0	0
	WHZ_3	0	0	0
	Maternal age	0	0	0
	Child's age at baseline	0	0	0
	Gender	0	0	0
	Family Education	0	0	0
	Family size	0	0	0
	Family Income	0	0	0

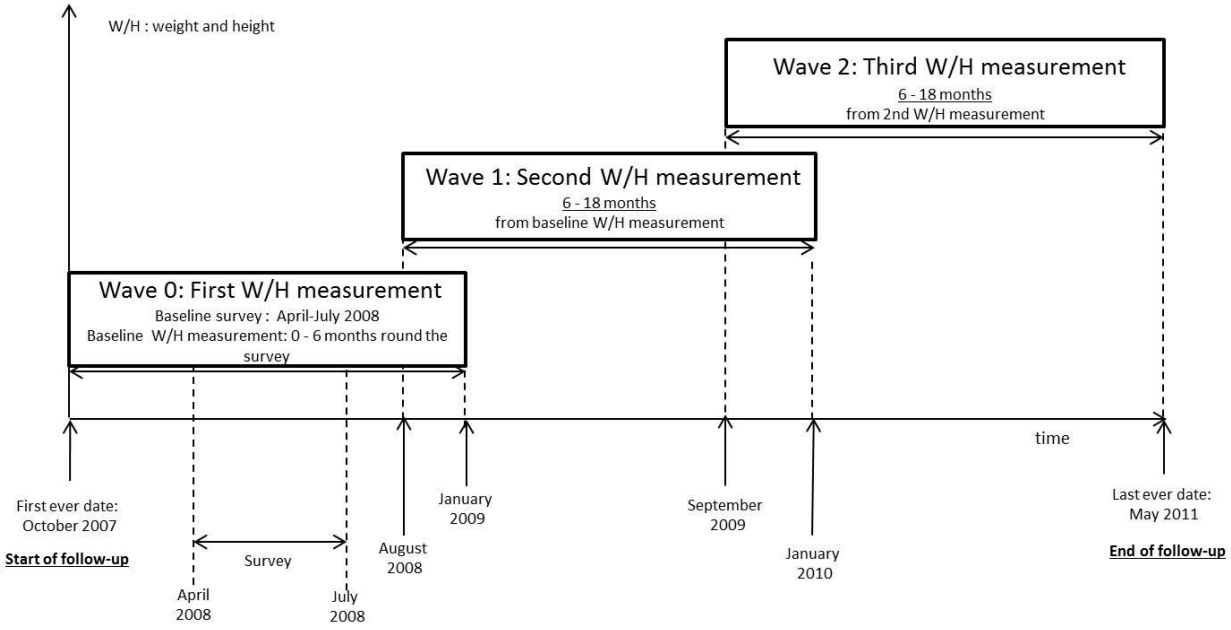
*Has some missing values

eTable 6.5 Mean WHZ score and population mean difference under hypothetical lifestyle interventions among WIC participants using the imputed dataset (n = 1054) (scenario 1 described in main manuscript)

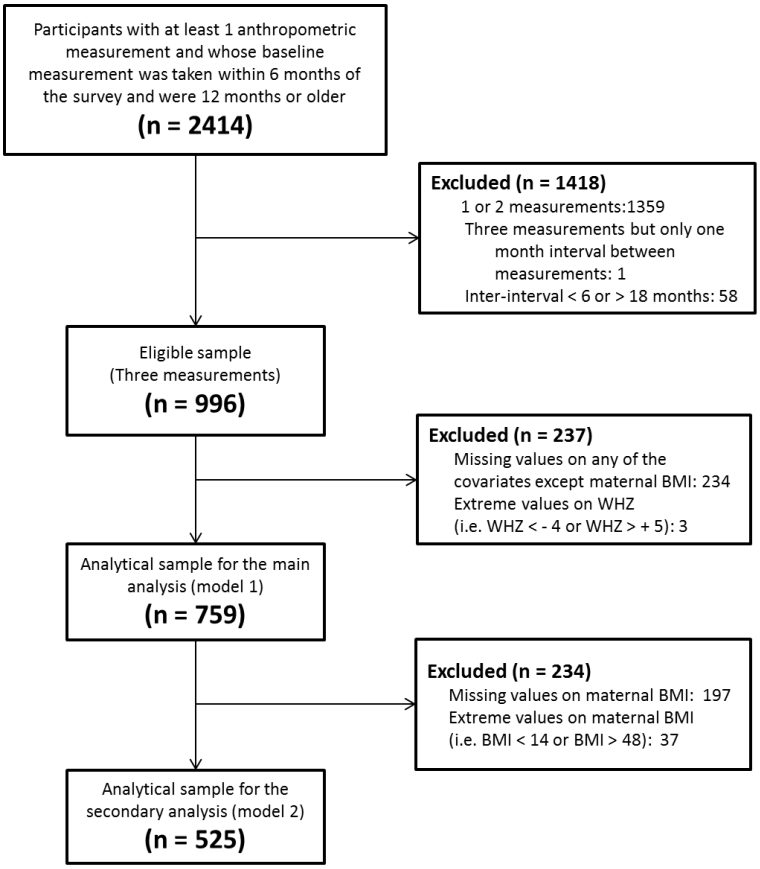
	Interventions	Model 1		Model 2*	
		Mean WHZ at the end of follow-up	**Population Mean Difference (95% CI)**	**Mean WHZ at the end of follow-up**	**Population Mean Difference (95% CI)**
(0)	No intervention, natural course	0.72 (0.70 to 0.74)	0	0.72 (0.70 to 0.74)	0
(1)	Breastfeed exclusively for at least 6 months	0.64 (0.60 to 0.67)	-0.08 (-0.11 to -0.05)	0.64 (0.60 to 0.67)	-0.08 (-0.11 to -0.05)
(2)	Watch TV for no more than one hour a day	0.7 (0.66 to 0.73)	-0.02 (-0.04 to 0.01)	0.70 (0.66 to 0.73)	-0.02 (-0.04 to 0.01)
(3)	Eat at least five fruits and vegetables/day	0.75 (0.72 to 0.77)	0.03 (0.02 to 0.04)	0.75 (0.72 to 0.77)	0.03 (0.02 to 0.04)
(4)	Eliminate SSB consumption	0.71 (0.68 to 0.73)	-0.01 (-0.02 to 0.00)	0.71 (0.68 to 0.73)	-0.01 (-0.02 to 0.00)
(5)	Play at the playground everyday	0.72 (0.67 to 0.76)	0.00 (-0.04 to 0.05)	0.72 (0.67 to 0.76)	0.00 (-0.04 to 0.05)
(6)	Eliminate fast-food consumption	0.71 (0.66 to 0.75)	-0.01 (-0.06 to 0.03)	0.71 (0.66 to 0.75)	-0.01 (-0.06 to 0.03)
(7)	Low-risk lifestyle intervention (1- 3 combined)	0.65 (0.60 to 0.69)	-0.07 (-0.11 to -0.03)	0.65 (0.60 to 0.69)	-0.07 (-0.11 to -0.03)
(8)	Low-risk lifestyle intervention (1 + 4 + 6 combined)	0.62 (0.56 to 0.67)	-0.10 (-0.15 to -0.05)	0.62 (0.56 to 0.67)	-0.10 (-0.15 to -0.05)
(9)	All interventions (1 - 6 combined)	0.63 (0.56 to 0.70)	-0.09 (-0.15 to -0.02)	0.63 (0.56 to 0.70)	-0.09 (-0.15 to -0.02)

* **Further adjusting for maternal BMI**

Scenario 2:



eFigure 6.1 Cohort flow diagram outlining the timing of subsequent measurement by wave



eFigure 6.2 Study flow diagram showing the inclusion of participants in the final cohort

eTable 6.6 Population mean difference under hypothetical lifestyle interventions among WIC participants using the 6-18 month interval restriction sample (eligible children n = 996)

	Interventions	Model 1 (n = 759)		Model 2* (n = 525)	
		Mean WHZ at the end of follow-up	**Population Mean Difference (95% CI)**	**Mean WHZ at the end of follow-up**	**Population Mean Difference (95% CI)**
(0)	No intervention, natural course	0.71 (0.64 to 0.79)	0	0.69 (0.59 to 0.79)	0
(1)	Breastfeed exclusively for at least 6 months	0.60 (0.46 to 0.74)	-0.11 (-0.23 to 0.00)	0.61 (0.45 to 0.77)	-0.08 (-0.22 to 0.05)
(2)	Watch TV for no more than one hour a day	0.70 (0.57 to 0.82)	-0.01 (-0.11 to 0.08)	0.71 (0.57 to 0.84)	0.01 (-0.09 to 0.12)
(3)	Eat at least five fruits and vegetables/day	0.73 (0.64 to 0.83)	0.02 (-0.03 to 0.07)	0.69 (0.57 to 0.80)	0.00 (-0.06 to 0.05)
(4)	Eliminate SSB consumption	0.73 (0.65 to 0.82)	0.02 (-0.03 to 0.07)	0.72 (0.61 to 0.83)	0.03 (-0.02 to 0.08)
(5)	Play at the playground everyday	0.74 (0.56 to 0.92)	0.03 (-0.13 to 0.20)	0.67 (0.45 to 0.88)	-0.03 (-0.23 to 0.17)
(6)	Eliminate fast-food consumption	0.66 (0.48 to 0.83)	-0.05 (-0.22 to 0.11)	0.57 (0.35 to 0.81)	-0.12 (-0.33 to 0.10)
(7)	Low-risk lifestyle intervention (1- 3 combined)	0.61 (0.44 to 0.78)	-0.10 (-0.25 to 0.05)	0.62 (0.43 to 0.80)	-0.07 (-0.24 to 0.09)
(8)	Low-risk lifestyle intervention (1 + 4 + 6 combined)	0.57 (0.36 to 0.78)	-0.14 (-0.34 to 0.07)	0.52 (0.27 to 0.78)	-0.17 (-0.42 to 0.08)
(9)	All interventions (1 - 6 combined)	0.61 (0.34 to 0.87)	-0.10 (-0.36 to 0.16)	0.50 (0.18 to 0.83)	-0.19 (-0.51 to 0.13)

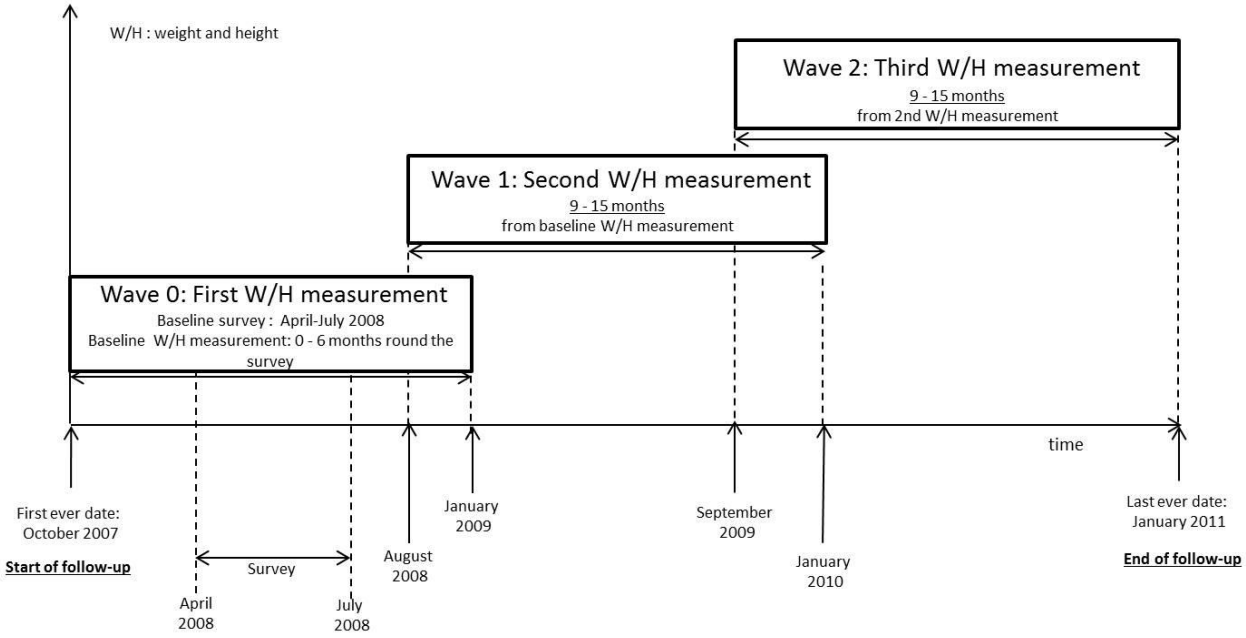
* **Further adjusting for maternal BMI**

eTable 6.7 Population mean difference under hypothetical lifestyle interventions among WIC participants using the imputed dataset with 6-18 month interval restriction (n = 996)

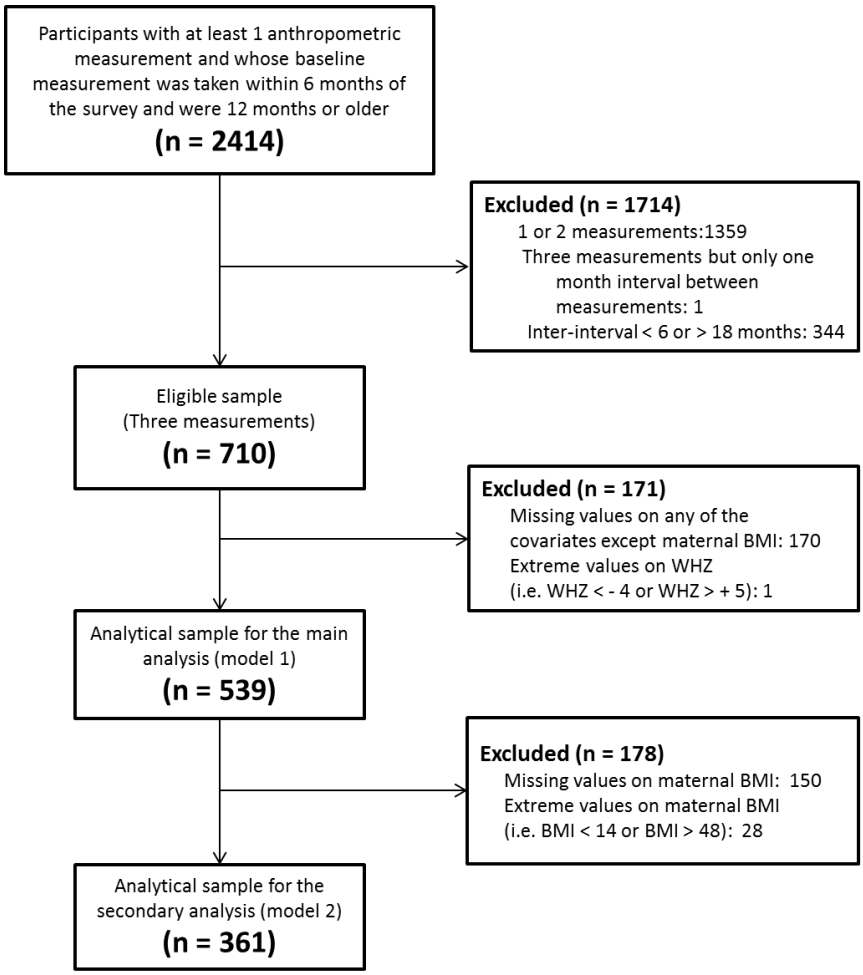
	Interventions	Model 1		Model 2*	
		Mean WHZ at the end of follow-up	**Population Mean Difference (95% CI)**	**Mean WHZ at the end of follow-up**	**Population Mean Difference (95% CI)**
(0)	No intervention, natural course	0.71 (0.68 to 0.73)	0	0.71 (0.68 to 0.73)	0
(1)	Breastfeed exclusively for at least 6 months	0.63 (0.60 to 0.67)	-0.07 (-0.10 to -0.04)	0.63 (0.60 to 0.67)	-0.07 (-0.10 to -0.04)
(2)	Watch TV for no more than one hour a day	0.67 (0.63 to 0.70)	-0.04 (-0.06 to -0.01)	0.67 (0.63 to 0.70)	-0.04 (-0.07 to -0.01)
(3)	Eat at least five fruits and vegetables/day	0.73 (0.71 to 0.76)	0.03 (0.02 to 0.04)	0.73 (0.71 to 0.76)	0.03 (0.02 to 0.04)
(4)	Eliminate SSB consumption	0.71 (0.68 to 0.73)	0.00 (-0.01 to 0.01)	0.71 (0.68 to 0.73)	0.00 (-0.01 to 0.01)
(5)	Play at the playground everyday	0.74 (0.69 to 0.79)	0.03 (-0.01 to 0.08)	0.74 (0.69 to 0.79)	0.03 (-0.01 to 0.08)
(6)	Eliminate fast-food consumption	0.68 (0.63 to 0.73)	-0.02 (-0.07 to 0.02)	0.68 (0.63 to 0.73)	-0.02 (-0.07 to 0.02)
(7)	Low-risk lifestyle intervention (1- 3 combined)	0.62 (0.58 to 0.67)	-0.08 (-0.12 to -0.04)	0.62 (0.58 to 0.67)	-0.08 (-0.12 to -0.04)
(8)	Low-risk lifestyle intervention (1 + 4 + 6 combined)	0.61 (0.55 to 0.67)	-0.09 (-0.15 to -0.04)	0.61 (0.55 to 0.67)	-0.09 (-0.15 to -0.04)
(9)	All interventions (1 - 6 combined)	0.63 (0.56 to 0.71)	-0.07 (-0.14 to 0.00)	0.63 (0.56 to 0.71)	-0.07 (-0.14 to 0.00)

*** Further adjusting for maternal BMI**

Scenario 3:



eFigure 6.3 Cohort flow diagram outlining the timing of subsequent measurement by wave



eFigure 6.4 Study flow diagram showing the inclusion of participants in the final cohort

eTable 6.8 Population mean difference under hypothetical lifestyle interventions among the WIC participants one-year-or-older-children cohort and using the 9-15 month interval restriction sample (eligible children n = 710)

	Interventions	Model 1 (n = 539)		Model 2* (n = 361)	
		Mean WHZ at the end of follow-up	**Population Mean Difference (95% CI)**	**Mean WHZ at the end of follow-up**	**Population Mean Difference (95% CI)**
(0)	No intervention, natural course	0.73 (0.64 to 0.82)	0	0.70 (0.59 to 0.83)	0
(1)	Breastfeed exclusively for at least 6 months	0.64 (0.49 to 0.81)	-0.09 (-0.22 to 0.05)	0.64 (0.45 to 0.84)	-0.06 (-0.23 to 0.11)
(2)	Watch TV for no more than one hour a day	0.69 (0.55 to 0.84)	-0.04 (-0.14 to 0.06)	0.72 (0.57 to 0.88)	0.02 (-0.11 to 0.15)
(3)	Eat at least five fruits and vegetables/day	0.75 (0.64 to 0.86)	0.02 (-0.04 to 0.08)	0.72 (0.60 to 0.86)	0.02 (-0.05 to 0.10)
(4)	Eliminate SSB consumption	0.74 (0.64 to 0.85)	0.02 (-0.04 to 0.07)	0.72 (0.59 to 0.85)	0.01 (-0.05 to 0.07)
(5)	Play at the playground everyday	0.80 (0.56 to 1.01)	0.07 (-0.14 to 0.27)	0.71 (0.44 to 0.96)	0.00 (-0.24 to 0.22)
(6)	Eliminate fast-food consumption	0.72 (0.49 to 0.95)	-0.01 (-0.21 to 0.19)	0.58 (0.29 to 0.87)	-0.12 (-0.39 to 0.16)
(7)	Low-risk lifestyle intervention (1- 3 combined)	0.63 (0.45 to 0.81)	-0.10 (-0.26 to 0.06)	0.68 (0.47 to 0.90)	-0.03 (-0.23 to 0.18)
(8)	Low-risk lifestyle intervention (1 + 4 + 6 combined)	0.65 (0.40 to 0.91)	-0.08 (-0.32 to 0.14)	0.53 (0.20 to 0.88)	-0.17 (-0.49 to 0.15)
(9)	All interventions (1 - 6 combined)	0.7 (0.39 to 1.01)	-0.03 (-0.34 to 0.28)	0.57 (0.17 to 0.99)	-0.14 (-0.52 to 0.28)

* **Further adjusting for maternal BMI**

eTable 6.9 Population mean difference under hypothetical lifestyle interventions among the WIC participants using the imputed dataset with the 9-15 month interval restriction (n = 710)

	Interventions	Model 1		Model 2*	
		Mean WHZ at the end of follow-up	**Population Mean Difference (95% CI)**	**Mean WHZ at the end of follow-up**	**Population Mean Difference (95% CI)**
(0)	No intervention, natural course	0.71 (0.69 to 0.74)	0	0.71 (0.69 to 0.74)	0
(1)	Breastfeed exclusively for at least 6 months	0.66 (0.62 to 0.71)	-0.05 (-0.09 to -0.01)	0.66 (0.62 to 0.71)	-0.05 (-0.09 to -0.01)
(2)	Watch TV for no more than one hour a day	0.67 (0.63 to 0.71)	-0.05 (-0.07 to -0.02)	0.67 (0.63 to 0.71)	-0.05 (-0.08 to -0.02)
(3)	Eat at least five fruits and vegetables/day	0.74 (0.71 to 0.77)	0.03 (0.01 to 0.05)	0.74 (0.71 to 0.77)	0.03 (0.01 to 0.05)
(4)	Eliminate SSB consumption	0.71 (0.69 to 0.74)	0.00 (-0.01 to 0.01)	0.71 (0.69 to 0.74)	0.00 (-0.01 to 0.01)
(5)	Play at the playground everyday	0.77 (0.71 to 0.83)	0.06 (0.00 to 0.12)	0.77 (0.71 to 0.83)	0.06 (0.00 to 0.12)
(6)	Eliminate fast-food consumption	0.70 (0.64 to 0.75)	-0.02 (-0.07 to 0.04)	0.70 (0.64 to 0.75)	-0.02 (-0.07 to 0.04)
(7)	Low-risk lifestyle intervention (1- 3 combined)	0.65 (0.59 to 0.70)	-0.07 (-0.12 to -0.02)	0.65 (0.59 to 0.70)	-0.07 (-0.11 to -0.02)
(8)	Low-risk lifestyle intervention (1 + 4 + 6 combined)	0.65 (0.59 to 0.71)	-0.07 (-0.12 to -0.01)	0.65 (0.59 to 0.71)	-0.07 (-0.12 to -0.01)
(9)	All interventions (1 - 6 combined)	0.69 (0.60 to 0.78)	-0.03 (-0.11 to 0.06)	0.69 (0.60 to 0.78)	-0.03 (-0.11 to 0.06)

* **Further adjusting for maternal BMI**

Chapter 7. General Discussion and Concluding Remarks

This dissertation endeavor offered us with an eminent topical and methodological opportunity to advance the state of science in obesity research. From a methodological perspective, this research study stemmed from the desire to answer highly relevant questions that were seldom answerable in the current scientific paradigm. In the current paradigm one is often confined to one particular dataset, one exposure/risk factor and one outcome. This new paradigm, however, calls for the integration and simulation of new data via systems science methods such as agent-based modeling. In this research, we integrated and synthesized the best available knowledge in order the have a clearer picture of obesity and diabetes in Los Angeles County. It was akin to finding and putting the pieces of a puzzle together in order to have a clearer image. This step is critical as it helps us reason about and compute if necessary the correct parameters. Once all the best data have been gathered and synthesized, we integrated them into models that encoded our understanding of the process under study. Secondly, we used the parameters and models to create a virtual world and laboratory that resembles reality. The researchers and policymakers can then experiment *in silico* potential new interventions by simulating counterfactual scenarios. In so doing, we can assess the potential efficacy and harm of candidate interventions and as a result avoid spending money on interventions that will prove unsuccessful in the population as a whole and in the long run. Such impact can then be evaluated with the help of rigorous methods that are also flexible, namely, the g-computation algorithm.[126]

From a topical and applied perspective, this dissertation offered many tangible contributions. First, this study sheds light on the mechanisms and underlying causal pathways behind the obesity and diabetes epidemic. It also sheds light on the risk factors that contributed to sustaining the epidemic and those that gave rise to health disparities in obesity and diabetes. Second, this

study provided policymakers with a better understanding of the current situation in light of the past and the current knowledge. This is important as it can help government officials better prepare for an eventual health care need. In the same way, many researchers have attempted to forecast the future burden of obesity and type 2 diabetes but very few have been able to do so within a single birth cohort. This is critical because most other forecasting endeavors do not take into account the specificity at the individual level and so fall short when it comes to designing interventions that are specific to different groups of people. Third, we foresaw that given the current prevalence of obesity and unhealthy behaviors across age groups, the incidence and prevalence of obesity and type 2 diabetes will continue to increase with age during an individual's life course. Fourth, since childhood obesity remains an independent risk factor of type 2 diabetes, one should start the prevention of type 2 diabetes of the latter should start early in childhood. Fifth, another intuition that arose was that treating the population as a whole may as some have suggested[180] prevented more burden than focusing on high-risk populations. Sixth, most health interventions aimed at halting the obesity or diabetes epidemic may yield modest to null effects if not sustained over time or done in combination with one another. Lastly, from this study, it appeared that spending more efforts to get people to breastfeed for at least six months in early childhood, exercise more, avoid fast-food and sugar-sweetened beverage consumption can greatly reduce the risk of obesity and type 2 diabetes.

References

1. World Health Organization (2012) Obesity and overweight. *Fact sheet N°311*. http://www.who.int/mediacentre/factsheets/fs311/en/.
2. Flegal KM, Carroll MD, Kit BK, et al. (2012) Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *JAMA* **307**, 491–7.
3. Ogden CL, Carroll MD, Kit BK, et al. (2014) Prevalence of childhood and adult obesity in the United States, 2011-2012. *JAMA* **311**, 806–14. American Medical Association.
4. Wang LY, Chyen D, Lee S, et al. (2008) The Association Between Body Mass Index in Adolescence and Obesity in Adulthood. *J. Adolesc. Heal.* **42**, 512–518.
5. Trasande L & Chatterjee S (2009) The impact of obesity on health service utilization and costs in childhood. *Obesity (Silver Spring).* **17**, 1749–1754.
6. Wang G & Dietz WH (2002) Economic burden of obesity in youths aged 6 to 17 years: 1979-1999. *Pediatrics* **109**, E81–E81.
7. Juonala M, Magnussen CG, Berenson GS, et al. (2011) Childhood adiposity, adult adiposity, and cardiovascular risk factors. *N. Engl. J. Med.* **365**, 1876–85.
8. Goran MI & Treuth MS (2001) ENERGY EXPENDITURE, PHYSICAL ACTIVITY, AND OBESITY IN CHILDREN. *Pediatr. Clin. North Am.*, 931–953.
9. Daniels SR, Jacobson MS, McCrindle BW, et al. (2009) American heart association childhood obesity: Research summit. In *Circulation*, vol. 119, pp. 2114–2123.
10. Prince SA, Kristjansson EA, Russell K, et al. (2011) A multilevel analysis of neighbourhood built and social environments and adult self-reported physical activity and body mass index in Ottawa, Canada. *Int. J. Environ. Res. Public Health* **8**, 3953–78.
11. Schonfeld-Warden N & Warden CH (1997) Pediatric obesity. An overview of etiology

and treatment. *Pediatr. Clin. North Am.* **44**, 339–361.

12. National Heart Lung and Blood Institute (2007) *Working Group Report on Future Research Directions in Childhood Obesity Prevention and Treatment*. Washington, DC: .
13. Huang TT, Drewnoksi A, Kumanyika S, et al. (2009) A systems-oriented multilevel framework for addressing obesity in the 21st century. *Prev. Chronic Dis.*, A82.
14. Valladares-Salgado A, Suárez-Sánchez F, Burguete-García AI, et al. (2014) [Epigenetics of childhood obesity and diabetes]. *Rev. Med. Inst. Mex. Seguro Soc.* **52 Suppl 1**, 88–93.
15. Daniels SR, Arnett DK, Eckel RH, et al. (2005) Overweight in children and adolescents: Pathophysiology, consequences, prevention, and treatment. *Circulation*, 1999–2012.
16. Newbold RR, Padilla-Banks E, Jefferson WN, et al. (2008) Effects of endocrine disruptors on obesity. *Int. J. Androl.* **31**, 201–8.
17. Kaur J (2014) A Comprehensive Review on Metabolic Syndrome. *Cardiol. Res. Pract.* **2014**, 943162.
18. Capodaglio P, Faintuch J & Liuzzi A (2013) *Disabling Obesity: From determinants to health care models*. [Capodaglio P, Faintuch J, Liuzzi A, editors]. Berlin, Heidelberg: Springer Berlin Heidelberg.
19. Hauner H (2010) Obesity and Diabetes. In *Textb. Diabetes*.
20. Felber J-P & Golay A (2002) Pathways from obesity to diabetes. *Int. J. Obes. Relat. Metab. Disord.* **26 Suppl 2**, S39-45.
21. Ahmad SI (2013) *Diabetes: An Old Disease, a New Insight*. [Ahmad SI, editor]. New York, NY: Springer New York.
22. National Center for Chronic Disease Prevention and Health Promotion (2014) *National Diabetes Statistics Report 2014 Estimates of Diabetes and Its Burden in the*

Epidemiologic estimation methods. Atlanta, GA: .

23. Verbeeten KC, Elks CE, Daneman D, et al. (2011) Association between childhood obesity and subsequent Type 1 diabetes: a systematic review and meta-analysis. *Diabet. Med.* **28**, 10–18.

24. Ouchi N, Kihara S, Funahashi T, et al. (2003) Obesity, adiponectin and vascular inflammatory disease. *Curr. Opin. Lipidol.* **14**, 561–566.

25. Lara-Castro C, Fu Y, Chung BH, et al. (2007) Adiponectin and the metabolic syndrome: mechanisms mediating risk for metabolic and cardiovascular disease. *Curr. Opin. Lipidol.* **18**, 263–270.

26. van Kruijsdijk RCM, van der Wall E & Visseren FLJ (2009) Obesity and cancer: the role of dysfunctional adipose tissue. *Cancer Epidemiol. Biomarkers Prev.* **18**, 2569–2578.

27. Prieto-Hontoria PL, P??rez-Matute P, Fern??ndez-Galilea M, et al. (2011) Role of obesity-associated dysfunctional adipose tissue in cancer: A molecular nutrition approach. *Biochim. Biophys. Acta - Bioenerg.*, 664–678.

28. Freedman DS, Dietz WH, Srinivasan SR, et al. (1999) The relation of overweight to cardiovascular risk factors among children and adolescents: the Bogalusa Heart Study. *Pediatrics* **103**, 1175–1182.

29. Kushi LH, Byers T, Doyle C, et al. (2006) American Cancer Society Guidelines on Nutrition and Physical Activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. *CA. Cancer J. Clin.* **56**, 254.

30. Serdula MK, Ivery D, Coates RJ, et al. (1993) Do Obese Children Become Obese Adults? A Review of the Literature. *Prev. Med. (Baltim).* **22**, 167–177.

31. Whitaker RC, Wright JA, Pepe MS, et al. (1997) Predicting obesity in young adulthood

from childhood and parental obesity. *N. Engl. J. Med.* **337**, 869–873.

32. Linke SE, Robinson CJ & Pekmezi D (2013) Applying Psychological Theories to Promote Healthy Lifestyles. *Am. J. Lifestyle Med.* **8**, 4–14.

33. Nixon C a, Moore HJ, Douthwaite W, et al. (2012) Identifying effective behavioural models and behaviour change strategies underpinning preschool- and school-based obesity prevention interventions aimed at 4-6-year-olds: a systematic review. *Obes. Rev.* **13 Suppl 1**, 106–17.

34. Almarzooqi MA & Nagy MC (2011) Childhood Obesity Intervention Programs : A Systematic Review. *Life Sci.* **8**, 45–60.

35. Glanz K & Rimer BK (2005) *Theory at a Glance: A Guide for Health Promotion Practice. Heal. (San Fr.*

36. Flynn MAT, McNeil DA, Maloff B, et al. (2006) Reducing obesity and related chronic disease risk in children and youth: a synthesis of evidence with 'best practice' recommendations. *Obes. Rev.* **7 Suppl 1**, 7–66.

37. Resnicow K & Vaughan R (2006) A chaotic view of behavior change: a quantum leap for health promotion. *Int. J. Behav. Nutr. Phys. Act.* **3**, 25.

38. Baranowski T (2006) Crisis and chaos in behavioral nutrition and physical activity. *Int. J. Behav. Nutr. Phys. Act.*, 27.

39. Hammond RA (2009) Complex systems modeling for obesity research. *Prev Chronic Dis* **6**, A97.

40. Wong RJ, Chou C & Ahmed A (2014) Long Term Trends and Racial/Ethnic Disparities in the Prevalence of Obesity. *J. Community Health.*

41. Braveman P (2006) Health disparities and health equity: concepts and measurement.

Annu. Rev. Public Health **27**, 167–94.

42. National Institute of Health (NIH) (2005) *Addressing disparities: the NIH program of action*. .

43. Dehlendorf C, Bryant A, Huddleston H, et al. (2010) Health Disparities: Definitions and Measurements. *Am. J. Obstet. Gynecol.* **202**, 212–213.

44. Krieger N, Williams DR & Moss NE (1997) Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu. Rev. Public Health* **18**, 341–378.

45. Summerbell C, Waters E, Edmunds L, et al. (2009) Interventions for preventing obesity in children (Review). *Cochrane database Syst. Rev.*, CD001871.

46. Maglio PP & Mabry PL (2011) Agent-based models and systems science approaches to public health. *Am. J. Prev. Med.* **40**, 392–4.

47. Skinner AC & Foster EM (2013) Systems science and childhood obesity: A systematic review and new directions. *J. Obes.*

48. Auchincloss AH & Diez Roux A V (2008) A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. *Am. J. Epidemiol.* **168**, 1–8.

49. Hui E, Xu A, Bo Yang H, et al. (2013) Obesity as the common soil of non-alcoholic fatty liver disease and diabetes: Role of adipokines. *J. Diabetes Investig.* **4**, 413–425.

50. Lloyd LJ, Langley-Evans SC & McMullen S (2012) Childhood obesity and risk of the adult metabolic syndrome: a systematic review. *Int. J. Obes. (Lond).* **36**, 1–11. Nature Publishing Group.

51. Skinner AC & Skelton JA (2014) Prevalence and Trends in Obesity and Severe Obesity Among Children in the United States, 1999-2012. *JAMA Pediatr.*

52. Fuglestad PT, Jeffery RW & Sherwood NE (2012) Lifestyle patterns associated with diet, physical activity, body mass index and amount of recent weight loss in a sample of successful weight losers. *Int. J. Behav. Nutr. Phys. Act.* **9**, 79.

53. Moyer VA & Butler M (2004) Gaps in the evidence for well-child care: a challenge to our profession. *Pediatrics* **114**, 1511–1521.

54. Wofford LG (2008) Systematic Review of Childhood Obesity Prevention. *J. Pediatr. Nurs.*, 5–19.

55. Stephens SK, Cobiac LJ & Veerman JL (2014) Improving diet and physical activity to reduce population prevalence of overweight and obesity: an overview of current evidence. *Prev. Med. (Baltim).* **62**, 167–78.

56. County of Los Angeles Department of Public Health (2012) *Trends in Obesity: Adult obesity continues to rise*. Los Angeles, CA: .

57. Los Angeles County Department of Public Health Office of Health Assessment and Epidemiology (2011) *Obesity and Related Mortality in Los Angeles County: A Cities and Communities Health Report*. Los Angeles, CA: .

58. County of Los Angeles Department of Public Health (2010) *Project RENEW: Fact sheet*. Los Angeles, CA: .

59. Bandura A (1977) *Social Learning Theory*. New York, NY: General Learning Press.

60. Egger G, Swinburn B & Rossner S (2003) Dusting off the epidemiological triad: could it work with obesity? *Obes. Rev.* **4**, 115–9.

61. McLeroy KR, Bibeau D, Steckler A, et al. (1988) An ecological perspective on health promotion programs. *Health Educ. Q.* **15**, 351–77.

62. Bronfenbrenner U (1994) Ecological models of human development. *Int. Encycl. Educ.* **3**,

37–43.

63. Willows ND, Hanley AJG & Delormier T (2012) A socioecological framework to understand weight-related issues in Aboriginal children in Canada. *Appl. Physiol. Nutr. Metab* **13**, 1–13.

64. Williams JE, Kabukuru A, Mayo R, et al. (2011) Commentary: A social-ecological perspective on obesity among Latinos. *Ethn. Dis.* **21**, 467–472.

65. Elder GH (1998) The Life Course as Developmental Theory. *Child Dev.* **69**, 1–12. Blackwell Publishing Ltd.

66. Kuh D, Ben-Shlomo Y, Lynch J, et al. (2003) Life course epidemiology. *J. Epidemiol. Community Health* **57**, 778–783.

67. Ben-Shlomo Y & Kuh D (2002) A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int. J. Epidemiol.*, 285–293.

68. Wahi G & Anand SS (2013) Race/Ethnicity, Obesity, and Related Cardio-Metabolic Risk Factors: A Life-Course Perspective. *Curr. Cardiovasc. Risk Rep.* **7**, 326–335.

69. Luke D a. & Stamatakis K a. (2012) Systems science methods in public health: dynamics, networks, and agents. *Annu. Rev. Public Health* **33**, 357–76.

70. Glass T a & McAtee MJ (2006) Behavioral science at the crossroads in public health: extending horizons, envisioning the future. *Soc. Sci. Med.* **62**, 1650–71.

71. SHIFT Obesity System inlfuence Diagram. http://www.shiftn.com/obesity/Full-Map.html (accessed January 2014).

72. Pearl J (1995) Causal diagrams for emprirical research. *Biometrika* **82**, 669–688.

73. Rothman K, Greenland S & Lash T (2008) *Modern Epidemiology*. 3rd ed. Philadelphia:

Lippincott Williams & Wilkins.

74. Greenland S, Pearl J & Robins JM (1999) Causal diagrams for epidemiologic research. *Epidemiology* **10**, 37–48.

75. Epstein JM & Axtell R (1996) *Growing artificial societies: Social science from the bottom up. Proc. fifth Annu. CCSC Northeast.* Washington, DC: Brookings Institution Press.

76. Miller JH & Page SE (2007) *Complex Adaptive Systems: An Introduction to computational models of social life*. Princeton, New Jersey: Princeton University Press.

77. Railsback SF & Grimm V (2012) *Agent-Based and Individual-Based Modeling*. Princeton, New Jersey: Princeton University Press.

78. Page SE (2008) Agent-based models. In: Durlauf SN, Blume LE, eds. The new Palgrave dictionary of economics. In *New Palgrave Dict. Econ.*, 2nd editio. New York, NY: Palgrave Macmillan.

79. Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. U. S. A.* **99 Suppl 3**, 7280–7.

80. Nilsson F & Darley V (2006) On complex adaptive systems and agent-based modelling for improving decision-making in manufacturing and logistics settings: Experiences from a packaging company. *Int. J. Oper. Prod. Manag.* **26**, 1351–1373. Emerald Group Publishing Limited.

81. Day TE, Ravi N, Xian H, et al. (2013) An Agent-Based Modeling Template for a Cohort of Veterans with Diabetic Retinopathy. *PLoS One* **8**, e66812.

82. Yang Y, Diez Roux A V, Auchincloss AH, et al. (2011) A spatial agent-based model for the simulation of adults' daily walking within a city_Appendix. *Am. J. Prev. Med.* **40**, 353–61.

83. Gorman DM, Mezic J, Mezic I, et al. (2006) Agent-based modeling of drinking behavior: a preliminary model and potential applications to theory and practice. *Am. J. Public Health* **96**, 2055–60.

84. Macal CM & North MJ (2010) Tutorial on agent-based modelling and simulation. *J. Simul.* **4**, 151–162. Palgrave Macmillan.

85. North MJ & Macal CM (2007) *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and simulation*. New York, NY: Oxford University Press.

86. Sargent RG (2007) Verification and Validation of simulation models. *Proc. 2007 Winter Simul. Conf.*, 124–137.

87. Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods--application to control of the healthy worker survivor effect. *Mathe Model* **7**, 1393–1512.

88. Taubman SL, Robins JM, Mittleman MA, et al. (2009) Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int. J. Epidemiol.* **38**, 1599–1611.

89. Danaei G, Pan A, Hu FB, et al. (2013) Hypothetical midlife interventions in women and risk of type 2 diabetes. *Epidemiology* **24**, 122–8.

90. Westreich D, Cole SR, Young JG, et al. (2012) The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Stat. Med.* **31**, 2000–9.

91. Robins JM & Hernan MA (2009) *Estimation of the causal effects of time-varying exposures*. *Verbeke G, Davidian M, Fitzmaurice G, al*, Longitidun. Boca Raton, FL:

Chapman and Hall/CRC.

92. Daniel RM, Cousens SN, De Stavola BL, et al. (2013) Methods for dealing with time-dependent confounding. *Stat. Med.* **32**, 1584–618.

93. Hernán MA & Robins JM (2006) Estimating causal effects from epidemiological data. *J. Epidemiol. Community Health* **60**, 578–586.

94. Westreich D & Cole SR (2010) Invited commentary: positivity in practice. *Am. J. Epidemiol.* **171**, 674-677-681.

95. Cole SR & Frangakis CE (2009) The consistency statement in causal inference: a definition or an assumption? *Epidemiology* **20**, 3–5.

96. Hedley AA, Ogden CL, Johnson CL, et al. (2004) Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002. *JAMA* **291**, 2847–50. American Medical Association.

97. te Velde SJ, Twisk JWR & Brug J (2007) Tracking of fruit and vegetable consumption from adolescence into adulthood and its longitudinal association with overweight. *Br. J. Nutr.* **98**, 431–8.

98. Bjelland M, Brantsæter AL, Haugen M, et al. (2013) Changes and tracking of fruit, vegetables and sugar-sweetened beverages intake from 18 months to 7 years in the Norwegian Mother and Child Cohort Study. *BMC Public Health* **13**, 793.

99. Singh AS, Mulder C, Twisk JWR, et al. (2008) Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obes. Rev.* **9**, 474–488.

100. Lopez RP & Hynes HP (2006) Obesity , physical activity , and the urban environment : public health research needs. *Environ. Heal. A Glob. Access Sci.* **10**, 1–10.

101. Papas M a, Alberg AJ, Ewing R, et al. (2007) The built environment and obesity.

Epidemiol. Rev. **29**, 129–43.

102. Popkin BM, Duffey K & Gordon-Larsen P (2005) Environmental influences on food choice, physical activity and energy balance. *Physiol. Behav.* **86**, 603–13.

103. Christakis NA & Fowler JH (2007) The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–9. Massachusetts Medical Society.

104. Must A, Spadano J, Coakley EH, et al. (1999) The disease burden associated with overweight and obesity. *JAMA* **282**, 1523–9.

105. Petersen M (2003) Economic costs of diabetes in the U.S. in 2002. *Diabetes Care* **26**, 917–932.

106. Siegfried R (2014) *Modeling and Simulation of Complex Systems*. Wiesbaden: Springer Fachmedien Wiesbaden.

107. Eddy DM & Schkessinger L (2003) A trial-validated model of diabetes. *Diabetes Care* **26**.

108. Eddy DM & Schlessinger L (2003) Validation of the Archimedes Diabetes Model. *Diabetes Care* **26**, 3102–3110.

109. United States Census Bureau (2016) Los Angeles County QuickFacts. http://www.census.gov/quickfacts/table/PST045215/00,06037 (accessed October 2016).

110. U.S. Census Bureau (2014) American Community Survey, 2010-2014 American Community Survey 5-year estimates. http://factfinder2.census.gov (accessed January 2016).

111. Chaparro MP, Whaley SE, Crespi CM, et al. (2014) Influences of the neighbourhood food environment on adiposity of low-income preschool-aged children in Los Angeles County: a longitudinal study. *J. Epidemiol. Community Health* **68**, 1027–1033.

112. Cohen-Cole E & Fletcher JM (2008) Is obesity contagious? Social networks vs.

environmental factors in the obesity epidemic. *J. Health Econ.* **27**, 1382–1387.

113. Wing R & Jeffery R (1999) Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance. *J. Consult. Clin.*

114. Wang Y & Chen H-J (2012) Use of Percentiles and Z -Scores in Anthropometry. *Handb. Anthr.*, 91–114.

115. CDC (2016) A SAS Program for the 2000 CDC Growth Charts (ages 0 to <20 years). *CDC*. http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm (accessed September 2016).

116. World Health Organization (WHO) (2015) WHO BMI-for-age growth charts. *WHO*. World Health Organization; http://www.who.int/growthref/who2007_bmi_for_age/en/ (accessed September 2016).

117. World Health Organization (2016) BMI classification. *2016*. http://apps.who.int/bmi/index.jsp?introPage=intro_3.html (accessed October 2016).

118. Walk Score (2016) Walk Score. https://www.walkscore.com/methodology.shtml (accessed January 2016).

119. Carr LJ, Dunsiger SI & Marcus BH (2010) Walk Score as a global estimate of neighborhood walkability. *Am. J. Prev. Med.* **39**, 460–463. Elsevier Inc.

120. Carr LJ, Dunsiger SI & Marcus BH (2011) Validation of Walk Score for estimating access to walkable amenities. *Br. J. Sports Med.* **45**, 1144–1148.

121. Brown SC, Pantin H, Lombard J, et al. (2013) Walk score: Associations with purposive walking in recent cuban immigrants. *Am. J. Prev. Med.* **45**, 202–206. Elsevier.

122. Wolch J, Wilson J & Fehrenbach J (2005) Parks and Park Funding in Los Angeles: An Equity-Mapping Analysis. *Urban Geogr.* **26**, 4–35.

123. Walls and Associates/Your Economy (2016) National Establishement Time-Series (NETS) Database by Walls & Associates. http://143.235.14.134/our-databases.iegc#NETS.

124. Mortensen LH, Siegler IC, Barefoot JC, et al. (2006) Prospective associations between sedentary lifestyle and BMI in midlife. *Obes. (Silver Spring)* **14**, 1462–1471.

125. Steyerberg E (2009) *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer.

126. Robins J (1986) A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7**, 1393–1512.

127. Center for Disease Control and Prevention (2012) *Breastfeeding Report Card — United States , 2012*. .

128. California Health Interview Survey (AskCHIS) (2016) 2009 California Health Interview Survey. http://ask.chis.ucla.edu/ (accessed January 2016).

129. Los Angeles County Department of Public Health Office of Health Assessment and Epidemiology (2016) Los Angeles Health And Nutrition Examination Survey (LAHANES 2011-2012). http://publichealth.lacounty.gov/.

130. National Center for Health Statistics CDC (2015) National Health and Nutrition Examination Survey (NHANES) - Questionnaires, Datasets, and Related Documentation. http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm.

131. Los Angeles County Department of Public Health Office of Health Assessment and Epidemiology (2012) *Trends in Diabetes: Time for Action*. .

132. Holford TR (1991) Understanding the effects of age, period, and cohort on incidence and

mortality rates. *Annu. Rev. Public Health* **12**, 425–427.

133. Giordano GN, Ohlsson H, Kendler KS, et al. (2014) Age, period and cohort trends in drug abuse hospitalizations within the total Swedish population (1975-2010). *Drug Alcohol Depend.* **134**, 355–361. Elsevier Ireland Ltd.

134. Shi L, van Meijgaard J & Fielding J (2011) Forecasting diabetes prevalence in California: a microsimulation. *Prev. Chronic Dis.* **8**, A80.

135. International Diabetes Federation (2015) *Diabetes Fact Sheet in North America and Caribbean*. .

136. Wang Y, Beydoun MA, Liang L, et al. (2008) Will All Americans Become Overweight or Obese? Estimating the Progression and Cost of the US Obesity Epidemic. *Obesity* **16**, 2323–2330.

137. Gregg EW, Zhuo X, Cheng YJ, et al. (2014) Trends in lifetime risk and years of life lost due to diabetes in the USA, 1985–2011: a modelling study. *Lancet Diabetes Endocrinol.* **2**, 867–874.

138. Kuo T, Robles B, Trogdon JG, et al. (2016) Framing the Local Context and Estimating the Health Impact of CPPW Obesity Prevention Strategies in Los Angeles County, 2010-2012. *J. Public Heal. Manag. Pract.* **22**, 360–369.

139. Nianogo RA & Arah OA (2015) Agent-based modeling of noncommunicable diseases: a systematic review. *Am. J. Public Health* **105**, e20-31. American Public Health Association.

140. De Kroon ML, Renders CM, Buskermolen MP, et al. (2011) The Terneuzen Birth Cohort. Longer exclusive breastfeeding duration is associated with leaner body mass and a healthier diet in young adulthood. *BMC Pediatr* **11**, 33.

141. Li X-H, Lin S, Guo H, et al. (2014) Effectiveness of a school-based physical activity intervention on obesity in school children: a nonrandomized controlled trial. *BMC Public Health* **14**, 1282.

142. Aune D, Norat T, Leitzmann M, et al. (2015) Physical activity and the risk of type 2 diabetes: A systematic review and dose-response meta-analysis. *Eur. J. Epidemiol.* **30**, 529–542. Springer Netherlands.

143. Forshee RA, Anderson PA & Storey ML (2008) Sugar-sweetened beverages and body mass index in children and adolescents: A meta-analysis. *Am. J. Clin. Nutr.* **87**, 1662–1671.

144. Imamura F, O'Connor L, Ye Z, et al. (2015) Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. *BMJ* **351**, h3576.

145. Bertoia ML, Mukamal KJ, Cahill LE, et al. (2015) Changes in Intake of Fruits and Vegetables and Weight Change in United States Men and Women Followed for Up to 24 Years: Analysis from Three Prospective Cohort Studies. *PLoS Med.* **12**, 1–20.

146. Wu Y, Zhang D, Jiang X, et al. (2015) Fruit and vegetable consumption and risk of type 2 diabetes mellitus: A dose-response meta-analysis of prospective cohort studies. *Nutr. Metab. Cardiovasc. Dis.* **25**, 140–147.

147. Pereira MA, Kartashov AI, Ebbeling CB, et al. (2005) Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. *Lancet* **365**, 36–42.

148. Pan A, Sun Q & Bernstein A (2011) Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *Am.*, 1–9.

149. Llewellyn A, Simmonds M, Owen CG, et al. (2016) Childhood obesity as a predictor of morbidity in adulthood: A systematic review and meta-analysis. *Obes. Rev.* **17**, 56–67.

150. Moore L V., Diez Roux A V., Nettleton JA, et al. (2008) Associations of the local food environment with diet quality - A comparison of assessments based on surveys and geographic information systems. *Am. J. Epidemiol.* **167**, 917–924.

151. Moore L V., Diez Roux A V., Nettleton JA, et al. (2009) Fast-Food consumption, diet quality, and neighborhood exposure to fast food. *Am. J. Epidemiol.* **170**, 29–36.

152. Giles-Corti B, Broomhall MH, Knuiman M, et al. (2005) Increasing walking: How important is distance to, attractiveness, and size of public open space? *Am. J. Prev. Med.* **28**, 169–176.

153. Los Angeles County Department of Public Health Office of Health Assessment and Epidemiology (2012) *Trends in Obesity: Adult Obesity Continues to Rise, September 2012*. Los Angeles, CA: .

154. Singh AS, Mulder C, Twisk JWR, et al. (2008) Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obes. Rev.* **9**, 474–88.

155. Glynn AN (2012) The Product and Difference Fallacies for Indirect Effects. *Am. J. Pol. Sci.* **56**, 257–269.

156. Pearl J (2012) The Mediation Formula : A guide to the assessment of causal pathways in nonlinear models Mediation : Direct and Indirect Effects. In *Causality Stat. Perspect. Appl.*, pp. 151–179 [Berzuini C, Dawid P, Bernardinelli L, editors]. Chichester, UK: John Wiley and Sons, Ltd.

157. Victora CG, Lessa Horta B, Loret de Mola C, et al. (2015) Association between breastfeeding and intelligence, educational attainment, and income at 30 years of age: a

prospective birth cohort study from Brazil. *Lancet Glob. Heal.* **3**, e199–e205.

158. Zhang YT, Laraia B a., Mujahid MS, et al. (2015) Does food vendor density mediate the association between neighborhood deprivation and BMI? *Epidemiology* **26**, 344–352.

159. Loret de Mola C, Hartwig FP, Gonçalves H, et al. (2016) Genomic ancestry and the social pathways leading to major depression in adulthood: the mediating effect of socioeconomic position and discrimination. *BMC Psychiatry* **16**, 308. BioMed Central.

160. Daniel RM, Cousens SN, De Stavola BL, et al. (2013) Methods for dealing with time-dependent confounding. *Stat. Med.* **32**, 1584–1618.

161. Rubin DB (1980) Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Source J. Am. Stat. Assoc.* **75**, 591–593.

162. Pearl J (2001) Direct and indirect effects. In *Proc. Seventeenth Conf. Uncertain. Artif. Intell.*, pp. 411–420. San Francisco: Morgan Kaufmann.

163. Vanderweele TJ, Vansteelandt S & Robins JM (2014) Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* **25**, 300–6.

164. Wang A & Arah OA (2015) G-computation demonstration in causal mediation analysis. *Eur. J. Epidemiol.* **30**, 1119–1127.

165. Al Mamun A, Cramb SM, O'Callaghan MJ, et al. (2009) Childhood Overweight Status Predicts Diabetes at Age 21 Years: A Follow-up Study. *Obesity* **17**, 1255–1261.

166. Liang Y, Hou D, Zhao X, et al. (2015) Childhood obesity affects adult metabolic syndrome and diabetes. *Endocrine* **50**, 87–92. Springer US.

167. Reilly JJ & Kelly J (2011) Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. *Int J Obes* **35**, 891–898.

168. Adolescents HAUS (2009) Prevalence of Pre-Diabetes and Its Association With Clustering of Cardiometabolic Risk Factors and. *Cardiovasc. Metab. risk* **32**.

169. Nichols GA, Hillier TA & Brown JB (2007) Progression from newly acquired impaired fasting glusose to type 2 diabetes. *Diabetes Care* **30**, 228–33.

170. Ogden CL, Carroll MD, Kit BK, et al. (2014) Prevalence of childhood and adult obesity in the United States, 2011-2012. *JAMA* **311**, 806–14.

171. Eberhart MS, Ogden C, Engelgau M, et al. (2008) Prevalence of Overweight and Obesity Among Adults with Diagnosed Diabetes --- United States, 1988--1994 and 1999--2002. *MMWR*. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5345a2.htm (accessed July 2014).

172. Waters E, de Silva-Sanigorski A, Hall BJ, et al. (2011) Interventions for preventing obesity in children. *Cochrane database Syst. Rev.*, CD001871.

173. Jeon CY, Lokken RP, Hu FB, et al. (2007) Physical activity of moderate intensity and risk of type 2 diabetes: A systematic review. *Diabetes Care*, 744–752.

174. Skinner AC & Skelton J a (2014) Prevalence and Trends in Obesity and Severe Obesity Among Children in the United States, 1999-2012. *JAMA Pediatr.*

175. Doak CM, Visscher TLS, Renders CM, et al. (2006) The prevention of overweight and obesity in children and adolescents: a review of interventions and programmes. *Obes. Rev.* **7**, 111–36.

176. Prevention D & Diabetes Prevention Program Research Group (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **346**, 393–403.

177. Murthy VH, Krumholz HM & Gross CP (2004) Participation in cancer clinical trials: race-

, sex-, and age-based disparities. *JAMA* **291**, 2720–6.

178. Danaei G, Pan A, Hu FB, et al. (2013) Hypothetical midlife interventions in women and risk of type 2 diabetes. *Epidemiology* **24**, 122–8.

179. Orr MG, Galea S, Riddle M, et al. (2014) Reducing racial disparities in obesity: simulating the effects of improved education and social network influence on diet behavior. *Ann. Epidemiol.* **24**, 563–569.

180. Rose G (1985) Sick individuals and sick populations. *Int. J. Epidemiol.* **14**, 32–38.

181. Montonen J, Järvinen R, Knekt P, et al. (2007) Consumption of sweetened beverages and intakes of fructose and glucose predict type 2 diabetes occurrence. *J. Nutr.* **137**, 1447–54. American Society for Nutrition.

182. Te L, Research Fellow M, Mallard Research Assistant S, et al. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. .

183. Chen L, Appel LJ, Loria C, et al. (2009) Reduction in consumption of sugar-sweetened beverages is associated with weight loss. *Am. J. Clin. Nutr.* **89**, 1299–306.

184. Pan A, Malik V, Hao T, et al. (2013) Changes in water and beverage intake and long-term weight changes: results from three prospective cohort studies. *Int. J. Obes.* **37**, 1378–1385.

185. Zheng M, Allman-Farinelli M, Heitmann BL, et al. (2015) Substitution of Sugar-Sweetened Beverages with Other Beverage Alternatives: A Review of Long-Term Health Outcomes. *J. Acad. Nutr. Diet.* **115**, 767–779.

186. Yan J, Liu L, Zhu Y, et al. (2016) The association between breastfeeding and childhood obesity: A meta-analysis. *World Rev. Nutr. Diet.* **114**, 110–111.

187. Nianogo RA, Wang MC, Wang A, et al. (2016) Projecting the impact of hypothetical early

life interventions on adiposity in children living in lowincome households. *Pediatr. Obes.*, 1–8.

188. Casazza K, Fontaine KR, Astrup A, et al. (2013) Myths, Presumptions, and Facts about Obesity. *N. Engl. J. Med.* **368**, 446–454.

189. Hancox RJ, Stewart AW, Braithwaite I, et al. (2015) Association between breastfeeding and body mass index at age 6-7 years in an international survey. *Pediatr. Obes.* **10**, 283–7.

190. Sallis JF, Saelens BE, Frank LD, et al. (2009) Neighborhood built environment and income: Examining multiple health outcomes. *Soc. Sci. Med.* **68**, 1285–1293.

191. Sterman JD (2006) Learning from Evidence in a Complex World. *Am. J. Public Health* **96**, 505–514.

192. Orr MG, Kaplan GA & Galea S (2016) Neighbourhood food, physical activity, and educational environments and black/white disparities in obesity: a complex systems simulation analysis. *J. Epidemiol. Community Health*, 862–867.

193. Silveira JAC, Colugnati FAB, Poblacion AP, et al. (2015) The role of exclusive breastfeeding and sugar-sweetened beverage consumption on preschool children's weight gain. *Pediatr. Obes.* **10**, 91–7.

194. Public Health Foundation Project WIC (2002) Data Mining Project. http://www.phfewic.org/projects/DataMining.aspx.

195. Crespi CM, Alfonso VH, Whaley SE, et al. (2012) Validity of child anthropometric measurements in the Special Supplemental Nutrition Program for Women, Infants, and Children. *Pediatr. Res.* **71**, 286–92.

196. Jones A, Charakida M, Falaschetti E, et al. (2012) Adipose and height growth through childhood and blood pressure status in a large prospective cohort study. *Hypertension* **59**,

919–25.

197. Kuczmarski RJ, Ogden CL, Guo SS, et al. (2002) 2000 CDC Growth Charts for the United States: methods and development. *Vital Health Stat. 11.*, 1–190.

198. Yan J, Liu L, Zhu Y, et al. (2014) The association between breastfeeding and childhood obesity: a meta-analysis. *BMC Public Health* **14**, 1267.

199. Kramer MS, Matush L, Vanilovich I, et al. (2007) Effects of prolonged and exclusive breastfeeding on child height, weight, adiposity, and blood pressure at age 6.5 y: evidence from a large randomized trial. *Am J Clin Nutr* **86**, 1717–1721.

200. Hernell O (2011) Human milk vs. cow's milk and the evolution of infant formulas. *Nestle Nutr. Workshop Ser. Pediatr. Program.* **67**, 17–28.

201. Rolland-Cachera MF, Deheeger M, Akrout M, et al. (1995) Influence of macronutrients on adiposity development: a follow up study of nutrition and growth from 10 months to 8 years of age. *Int. J. Obes. Relat. Metab. Disord.* **19**, 573–8.

202. Casabiell X, Piñeiro V, Tomé MA, et al. (2013) Presence of Leptin in Colostrum and/or Breast Milk from Lactating Mothers: A Potential Role in the Regulation of Neonatal Food Intake. *J. Clin. Endocrinol. Metab.* Endocrine Society.

203. Birch LL & Fisher JO (1998) Development of eating behaviors among children and adolescents. *Pediatrics* **101**, 539–549.

204. World Health Organization (2001) Global strategy for infant and young child feeding The optimal duration of exclusive breastfeeding. 1–5.

205. Nyberg G, Sundblom E, Norman Å, et al. (2015) Effectiveness of a universal parental support programme to promote healthy dietary habits and physical activity and to prevent overweight and obesity in 6-year-old children: the Healthy School Start Study, a cluster-

randomised controlled trial. *PLoS One* **10**, e0116876. Public Library of Science.

206. Shearrer GE, Whaley SE, Miller SJ, et al. (2015) Association of gestational diabetes and breastfeeding on obesity prevalence in predominately Hispanic low-income youth. *Pediatr. Obes.* **10**, 165–71.

207. Ergin I, Hassoy H, Tanik FA, et al. (2010) Maternal age, education level and migration: socioeconomic determinants for smoking during pregnancy in a field study from Turkey. *BMC Public Health* **10**, 325.

208. Makgoba M, Savvidou MD & Steer PJ (2012) An analysis of the interrelationship between maternal age, body mass index and racial origin in the development of gestational diabetes mellitus. *BJOG* **119**, 276–82.

209. Sanzone LA, Lee JY, Divaris K, et al. (2013) A cross sectional study examining social desirability bias in caregiver reporting of children's oral health behaviors. *BMC Oral Health* **13**, 24.

210. Daniels SR, Hassink SG & The American Academy of Pediatrics Committee on Nutrition (2015) The Role of the Pediatrician in Primary Prevention of Obesity. *Pediatrics* **136**, e275-92.

211. Barlow SE & the Expert Committee (2007) Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. *Pediatrics* **120 Suppl**, S164-92.

212. The American Academy of Pediatrics Committee on Public Education (2001) Children, Adolescents, and Television. *Pediatrics* **107**, 423–426.

213. Wang X, Ouyang Y, Liu J, et al. (2014) Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response

meta-analysis of prospective cohort studies. *BMJ* **349**, g4490–g4490.