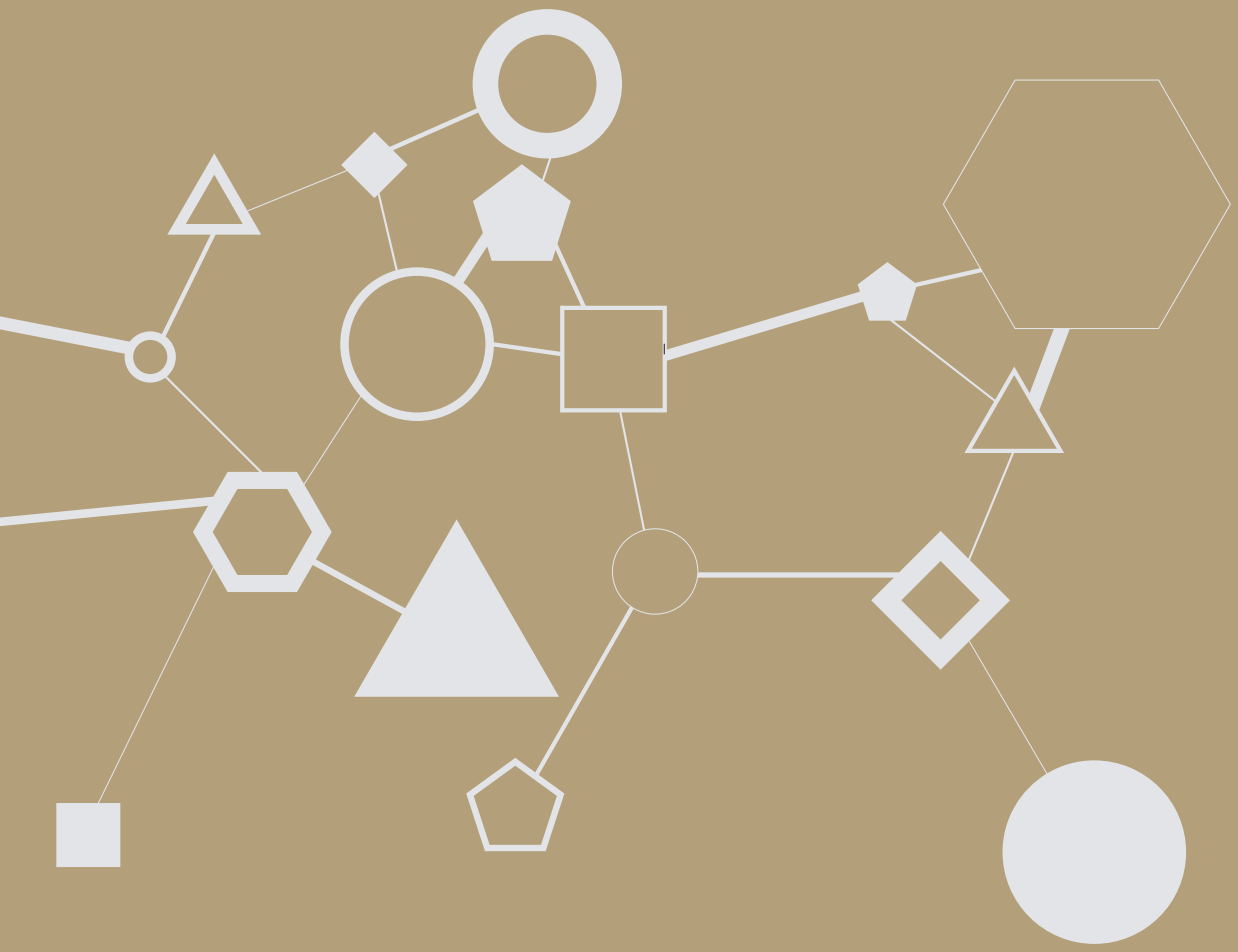
Introducing Systems Approaches in Health Behavioral Research



David J. Blok

INTRODUCING SYSTEMS APPROACHES IN HEALTH BEHAVIORAL RESEARCH

DAVID J. BLOK

Introducing Systems Approaches in Health Behavioral Research
Thesis, Erasmus MC, University Medical Center Rotterdam

ISBN: 978-94-6332-347-5

Design: David J. Blok
Print: GVO drukkers & vormgevers B.V.

This thesis was financially supported by the Department of Public Health and the Erasmus MC.

Introducing Systems Approaches in Health Behavioral Research

Introduceren van systeemmethoden in onderzoek naar gezondheidsgerelateerde gedragingen

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

dinsdag 5 juni 2018 om 11:30 uur

David Johannes Blok
geboren te Rotterdam

Erasmus University Rotterdam

Promotiecommissie:

Promotoren:	Prof.dr. J.H. Richardus
	Prof.dr. F.J. van Lenthe
	Prof.dr. S.J. de Vlas
Overige leden:	Prof.dr. O.H. Franco Duran
	Prof.dr. E.F.C. van Rossum
	Prof.dr. J. Wallinga

Contents

1

GENERAL INTRODUCTION

The epidemiology of health behaviors

Health behaviors are major determinants of morbidity and mortality.[1] Smoking, unhealthy diet, physical inactivity and as a result obesity, are among the top leading risk factors of non-communicable diseases such as type 2 diabetes, cardiovascular disease and several types of cancer (see Figure 1).[2] They cause more than two-thirds of all new cases of non-communicable diseases, and increase the risk of complications in people with those diseases.[1-4] Smoking remains the largest avoidable health risk in the general population, killing around 6 million people each year worldwide and contributing to around 6% of global disability adjusted life years (DALYs).[2,4-6] The prevalence of obesity increased dramatically in the past decades,[3] and a high body-mass index contributes to more than 5% of global DALYs. Low fruit and low vegetable intake is one of the leading risk factors for mortality with approximately 5.2 million deaths globally in 2013.[7,8] Insufficient physical activity is the cause of around 3.2 million deaths yearly.[9] Despite various efforts to reduce unhealthy behaviors, still more than 20% of the global population smokes,[10] around 13% is obese,[3] approximately a quarter of the adult population does not fulfill the guidelines for physical activity (i.e. at least 150 minutes per week),[11] and less than a quarter of the population meets the recommendations of fruit and vegetable consumption.[12,13]

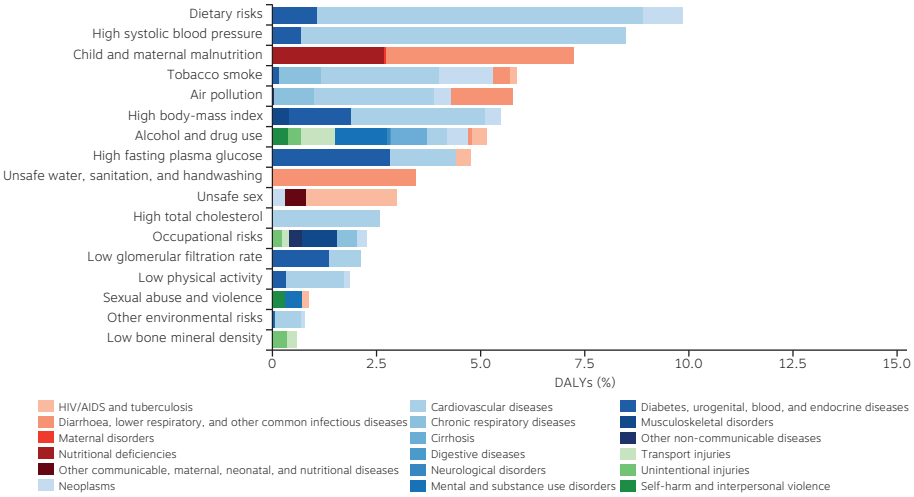


Figure 1. Burden of disease attributable to leading risk factors in 2013. It is expressed as a percentage of the global disability-adjusted life-years. Reproduced with permission from *Global Burden of Disease (GBD) Collaborators (2015)*.[2]

Prevalence rates of unhealthy behaviors vary between socioeconomic groups. In general, lower socioeconomic groups are at an increased risk of unhealthy behaviors compared to higher socioeconomic groups, regardless of the measure of socioeconomic status used (e.g. income, education, or occupation). In most countries smoking is far more common among individuals in lower socioeconomic groups.[6,14] Persons in lower socioeconomic groups are also at an increased risk of lower levels of physical activity, fruit and vegetables consumption, and higher levels of fat intake compared to those in higher socioeconomic groups.[15-17] Inequalities in obesity and overweight are known to be large and persistent in Western countries.[18] As a result, morbidity and mortality rates are higher in lower socioeconomic groups as compared to higher socioeconomic groups.[6,19,20] The higher prevalence of unhealthy behaviors among lower socioeconomic groups as compared to higher socioeconomic groups is among the main reasons of socioeconomic inequalities in health.[19,21] For example, smoking constitutes the single most important contributor to socioeconomic inequalities in mortality among men.[6,20]

Both relative and absolute inequalities in smoking prevalence and physical

inactivity have widened in the past decades in Western countries.[22,23] Trends in absolute socioeconomic inequalities in obesity showed either a stable or widening trend, depending on the country.[24,25] Reducing socioeconomic inequalities in health behaviors therefore remains a major challenge in public health. Thus far, little is known about how to do this effectively.[26] While modest changes in health behaviors can be achieved with theoretically informed interventions, the long term impact and the translation into health improvements at a population level are poorly understood.[27,28]

An ecological perspective of health behaviors

Socioeconomic inequalities in health behaviors are believed to result from selection processes (whereby health behaviors determine socioeconomic status), and social causation (whereby socioeconomic status has an indirect effect on health behaviors through an unequal distribution of determinants of behaviors). The latter mechanism is generally seen as the dominant one. For a long time, individual cognitive factors derived from behavioral change theories, were considered the most important determinants of health behaviors.[29,30] According to *Theory of Planned Behavior* (TPB), health behaviors are determined by intentions, which in turn are determined by attitudes, self-efficacy and subjective norms. Research has shown that self-efficacy is consistently associated with smoking, dietary intake and physical activity.[31-33] These factors are also known to vary by socioeconomic groups whereby lower socioeconomic groups for example have a lower attitude towards healthy behavior, which contributes to socioeconomic inequalities in smoking and obesity.[34] However, it is not easy to understand why socioeconomic groups differ in individual cognitive factors, if not determined by shared underlying factors.[35]

Public health scholars increasingly recognize that determinants of health behaviors cannot be fully understood in isolation of the context in which behaviors are shaped and sustained.[27] Therefore, they adopted an ecological approach, which emphasizes the larger physical and social context of behavior.[36,37] Features of the physical environment may constrain, reward or induce the behavior of individuals.[27] Indeed, access to supermarkets or lower accessibility to takeaway outlets was associated with healthier food consumption and a lower prevalence of

obesity, whereas the availability of parks was positively associated with physical activity.[38-40] Similarly, the social environment provides opportunities for sharing norms around behaviors, social support for behavioral decisions, and social influence.[41] It is for such reasons that the social environment is important for smoking cessation and weight loss.[42,43] Environmental factors may also contribute to socioeconomic inequalities in unhealthy behaviors. Lower socioeconomic groups may reside more often in neighborhoods less supportive for certain health behaviors, including poorer access to facilities, and less favorable social circumstances.[44,45] To make it even more complex, health behaviors may result from interactions between features of the social and physical environment, and individual factors.[27] Environments may reinforce individual cognitive factors: stronger intentions to sports, for example, were associated with better availability of sports facilities.[46] A more contextual understanding of smoking, obesity, diet and physical activity would therefore advance the effectiveness of public health policies and interventions.[27] In the past decades research on environmental factors for health behaviors has primarily focused on the role of the physical environment, while the social environment received less attention.

The influence of social networks

The importance of social networks on health and health behaviors is now widely recognized.[47] Network theories assume that the social network is largely responsible for individual behavior and attitudes, by shaping the flow of resources or information that provide opportunities and constraints on behaviors.[41] Social networks influence health behaviors through four pathways.[47] The first pathway is social support, which includes the provision of emotional, instrumental, appraisal and informational support to others.[48] Social support is known to be important for smoking cessation and weight loss.[42,43] Secondly, social networks provide opportunities for social engagement or participation, which gives a person a sense of value through meaningful social roles (e.g. parental roles) and interpersonal attachment. No social engagement or social isolation has been associated with a higher prevalence of smoking.[49] Thirdly, social networks provide access to resources and new information. Both close and weak (i.e. not close) ties are important to facilitate the diffusion of resources and information.[50]

The last and often ignored pathway of social networks is social influence, which is the process of mutual influence taking place in the network.[47] Marsden stated that proximity of two people in social networks is associated with the occurrence of interpersonal influence between these persons.[51] Social influence does not require deliberate attempts to change behavior, nor needs face-to-face contact.[41] Shared norms might be an important source of social influence.[47] Being within a similar environment without active social interaction, such as at work or in the same living area, may already be sufficient to influence behaviors.

The public health relevance of social networks changed with key papers on smoking, obesity, and other health risk factors by *Christakis and Fowler*.[52,53] Using the Framingham Heart Study from 1971 to 2003, they showed the dynamics of smoking and obesity in social networks. They nicely demonstrated that obese persons and smokers tend to cluster in the network (see Figure 2). In addition, they concluded that the risk of smoking increased by 61% if a person is socially close to a smoker. This risk differed with the type of social tie or relationship: close friends and spouses had the highest impact on participants' smoking behavior.[53] Similarly, the risk of a participant to become obese was about 57% if he or she had a friend who became obese, and 40% if he or she had a sibling who became obese.[52] These studies do not only support the idea of social contagion or a person-to-person spread, but also suggest that different social ties may have a different impact on the spread of health-behavior.

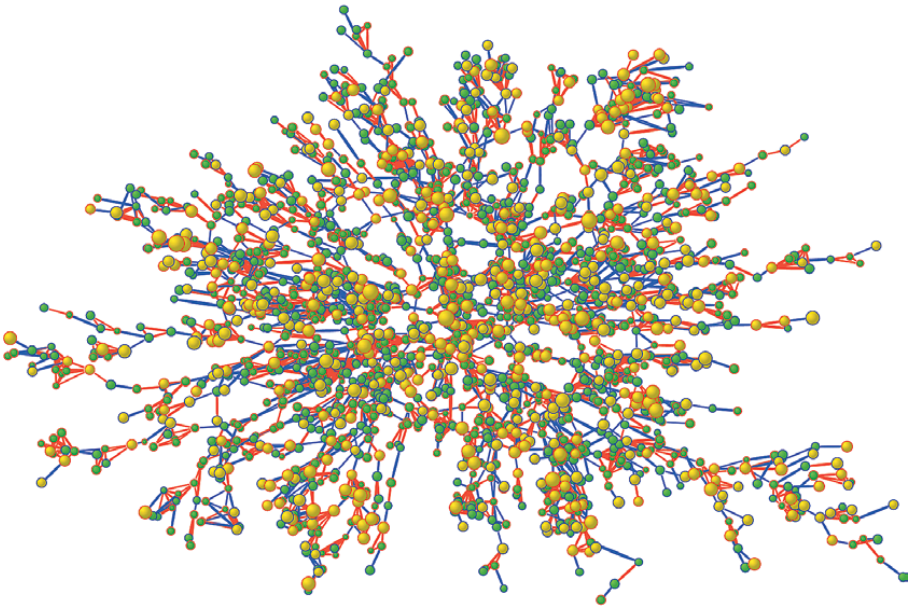


Figure 2. Clustering of obese persons in a large social network. Each circle represents one person in the data. The color indicates the person's obesity status: yellow is an obese person and green is a non-obese person. Red circle borders denote women, and blue circle borders denote men. Data are from the Framingham Heart Study in the year 2000. Reproduced with permission from *Christakis et al. (2007)*,[52] Copyright Massachusetts Medical Society.

An important consideration when investigating the influence of social networks in observational studies is to distinguish between the impact of social influence and homophily.[54,55] Homophily (or selection) is the tendency of people to select others with similar traits e.g. age, socioeconomic position, smoking behavior or other behaviors. Generally a contact between similar individuals is more likely than between dissimilar individuals.[56] Various methods have been proposed to account for (unmeasured) homophily including controlling for previous behavioral status (e.g. smoking status) in statistical modeling,[57,58] and the use of simulation modeling, such as the actor-based models for network dynamics.[59]

Christakis and Fowler suggest a similarity with the spread of infectious diseases, which is a field that has a long history in considering dynamic systems to describe spread (of infection) within social networks.[52,53,57] It is fascinating to consider the wider common ground between the spread of infections and health behaviors.

If, in analogy to infectious diseases, health-related behavior can be considered contagious, then models of infectious diseases could be useful for health-related behaviors as well. A first attempt has been made by *Hill et al.* who modelled the obesity epidemic as an infectious disease.[60] The model simulates transitions between two compartments, non-obese and obese, and predicts a long-term obesity prevalence of around 42%. Key features of this model are its ability to investigate the relative importance of social transmission and to make long-term predictions. Further exploiting models of infectious disease could enhance our understanding of the spread of unhealthy behaviors, as well as providing new targets for intervention.

Evidence that smoking and obesity spread in social networks provides opportunities to take advantage of the network to prevent such behaviors. Smoking cessation programs and weight-loss interventions that provide peer support are known to be more successful than those that do not.[42,43] Alternatively influential individuals ("role models") could be targeted to maximize population-level behavior change.[61] For example, randomized controlled trials of smoking cessation interventions that target students based on their network position have documented peer effects.[62,63] Also, public health interventions might be more (cost-)effective than initially thought, because health improvements in one person might spread to others.[52,53]

Systems thinking

Health problems arise from a complex causal web in which determinants of health and health behaviors mutually influence each other over life courses. Therefore scholars have recently advocated treating health problems as a "system".[33,64-66] In a system, problems are not explained by an understanding of their components alone.[67] A key feature of systems thinking is the recognition that health of individuals at the population-level emerges from the behaviors of heterogeneous individuals and the interactions of individuals with each other and with environments.[64] Behaviors are known to be sensitive to initial conditions, such as socioeconomic position (both of the individual and e.g. his/her parents) or the environmental state, and may adapt over time to the changes in other individuals (i.e. social network) and the physical environment.[66] These interactions over

time occur in complex ways, which enforce many of the mechanisms of health and health behaviors For example, membership of social networks is based on personal preferences and personal characteristics.[56] At the same time, networks may influence health behaviors, for example through social influence or social support during life.[41] As a result small behavioral changes can potentially have large system-level or population-level effects.

Another feature of systems thinking is the presence of positive or negative feedback loops, where determinants can modulate health behaviors as much as health behaviors can modulate determinants.[64,68] For example, the availability of places to be physical active promotes physical activity, but new sports facilities are more likely to locate in areas where individuals are known to be active.[69] Also, improvements in health behaviors in one person might influence socially-close other persons to improve their health behaviors, mutually reinforcing a positive feedback loop.[52,53] Hence, environments may influence health behaviors, and people may influence their environment. These dynamics are often not investigated or even considered in public health, although they might have considerable implications for behavioral research.

Dynamics of population behaviors and health also feature nonlinearity. Changes in risk factors are not always proportional to the changes in behaviors. Yet, the mainstay approach reduces the system to a series of isolated and independent effect measures that are merely associations [66]. This approach has been criticized for its inability to identify causal factors, because of the complex interactions and the lack of a good counterfactual.[27,70] In quasi-experimental studies the intervention and control group may still differ in many respects, making it difficult to infer what would have happened to the control group had it been exposed to the intervention.[70] Many public health problems, such as the obesity epidemic, have proven to be difficult to solve. Despite numerous intervention studies, an effective solution is not yet available.[70] Also very little progress has been made in eliminating inequalities.[64] One possibility of this is that the underlying and structural causes have not yet been sufficiently addressed. Since it is increasingly clear that health problems arise from complex multilevel processes, health behaviors and also health should be studied in a system using system approaches.[64,69,70]

Systems approaches have been introduced successfully in many fields of research, such as economics and political sciences, but this paradigm has hardly entered public health thus far.[65,71,72] Systems approaches have the potential to take into account all elements of the system to generate macro-level patterns from lower level processes, including feedback loops and dynamic interactions between individuals and between individuals and their environment.[64,69,73] It can also inform our knowledge about how policies or interventions influence health behaviors. It could move the field of behavioral research forward in three important ways: (1) promoting the development of more sophisticated dynamic conceptual models to understand the causes of health behaviors; (2) exploring the long-term effects of various interventions in the context of dynamic interactions; (3) promoting the collection of new types of data.[64] Systems approaches that are commonly used include systems dynamics, network analysis and agent-based modeling.[65] These methods are to some extent overlapping. Agent-based modeling is particularly promising, because it is the only tool that can dynamically account for interactions between heterogeneous agents and their environment.[65]

Agent-based modeling

Agent-based modeling (ABM), in other disciplines also called individual-based modeling, is a computational simulation method with the aim to represent the complexities of real-life processes at the level of individuals (agents) and to explore how these will behave in the future.[74] These processes can be described by rules and interactions among individuals and between individuals and environments, which influence their behaviors.[75,76] ABM typically facilitates a bottom-up approach, which means that phenomena observed at the population level are the result of underlying individual decisions that are explicitly modelled, while accounting for nonlinearity, interactions, and feedbacks.[64,70,75] It also provides a natural description of a system (i.e. close to reality) and is relatively flexible compared to other (e.g. deterministic) modeling approaches. ABM allows alterations or variations on macro group levels, sub groups or single agent level. This makes it very suitable to test the impact of different real-life policy and intervention scenarios.[64]

Agents in the model can be any entity, usually individuals, but also households and

facilities, as we will explore in this thesis. Each agent is characterized by a set of attributes (e.g. age, sex, income level). Agents are autonomous, interdependent, heterogeneous, adaptive and follow simple rules.[75] These behavioral rules describe how an agent interacts with other agents and the environment. Using transition probabilities, an ABM can simulate changes in state and behavior of each agent. These agents can adapt their behaviors in response to changes in behaviors of other agents and to changes in their environment due to for example interventions. Agents can also be clustered into groups at different levels such as households, social networks or neighborhoods.

One of the main challenges of the application of ABMs is the balance between the level of complexity or detail and model parsimony. The process of modeling should as much as possible be tailored to the research questions of interest to avoid unnecessary complexity.[75] Another challenge is the validation of these models. ABMs are difficult to validate completely and it can be a challenge to identify all relevant data to parameterize a model, which affect the quality of forecasting abilities. Generally, parameters of the model are quantified using real data or calibrated against real world observations.[68,69] Sensitivity and uncertainty analyses are often essential ingredients for studies using ABM, to express the consequences of the lack of (or uncertainty in) parameter quantifications. Finally, it can be computationally intensive and therefore time consuming to run ABMs.

Modelling health behaviors or socioeconomic inequalities in health behaviors as a system may help tackling two major challenges regarding interventions. Firstly, very little is known about how to reduce unhealthy behaviors or socioeconomic inequalities in health behaviors.[27] As mentioned earlier, the causal impact of interventions is poorly understood.[64,70] An important reason is that randomization of environmental factors is almost impossible, precluding causal inferences. Trials or observational studies always face the fundamental problem of a missing counterfactual.[27,70] With a model, the effect of different interventions can be studied in the same population.[64,70,72] Secondly, interventions cannot reasonably result in an observable reduction of unhealthy behaviors or inequalities in health behaviors in the short run, so that only (very expensive) long-term studies could provide real evidence on their eventual impact.[27,28] To identify effective policies to reduce inequalities in health behaviors, ABMs can be of very useful because they

can project possible long-term impact of interventions. However, the value of such long-term predictions remains modest, as several critical (fixed) assumptions may change in the future, such as economic and medical developments, as well as demographic trends.

Although ABM has been successfully adopted in many fields of research, it has hardly entered public health thus far, with the important exception of infectious disease epidemiology.[65,69] In infectious disease epidemiology, individual-based models are used to predict the spread of disease and the impact of control. At Erasmus MC it has been used for HIV, leprosy and worm infections, such as onchocerciasis.[77-79] Within the field of social epidemiology, ABMs are being recognized as a tool to assess the impact of various policies or interventions. Recent ABM studies have focused on dietary behaviors, social networks and obesity, and daily walking.[80-84] These studies lack the sophistication of work on infectious diseases, due to the short research history and lack of data.

Aims and objectives

The aims of this thesis are twofold: (1) to explore and quantify the importance of social networks as a determinant of health behaviors, and (2) to investigate the usefulness of agent-based models as a tool for assessing the impact of interventions to reduce socioeconomic inequalities in health behaviors.

First, we investigated to which extent there is a common ground between the spread of infections and unhealthy behaviors, and how experiences from infectious disease modeling could be useful for the field of social epidemiology. We also used existing and new data sources to analyze the influence of social networks on smoking, sports participation and overweight.

The second part of this thesis focuses on the application of systems approaches through agent-based modeling. To overcome the limitations of traditional methods and to be able to evaluate which policies or interventions have potentially the highest impact on reducing socioeconomic inequalities in health behaviors, we developed two agent-based models as proof of concepts. These models focus on dietary behaviors and sports participation, accounting for dynamic interaction between individuals or households and food shops and sports facilities,

respectively. Interventions that were evaluated target both individual as well as environmental factors.[37]

In summary, the specific objectives of this thesis are:

1. To investigate to which extent the spread of unhealthy behaviors and infectious diseases share similarities and how infectious disease modeling could be applied for health behavioral research.
2. To quantify the associations between social networks and smoking, sports participation and overweight, and whether these associations vary by type of social network tie.
3. To develop two agent-based models to explore the potential impact of interventions aimed at reducing socioeconomic inequalities in food consumption and sports participation.

Overview of this thesis

The first objective is addressed in *Chapter 2*, which gives an overview of several similarities between the spread of unhealthy behaviors and of infectious diseases. We also discuss the implications of the findings for the field of social epidemiology. *Chapters 3* and *4* address the second objective. *Chapter 3* looks into whether neighborhood prevalence of health-related behaviors is a risk factor of smoking, sports participation and becoming overweight in Eindhoven, the Netherlands. In *Chapter 4*, we assess the influences of social networks on smoking cessation and smoking relapse in the Netherlands using data from the *LISS* panel. *Chapters 5* and *6* address the third objective. In *Chapter 5*, we introduce the first agent-based model within the *HEBSIM* (*He*alth *B*ehavior *Sim*ulation) suite. This model describes income inequalities in food consumption, taking into account the interaction with the physical environment. It has been quantified using data from the *GLOBE* study in Eindhoven. Using this model, we assess the impact of various interventions that may reduce income inequalities in food consumption. The second model in the *HEBSIM* suite, which describes income inequalities in sports participation, is presented in *Chapter 6*. This model accounts for both interaction with the physical and social environment. In this chapter, we assess the impact of individual and environmental interventions on reducing income inequalities in

sports participation. Finally, *Chapter 7* contains a critical appraisal of the main findings of this thesis and a discussion on how to bring systems approaches in health behavioral research forward. This thesis is concluded by summaries in English and Dutch.

References

1. Beaglehole R, Bonita R, Horton R, et al. Priority actions for the non-communicable disease crisis. Lancet 2011; 377(9775): 1438-47.
2. Collaborators GBDRF, Forouzanfar MH, Alexander L, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015; 386(10010): 2287-323.
3. World Health Organization. Obesity and overweight, Fact sheet No 311. 2015. http://www.who.int/mediacentre/factsheets/fs311/en/ (accessed May 3 2016).
4. US Department of Health and Human Services. The health consequences of smoking - 50 years of progress. Rockville, MD: US Department of Health and Human Services,, 2014.
5. World Health Organization. Tobacco, Fact sheet No 339. 2015. http://www.who.int/mediacentre/factsheets/fs339/en/ (accessed May 3 2016).
6. Jha P, Peto R, Zatonski W, Boreham J, Jarvis MJ, Lopez AD. Social inequalities in male mortality, and in male mortality from smoking: indirect estimation from national death rates in England and Wales, Poland, and North America. Lancet 2006; 368(9533): 367-70.
7. Wang X, Ouyang Y, Liu J, et al. 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 2014; 349: g4490.
8. World Health Organization. Increasing fruit and vegetable consumption to reduce the risk of noncommunicable diseases. 2015. http://www.who.int/elena/titles/fruit_vehetables_ncds/en/ (accessed May 3 2016).
9. World Health Organization. Physical Inactivity: A Global Public Health Problem. 2016. http://www.who.int/dietphysicalactivity/factsheet_inactivity/en/ (accessed May 3 2016).
10. World Health Organization. WHO report on the global tobacco epidemic, 2015: Raising taxes on tobacco. Luxembourg: World Health Organization, 2015.

11. World Health Organization. Global status report on nocommunicable diseases 2014. Switzerland: WHO, 2014.
12. Moore LV, Thompson FE. Adults Meeting Fruit and Vegetable Intake Recommendations - United States, 2013: Centers fo Disease Control and Prevention, 2015.
13. Statistics Netherlands. Nederland eet onvoldoende groente, fruit en vis. 2015. https://www.cbs.nl/nl-nl/nieuws/2015/17/nederland-eet-onvoldoende-groente-fruit-en-vis (accessed May 3 2016).
14. Johnson W, Kyvik KO, Mortensen EL, Skytthe A, Batty GD, Deary IJ. Does education confer a culture of healthy behavior? Smoking and drinking patterns in Danish twins. Am J Epidemiol 2011; 173(1): 55-63.
15. Giskes K, Avendano M, Brug J, Kunst AE. A systematic review of studies on socioeconomic inequalities in dietary intakes associated with weight gain and overweight/obesity conducted among European adults. Obes Rev 2010; 11(6): 413-29.
16. Irala-Estevez JD, Groth M, Johansson L, Oltersdorf U, Prattala R, Martinez-Gonzalez MA. A systematic review of socio-economic differences in food habits in Europe: consumption of fruit and vegetables. Eur J Clin Nutr 2000; 54(9): 706-14.
17. Giskes K, Turrell G, Patterson C, Newman B. Socio-economic differences in fruit and vegetable consumption among Australian adolescents and adults. Public Health Nutr 2002; 5(5): 663-9.
18. Devaux M, Sassi F. Social inequalities in obesity and overweight in 11 OECD countries. Eur J Public Health 2013; 23(3): 464-9.
19. Mackenbach JP, Stirbu I, Roskam AJ, et al. Socioeconomic inequalities in health in 22 European countries. N Engl J Med 2008; 358(23): 2468-81.
20. Stringhini S, Sabia S, Shipley M, et al. Association of socioeconomic position with health behaviors and mortality. JAMA 2010; 303(12): 1159-66.
21. Mackenbach JP, Bos V, Andersen O, et al. Widening socioeconomic inequalities in mortality in six Western European countries. Int J Epidemiol 2003; 32(5): 830-7.
22. Giskes K, Kunst AE, Benach J, et al. Trends in smoking behaviour between 1985 and 2000 in nine European countries by education. J Epidemiol Community Health 2005; 59(5): 395-401.

23. Harper S, Lynch J. Trends in socioeconomic inequalities in adult health behaviors among U.S. states, 1990-2004. Public Health Rep 2007; 122(2): 177-89.
24. Charafeddine R, Van Oyen H, Demarest S. Trends in social inequalities in obesity: Belgium, 1997 to 2004. Prev Med 2009; 48(1): 54-8.
25. Sassi F, Devaux M, Cecchini M, Rusticelli E. The Obesity Epidemic: Analysis of Past and Projected Future Trends in Selected OECD Countries. OECD Health Working Papers 2009; 45.
26. Beenackers MA, Nusselder WJ, Oude Groeniger J, van Lenthe FJ. Het terugdringen van gezondheidsachterstanden: een systematisch overzicht van kansrijke en effectieve interventies: Department of Public Health, Erasmus MC, 2015.
27. Glass TA, McAtee MJ. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. Soc Sci Med 2006; 62(7): 1650-71.
28. Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? Ann Behav Med 2004; 27(1): 3-12.
29. Azjen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991; 50(2): 179-211.
30. Gebhardt WA, Maes S. Integrating social-psychological frameworks for health behavior research. Am J Health Behav 2001; 25(6): 528-36.
31. Tyas SL, Pederson LL. Psychosocial factors related to adolescent smoking: a critical review of the literature. Tob Control 1998; 7(4): 409-20.
32. Shaikh AR, Yaroch AL, Nebeling L, Yeh MC, Resnicow K. Psychosocial predictors of fruit and vegetable consumption in adults a review of the literature. Am J Prev Med 2008; 34(6): 535-43.
33. Craggs C, Corder K, van Sluijs EM, Griffin SJ. Determinants of change in physical activity in children and adolescents: a systematic review. Am J Prev Med 2011; 40(6): 645-58.
34. Wardle J, Steptoe A. Socioeconomic differences in attitudes and beliefs about healthy lifestyles. J Epidemiol Community Health 2003; 57(6): 440-3.

35. Droomers M, Schrijvers CT, Mackenbach JP. Educational differences in the intention to stop smoking: explanations based on the Theory of Planned Behaviour. Eur J Public Health 2004; 14(2): 194-8.
36. Susser M, Susser E. Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. Am J Public Health 1996; 86(5): 674-7.
37. Sallis JF, Owen N, Fisher EB. Ecological modesl of health behavior. In: Glanz K, Rimer BK, Viswanath K, eds. Health behavior and health education. 4th ed. San Francisco: Jossey-Bass; 2008.
38. Wilson LM, Avila Tang E, Chander G, et al. Impact of tobacco control interventions on smoking initiation, cessation, and prevalence: a systematic review. J Environ Public Health 2012; 2012: 961724.
39. Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? Obes Rev 2011; 12(5): e95-e106.
40. Sallis JF, Cerin E, Conway TL, et al. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. Lancet 2016.
41. Berkman LF, Krishna A. Social network epidemiology. In: Berkmand LF, Kawachi I, Glymour M, eds. Social epidemiology. New York, NY: Oxford University Press; 2014.
42. May S, West R. Do social support interventions ("buddy systems") aid smoking cessation? A review. Tob Control 2000; 9(4): 415-22.
43. Wing RR, Jeffery RW. Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance. J Consult Clin Psychol 1999; 67(1): 132-8.
44. Kamphuis CB, Mackenbach JP, Giskes K, Huisman M, Brug J, van Lenthe FJ. Why do poor people perceive poor neighbourhoods? The role of objective neighbourhood features and psychosocial factors. Health Place 2010; 16(4): 744-54.
45. van Lenthe FJ, Brug J, Mackenbach JP. Neighbourhood inequalities in physical inactivity: the role of neighbourhood attractiveness, proximity to local facilities and safety in the Netherlands. Soc Sci Med 2005; 60(4): 763-75.

46. Prins RG, van Empelen P, Te Velde SJ, et al. Availability of sports facilities as moderator of the intention-sports participation relationship among adolescents. Health Educ Res 2010; 25(3): 489-97.
47. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Soc Sci Med 2000; 51(6): 843-57.
48. Weiss RS. The provisions of social relationships. In: Rubin Z, ed. Doing unto others. Englewood Cliffs, NJ: Prentice Hall; 1974.
49. Choi HJ, Smith RA. Members, isolates, and liaisons: meta-analysis of adolescents' network positions and their smoking behavior. Subst Use Misuse 2013; 48(8): 612-22.
50. Granovetter M. The strength of weak ties. Am J Sociol 1973; 78: 1360-80.
51. Marsden PV, Friedkin NE. Network studies of social influence. In: Wasserman S, Galaskiewicz J, eds. Advances in social network analysis: research in the social and behavioral sciences. Thousand Oaks, CA: Sage; 1994.
52. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007; 357(4): 370-9.
53. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008; 358(21): 2249-58.
54. Cohen-Cole E, Fletcher JM. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. J Health Econ 2008; 27(5): 1382-7.
55. Shalizi CR, Thomas AC. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. Sociol Methods Res 2011; 40(2): 211-39.
56. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Soc 2001; 27: 415-44.
57. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. Stat Med 2013; 32(4): 556-77.
58. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology 2011; 22(1): 42-52.
59. Snijders TAB, Van de Bunt GG, Steglich CEG. Introduction to stochastic actor-based models for network dynamics. Social Networks 2010; 32: 44-60.

60. Hill AL, Rand DG, Nowak MA, Christakis NA. Infectious disease modeling of social contagion in networks. PLoS Comput Biol 2010; 6(11): e1000968.
61. Valente TW. Network interventions. Science 2012; 337(6090): 49-53.
62. Campbell R, Starkey F, Holliday J, et al. An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial. Lancet 2008; 371(9624): 1595-602.
63. Valente TW, Ritt-Olson A, Stacy A, Unger JB, Okamoto J, Sussman S. Peer acceleration: effects of a social network tailored substance abuse prevention program among high-risk adolescents. Addiction 2007; 102(11): 1804-15.
64. Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health 2011; 101(9): 1627-34.
65. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. Annu Rev Public Health 2012; 33: 357-76.
66. Resnicow K, Page SE. Embracing chaos and complexity: a quantum change for public health. Am J Public Health 2008; 98(8): 1382-9.
67. Gallagher R, Appenzeller T. Beyond Reductionism. Science 1999; 28.
68. El-Sayed AM, Scarborough P, Seemann L, Galea S. Social network analysis and agent-based modeling in social epidemiology. Epidemiol Perspect Innov 2012; 9(1): 1.
69. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. Am J Epidemiol 2008; 168(1): 1-8.
70. Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol 2010; 39(1): 97-106.
71. Speybroeck N, Van Malderen C, Harper S, Muller B, Devleesschauwer B. Simulation models for socioeconomic inequalities in health: a systematic review. Int J Environ Res Public Health 2013; 10(11): 5750-80.
72. Galea S, Hall C, Kaplan GA. Social epidemiology and complex system dynamic modelling as applied to health behaviour and drug use research. Int J Drug Policy 2009; 20(3): 209-16.
73. Marshall BD, Galea S. Formalizing the role of agent-based modeling in causal inference and epidemiology. Am J Epidemiol 2015; 181(2): 92-9.
74. Law AM, Kelton WD. Simulation modeling & analysis: McGraw-Hill; 1991.

75. Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 2002; 99 Suppl 3: 7280-7.
76. Abdou M, Hamill L, Gilbert N. Designing and Building an Agent-Based Model. In: Heppenstall AJ, Crooks AT, See LM, Batty M, eds. Agent-Based Models of Geographical Systems: Springer; 2012: 141-65.
77. Coffeng LE, Stolk WA, Hoerauf A, et al. Elimination of African onchocerciasis: modeling the impact of increasing the frequency of ivermectin mass treatment. PLoS One 2014; 9(12): e115886.
78. Fischer EA, de Vlas SJ, Habbema JD, Richardus JH. The long-term effect of current and new interventions on the new case detection of leprosy: a modeling study. PLoS Negl Trop Dis 2011; 5(9): e1330.
79. van der Ploeg CPB, Van Vliet C, De Vlas SJ, et al. STDSIM: A microsimulation model for decision support in STD control. Interfaces 1998; 28: 84-100.
80. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. A spatial agent-based model for the simulation of adults' daily walking within a city. Am J Prev Med 2011; 40(3): 353-61.
81. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. Exploring walking differences by socioeconomic status using a spatial agent-based model. Health Place 2012; 18(1): 96-9.
82. Zhang D, Giabbanelli PJ, Arah OA, Zimmerman FJ. Impact of different policies on unhealthy dietary behaviors in an urban adult population: an agent-based simulation model. Am J Public Health 2014; 104(7): 1217-22.
83. Zhang J, Tong L, Lamberson PJ, Durazo-Arvizu RA, Luke A, Shoham DA. Leveraging social influence to address overweight and obesity using agent-based models: the role of adolescent social networks. Soc Sci Med 2015; 125: 203-13.
84. Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential segregation. Am J Prev Med 2011; 40(3): 303-11.

2

UNHEALTHY BEHAVIOR IS CONTAGIOUS: AN INVITATION TO EXPLOIT MODELS FOR INFECTIOUS DISEASES

D.J. BLOK, P. VAN EMPELEN, F.J. VAN LENTHE, J.H. RICHARDUS, S.J. DE VLAS

Published in: Epidemiology & Infection (2013), Volume 141, Issue 3, Pages 667-669

Abstract

We argue that the spread of unhealthy behavior shows marked similarities with infectious diseases. It is therefore interesting and challenging to use infectious disease methodologies for studying the spread and control of unhealthy behavior. This would be a great addition to current methods, because it allows taking into account the dynamics of individual interactions and the social environment at large. In particular, the application of individual-based modeling holds great promise to address some major public health questions.

Over the years many theories have been developed to explain why people engage in certain unhealthy behaviors and how these spread in populations.[1] These theories share the idea that behavior is in some way influenced by social contacts. Yet, empirical studies of unhealthy behaviors generally investigate behavioral change processes from an individual perspective and until recently paid little attention to social environmental influences on behavior. An intriguing exception is work from *Christakis & Fowler*,[2,3] who showed that both smoking and obesity spread from person-to-person, that the type of contact matters, and that groups can be distinguished within a social network. This lead to the idea that unhealthy behavior is contagious and that it spreads in populations like an infectious disease. This has been suggested before conceptually,[4] but there is a need to further operationalize this concept in ways that can be tested scientifically. Basically, adopting unhealthy behavior is analogous to acquiring, say, influenza from a family member. Moreover, influenza tends to cluster in schools, which can also be observed for unhealthy behaviors. Tuberculosis and leprosy are even better examples of infectious diseases that show similarities with unhealthy behaviors: they cluster in households and communities, only a minority of those exposed eventually develop disease, and clinical signs may not be visible until several years after infection. Although there is considerable evidence that the spread of behaviors is explained by social influence, it is also true that similarity of behaviors observed in social networks may to some extent be the result of the tendency of people to select others with similar behaviors (homophily). Yet, it is difficult to disentangle homophily from social influence.[5,6]

Apart from contagiousness, other concepts and underlying mechanisms can be identified that are comparable for unhealthy behavior and infectious diseases. First, an important concept in infectious diseases is heterogeneity, which can concern individual susceptibility to infection, infectiousness of a patient, and mixing patterns in the population.[7] Heterogeneity in susceptibility resembles variation in adopting unhealthy behaviors, such as stated in the theory of *Diffusion of Innovations*,[8] which indicates that some people are more susceptible to adopt a behavior than others. The rate of adoption further depends on the number of people in the social network that engage in a certain behavior. Each individual has his/her own adoption threshold. For instance, some people are more self-

efficacious than others, resulting in different levels of resilience. Heterogeneity in infectiousness can be compared with variation in social influence: position within networks, closeness of relationships, and number of contacts may explain why some people are more influential than others.[9] Heterogeneity concerning mixing patterns reflects that individuals tend to cluster within populations, e.g. according to age group or socioeconomic position. Second, a mechanism strongly related to heterogeneity is the presence of so-called super-spreaders. These are individuals that accelerate dissemination of an infection in a population, because of a prominent role in the contact network (i.e. many contacts) and/or high infectiousness. This greatly resembles opinion leaders or peer-role models, which are early adopters and can easily spread behaviors to others, due to their persuasiveness and high number of social contacts.[8] Third, vaccination is another concept that both fields share. Vaccination induces immunity, reduces the number of susceptible people, and reduces the risk of infectious diseases. In a similar way, social inoculation provides resistance to unhealthy behavior by emphasizing refusal skills, and thus reducing the risk of adopting a behavior.[10] Although vaccination and social inoculation are not exactly the same, they serve the same purpose. A fourth comparable mechanism is the influence of physical environmental factors. The physical environment promotes or discourages the spread of infections and behaviors in social networks through, e.g. climate and availability of fast-food, respectively. However, the availability of fast-food can also trigger a person to start unhealthy eating without any influence from the social environment.

The fact that the principles of infectious diseases and unhealthy behavior show a remarkable resemblance challenges us to study unhealthy behavior as an infectious disease. Infectious disease epidemiology has been studied for decades using sophisticated methods, in particular mathematical modeling, to analyze spread within populations, to predict the course of epidemics, and to evaluate interventions. As a major innovative step, *Hill et al.*[11] recently modeled the obesity epidemic as an infectious disease, using data from the Framingham Heart Study cohort.[12] The model mimics transitions between two compartments, i.e. susceptible (non-obese) and infected (obese) individuals. It also allows for possible spontaneous infections not resulting from contacts. The study concludes

that the obesity epidemic is driven by both contagious and spontaneous infection and will stabilize at 42% of the population being obese within the next 50 years. However, as the authors indicate in an earlier paper,[13] the proposed compartmental model is rather simplistic and does not take into account possible heterogeneities.

A major enhancement would be to go from compartmental modeling to a more comprehensive and realistic approach. Individual-based modeling is particularly useful to realistically model networks and individual heterogeneities. It simulates life-histories of individuals and specific interactions between individuals over time. Events, such as birth, death, relationship formation, transfer between social/risk groups, and acquisition of infection (behavior), are modeled through chance processes. Another advantage is that it is more suitable for analyzing the impact of interventions aimed at certain groups, such as households or schools. Individual-based modeling has proven to be very useful for practical decision making in infectious disease control, starting with the *ONCHOSIM* model for river blindness control in West Africa.[14,15] A more relevant model for sexually transmitted diseases, *STDSIM*, explicitly models individual contacts (sexual relationships) and formation of (sexual) networks.[16] Another recent example is the *SIMCOLEP* model for leprosy,[17] in which the formation of and movement between households is modeled.

The application of individual-based modeling holds great promise to address some of the major questions in public health regarding health-related behaviors. Why are some people more open for unhealthy behaviors than other people? What are major determinants causing the adoption of certain behaviors? How can we best prevent unhealthy behavior or promote behavioral change? These questions can only be answered adequately when taking into account the social context in which behaviors take place. Until now behavioral studies have mainly focused on the individual in a static environment. The introduction of infectious disease methodology and in particular individual-based modeling would be a great addition, because it takes into account the dynamics of individual interactions and the social environment at large. This may result in new or revised interventions and policies. For instance, community interventions for behavioral change that only show small individual effects may eventually have substantial indirect public health effects. In contrast, some interventions with large individual

effects may ultimately have a small impact on the population, due to a limited reach. Individual-based modeling in particular allows translating individual effects to population impact. Moreover, infectious disease modeling provides useful key concepts, such as the basic reproduction number (R_0), i.e. the average number of successful transmissions per infectious person in a fully susceptible population. An outbreak of, e.g. smoking in a non-smoking population will occur if $R_0>1$, which indicates that each smoker will on average trigger at least one other individual to start smoking. The goal is to reduce R_0 to below 1, to stop further spreading of smoking.

In conclusion, the spread of unhealthy behavior shows marked similarities with infectious diseases, and hence embracing existing infectious disease methods is beneficial. A first attempt to apply infectious disease modeling for unhealthy behaviors has now been published, but there is substantial room for improvement by including the dynamics and heterogeneities of social networks. The field of research aimed at studying health-related behaviors and at developing interventions and policies to promote health behaviors may benefit substantially from further exploiting models for infectious diseases, in particular individual-based models.

References

1. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010; 31: 399-418.
2. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007; 357(4): 370-9.
3. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008; 358(21): 2249-58.
4. Gladwell M. The tipping point: How little things can make a big difference. New York: Little, Brown and Company; 2000.
5. Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. Sociol Method Res 2011; 40(2): 211-39.
6. VandeWeele TJ. Sensitivity analysis for contagion effects in social networks. Sociol Method Res 2011; 40(2): 240-55.
7. Anderson RM, May RM. Infectious Diseases of Humans: Dynamics and Control: Oxford University Press; 1991.
8. Rogers EM. Diffusion of Innovations. 5th ed: New York: The Free Press; 2003.
9. Valente TW. Social Networks and Health: Oxford University Press; 2010.
10. Evans RI, Raines BE, Hanselka L. Developing data-based communications in social psychological research: adolescent smoking prevention. J Appl Soc Psych 1984; 14: 289-95.
11. Hill AL, Rand DG, Nowak MA, Christakis NA. Infectious disease modeling of social contagion in networks. PLoS Comput Biol 2010; 6(11): e1000968.
12. Dawber TR. The Framingham Study: the epidemiology of atherosclerotic disease. Cambridge, MA: Harvard Univ. Press; 1980.
13. Hill AL, Rand DG, Nowak MA, Christakis NA. Emotions as infectious diseases in a large social network: the SISa model. Proc Biol Sci 2010; 277(1701): 3827-35.
14. Plaisier AP, van Oortmarssen GJ, Habbema JD, Remme J, Alley ES. ONCHOSIM: a model and computer simulation program for the transmission and control of onchocerciasis. Comput Methods Programs Biomed 1990;

31(1): 43-56.
15. Remme JH. Research for control: the onchocerciasis experience. Trop Med Int Health 2004; 9(2): 243-54.
16. Korenromp EL, Van Vliet C, Bakker R, De Vlas SJ, Habbema JDF. HIV spread and partnership reduction for different patterns of sexual behaviour - a study with the microsimulation model STDSIM. Math Pop Studies 2000; 8: 135-73.
17. Fischer E, De Vlas S, Meima A, Habbema D, Richardus JH. Different mechanisms for heterogeneity in leprosy susceptibility can explain disease clustering within households. PLoS ONE 2010; 5(11).

3

CHANGES IN SMOKING, SPORTS PARTICIPATION AND OVERWEIGHT: DOES NEIGHBORHOOD PREVALENCE MATTER?

D.J. BLOK, S.J. DE VLAS, P. VAN EMPELEN, J.H. RICHARDUS, F.J. VAN LENTHE

Published in: Health & Place (2013), Volume 23, Pages 33-38

Abstract

We investigated whether the prevalence of health-related behaviors and overweight in neighborhoods is associated with changes in smoking, sports participation and overweight over 13 years of follow-up in Dutch adults residing in 86 neighborhoods of Eindhoven in 1991. We showed that living in neighborhoods with a high prevalence of non-smoking, no sports participation and overweight increased the odds of quitting smoking, quitting sports and becoming overweight. After adjustments for age, gender, education and neighborhood deprivation this association remained significant for becoming overweight. Neighborhood prevalence of health-related behaviors and overweight appears to be a currently neglected but relevant determinant of changes in health-related behaviors.

Introduction

The relevance of area characteristics for health and health-related behaviors is now well accepted.[1-4] Studies have shown that neighborhood characteristics are associated with overweight and health-related behaviors, such as smoking and physical inactivity.[5-7] In search for specific contextual determinants of these behaviors, much emphasis has been placed on physical environmental characteristics, such as accessibility and availability of facilities.[6,8] The importance of the social environment has also been considered: neighborhood social cohesion is for example commonly linked to physical activity.[9-12] A contextual determinant of health behaviors that has surprisingly little been studied is the prevalence of health-related behaviors.

The reasoning behind examining neighborhood prevalence of health-related behaviors stems from the idea that healthy and unhealthy behaviors spread from person-to-person. Individuals interact with each other and therefore influence other people's behavior, for example through peer pressure, conscious or unconscious copying of behavior (social mimicry).[13,14] This has been supported by behavioral theories as well as empirical evidence. For instance, according to the *Social Learning Theory*[15] people may quit smoking, because they watch other people in their environment quit smoking and consider this behavior favorable (observational learning). According to the theory of *Diffusion of Innovations*[16] adoption of a certain behavior spreads through social networks and depends on the number of people in the environment that engage in a certain behavior. In addition, empirical studies have shown that both smoking cessation and obesity (or norms associated with obesity) spread from person-to-person in a social network.[17,18]

The purpose of this study is to investigate the association between neighborhood prevalence of health-related behaviors at baseline and changes in these behaviors during follow-up. In order to rule out neighborhood level confounding,[19] adjustments will be made for neighborhood deprivation, because it is associated with neighborhood prevalence of health-related behaviors as well as behavioral change over time and may capture physical and social contextual factors related to deprivation.[5,7]

The Dutch *GLOBE* study is a prospective cohort study, which provides information on smoking, sports participation and overweight for a large sample over a period of 13 years in the city of Eindhoven.[20] It provides a unique opportunity to investigate the importance of the prevalence of health-related behaviors and overweight for subsequent (behavioral) changes. In three different studies we hypothesized that: (1) smokers living in a neighborhood with a high prevalence of non-smokers are more likely to quit during follow up, (2) participants in sports living in neighborhoods with a high prevalence of persons not participating in sports are more likely to quit during follow up and (3) normal weight persons living in neighborhoods with a high prevalence of overweight are more likely to become overweight during follow up.

Methods

Study population

Longitudinal data were obtained from the Dutch prospective *GLOBE* study. The area of study included the city of Eindhoven, which was the fifth largest city of The Netherlands with approximately 135,000 inhabitants between the age of 15 and 75 years in 1991. The city has 116 neighborhoods, of which 86 are predominantly residential neighborhoods.

Baseline data were collected in 1991 using postal questionnaires. An a-select sample of 27,070 non-institutionalized subjects between the age of 15 and 75 years living in or near the city of Eindhoven were selected to participate. The response was 70.1%, resulting in 18,973 respondents,[20] of which 10,450 persons resided in 86 neighborhoods of the city of Eindhoven. On average, these neighborhoods had 121.5 respondents (min=5, max=386). In the wave of data collection in 2004, an additional subsample was invited of participants who resided in the city of Eindhoven in 1991 and who still resided there in 2004.[21] This resulted in 2837 respondents living in the city of Eindhoven in both 1991 and 2004.

Only non-institutionalized respondents with valid measurements (i.e. no missing or impossible values) on the outcomes in 1991 and 2004 were included. The age range at follow-up was 28–88 years. Three studies were conducted: (1) smoking cessation, (2) quitting sports participation, and (3) becoming overweight. The

study population of the three studies consisted of (1) smokers at baseline (n=760), (2) respondents who participated in sports at baseline (n=1317) and (3) respondents without overweight at baseline (n=1674).

Measures

Smoking status, sport participation and overweight status were obtained from the 1991 and 2004 postal questionnaires. All remaining measures were obtained from the baseline (1991) postal questionnaire.

Smoking status

Self-reported smoking status was measured by asking respondents the following question: "Do you smoke?" Respondents could answer with "yes", "no, but used to smoke" (former smokers), and "never smoked before". Based on information about the amount of cigarettes smoked per day, those reporting at least 1 cigarette per day were considered smokers. The outcome of interest was change in smoking status, i.e. quitting smoking. Answers were categorized into: unchanged behavior (continuing smoking) and quitting smoking.

Sports participation

In 1991, sports participation was measured through a single question: "Do you participate in sports?" Respondents could answer with "no", "yes, <1 hour/week", "yes, 1–2 hours/week", "yes, >2 hours/week". In 2004, the standardized and validated SQUASH questionnaire was used.[22] Respondents could record up to four different sport activities in an open question. For each activity, the frequency, the average duration and the intensity were reported. Sports participation was dichotomized into "yes" for respondents who participated in sports weekly and "no" for those who did not participate in sports weekly. The outcome of interest was the change in sports participation. Answers were categorized into: unchanged (continuing doing sports) and quitting sports participation.

Overweight status

Information about body height (cm) and body weight (kg) was obtained through self-reported open questions. Body mass index (BMI) was calculated as: weight

(kg)/height (m)^2. Overweight was defined as BMI≥25.[23] The outcome of interest was change in overweight status. Answers were categorized into: unchanged (no overweight), becoming overweight.

Neighborhood prevalence of non-smoking, no sports participation and overweight

Calculations for neighborhood prevalence of non-smoking (i.e. former and never smokers together), no sports participation and overweight were based on the total eligible population at baseline that lived in the city of Eindhoven. This included 10,239 respondents for smoking, 10,298 respondents for sports participation, and 10,092 respondents for overweight. Based on these sample sizes, unstandardized prevalence rates of non-smoking, no sports participation and overweight at baseline were calculated for the 86 neighborhoods in the city of Eindhoven. Neighborhood prevalence of non-smoking ranged from 40.5% to 90%, no sports participation from 0% to 78.2% and overweight from 10% to 60%. These prevalence rates were not standardized, because the proposed mechanism is based on what individuals experience in their environment without taking age and sex distributions into account. This measure was further categorized into quartiles, each with 25% of neighborhoods (see footnote in Table 1).

Age, gender and education

Respondents provided information on age, gender, and educational level at baseline. Educational level was measured by self-reported questions about the respondent's highest attained level of education. Responses were categorized as follows: lower (primary and lower secondary), middle (higher secondary), higher (tertiary). Educational level has proven to be a good indicator of socioeconomic status in the Netherlands.[7,24]

Neighborhood deprivation

Neighborhood deprivation at baseline is measured following van *Lenthe and Mackenbach (2002)*.[7] Neighborhoods were ranked based on: the percentage of subjects with primary school as highest attained educational level per neighborhood (mean=22.6%, min=0%, max=44.1%); the percentage of subjects that are unskilled manual workers per neighborhood (mean=15.1%, min=0%, max=31%);

the percentage unemployed subjects per neighborhood (mean=11.4%, min=0%, max=29.1%). Quartiles were constructed using the summed rankings.[7]

Statistical analysis

To examine the association between neighborhood prevalence of non-smoking, no sports participation and overweight and subsequent (behavioral) changes, multilevel modeling was used. By doing so, the hierarchical structure of the data, where individuals (level 1) were nested in neighborhoods (level 2), was taken into account. First unadjusted analyses were performed to identify the association between (1) quartiles of neighborhood prevalence of non-smoking at baseline and quitting smoking during follow up, (2) quartiles of neighborhood prevalence of no sports participation and quitting sports, and (3) quartiles of neighborhood prevalence of overweight and becoming overweight. The lowest quartile of neighborhood prevalence rates was taken as reference category. In a subsequent model age, gender and educational level were added as possible confounders. The final model was additionally adjusted for neighborhood deprivation. To test whether a linear trend was present between neighborhood prevalence rates and subsequent changes, similar analyses were conducted treating neighborhood prevalence of non-smoking, no sports participations and overweight as continuous variables. All analyses were performed with the statistical package R (version 2.15.0), using the lme4 package.[25,26]

Results

Characteristics

Table 1 shows the baseline characteristics and the percentage of respondents who changed behaviors at follow-up for each study. In the total study population of each study more than one-third changed behaviors between 1991 and 2004 (study 1: 43.7%, study 2: 37.8%, and study 3: 34.5%). The mean age of participants in each study was approximately 45 years and the majority was classified as low educated at baseline. In each of the three studies, an increasing percentage of people changing behaviors were found by increasing quartiles of neighborhood prevalence rates at baseline. In all studies, the lowest quartile corresponded with the lowest percentage of behavioral change.

Table 1. Baseline characteristics and changes in smoking, sports participation and overweight during follow-up

	Study 1: Smoking cessation			Study 2: Quitting sports participation			Study 3: Becoming overweight		
	Total	Quitting smoking		Total	Quitting sports		Total	Overweight	
	N	N	%	N	N	%	N	N	%
Total sample	760	332	43.7	1317	498	37.8	1674	577	34.5
Neighborhoods	79			79			84		
Neighborhood prevalence rates[a]									
1–Lowest	179	68	38.0	229	55	24.0	285	79	27.7
2	171	70	40.9	466	176	37.8	542	202	37.3
3	314	150	47.8	410	177	43.2	533	189	35.5
4–Highest	96	44	45.8	212	90	42.5	314	107	34.1
Age, mean (sd)	44.8 (12.8)			45.3 (13.8)			44.7 (14.5)		
Gender									
Male	378	166	43.9	614	213	34.7	724	278	38.4
Female	382	166	43.5	703	285	40.5	950	299	31.5
Education									
1–Lower	455	193	42.4	655	293	44.7	823	287	34.9
2	171	72	42.1	321	115	35.8	415	152	36.6
3–Higher	134	67	50.0	341	90	26.4	436	138	31.7
Neighborhood deprivation									
1–Least deprived	119	56	47.1	292	73	25.0	323	112	34.7
2	240	111	46.3	423	158	37.4	558	187	33.5
3	180	75	41.7	295	129	43.7	384	142	37.0
4–Most deprived	221	90	40.7	307	138	45.0	409	136	33.3

[a] *study 1: neighborhood prevalence of non-smoking (quartile range: low [40.5% – 57.8%], 2 [57.8% – 63.5%], 3 [63.5% – 67.8%], high [67.8% – 90.0%]); study 2: neighborhood prevalence of no sports participation (quartile range: low [0% – 46.5%], 2 [46.5% – 55.8%], 3 [55.8% – 62.8%], high [62.8% – 78.2%]); study 3: neighborhood prevalence of overweight (quartile range: low [10.0% – 30.0%], 2 [30.0% – 36.4%], 3 [36.4% – 42.3%], high [42.3% – 60.0%])*

Study 1: Smoking cessation

Unadjusted results showed that odds ratios of quitting smoking during follow-up increased with higher neighborhood prevalence of non-smoking at baseline (see Table 2). Differences were only statistically significant between the lowest and second-highest quartile of neighborhood prevalence of non-smoking. People living in the second-highest neighborhood prevalence quartile showed a 49% (95% CI: 1.03–2.17) higher odd to quit smoking compared to those in the lowest quartile. After adjustments for age, gender, education as well as neighborhood deprivation, associations were no longer statistically significant. A significant trend was only present in the unadjusted regression analysis (P=0.045).

Table 2. Unadjusted and adjusted associations between neighborhood prevalence of non-smoking at baseline and quitting smoking during follow-up

Dependent variable: quitting smoking	Model 1				Model 2				Model 3			
N = 760	OR	95% CI		P	OR	95% CI		P	OR	95% CI		P
Neighborhood prevalence of non-smoking												
1–Lowest	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
2	1.13	0.74	1.74	0.573	1.08	0.70	1.67	0.718	1.08	0.69	1.70	0.737
3	1.49	1.03	2.17	0.036	1.38	0.94	2.02	0.099	1.38	0.89	2.13	0.152
4–Highest	1.38	0.84	2.28	0.208	1.23	0.73	2.05	0.436	1.25	0.70	2.22	0.446
P-trend[a]				0.045				0.165				0.229
Age	-	-	-	-	1.02	1.01	1.03	0.004	1.02	1.01	1.03	0.004
Gender												
Male	-	-	-	-	1.00	-	-	-	1.00	-	-	-
Female	-	-	-	-	1.13	0.83	1.54	0.426	1.13	0.83	1.54	0.428
Education												
1–Lower	-	-	-	-	0.67	0.44	1.01	0.053	0.66	0.43	1.02	0.062
2	-	-	-	-	0.72	0.45	1.14	0.161	0.71	0.44	1.15	0.164
3–Higher	-	-	-	-	1.00	-	-	-	1.00	-	-	-
Neighborhood deprivation												
1–Least deprived	-	-	-	-	-	-	-	-	1.00	-	-	-
2	-	-	-	-	-	-	-	-	1.06	0.65	1.73	0.808
3	-	-	-	-	-	-	-	-	1.04	0.60	1.79	0.896
4–Most deprived	-	-	-	-	-	-	-	-	1.03	0.60	1.78	0.904

[a] *P-trend: neighborhood prevalence rate of non-smoking treated as continuous variable*

Study 2: Quitting sports participation

Odds ratios of quitting sports increased with higher neighborhood prevalence of no sports participation (see Table 3). Model 1 showed statistically significant differences for all quartiles compared to the lowest quartile. Also, after adjustments for age, gender and education, results were still statistically significant (second vs. lowest: odds ratio=1.92, 95% CI: 1.34–2.75; third vs. lowest: odds ratio=2.16, 95% CI: 1.50–3.12; highest vs. lowest: odds ratio=2.21, 95% CI: 1.46–3.34). However, after further adjustment for neighborhood deprivation, this association no longer remained statistically significant. A significant linear trend was observed in the unadjusted model (P=0.000) and the model that was adjusted for age, gender and education (P=0.001).

Table 3. Unadjusted and adjusted associations between neighborhood prevalence of no sports participation at baseline and quitting sports during follow-up

Dependent variable: quitting sports	Model 1				Model 2				Model 3			
N = 1317	OR	95% CI		P	OR	95% CI		P	OR	95% CI		P
Neighborhood prevalence of no sports participation												
1–Lowest	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
2	1.92	1.34	2.74	0.000	1.92	1.34	2.75	0.000	1.47	0.93	2.33	0.098
3	2.40	1.68	3.45	0.000	2.16	1.50	3.12	0.000	1.38	0.81	2.35	0.234
4–Highest	2.33	1.55	3.51	0.000	2.21	1.46	3.34	0.000	1.33	0.74	2.39	0.342
P- trend[a]				0.000				0.001				0.995
Age	-	-	-	-	1.01	1.00	1.02	0.005	1.01	1.00	1.02	0.003
Gender												-
Male	-	-	-	-	1.00	-	-	-	1.00	-	-	-
Female	-	-	-	-	1.12	0.88	1.41	0.369	1.12	0.88	1.42	0.350
Education												
1–Lower	-	-	-	-	1.84	1.35	2.50	0.000	1.74	1.27	2.39	0.001
2	-	-	-	-	1.49	1.06	2.09	0.021	1.46	1.04	2.06	0.029
3–Higher	-	-	-	-	1.00	-	-	-	1.00	-	-	-
Neighborhood deprivation												
1–Least deprived	-	-	-	-	-	-	-	-	1.00	-	-	-
2	-	-	-	-	-	-	-	-	1.38	0.90	2.12	0.142
3	-	-	-	-	-	-	-	-	1.69	1.02	2.79	0.041
4–Most deprived	-	-	-	-	-	-	-	-	1.87	1.10	3.15	0.020

[a] *P-trend: neighborhood prevalence rate of no sports participation treated as continuous variable*

Study 3: Becoming overweight

The proportion of normal weight persons becoming overweight increased with neighborhood prevalence of overweight (see Table 4). Model 1 showed significant differences between the lowest and second-lowest quartile (odds ratio=1.57, 95% CI: 1.12–2.19) and between the lowest and second-highest quartile (odds ratio=1.43, 95% CI: 1.02–1.99). After adjustments for age, gender, education and neighborhood deprivation the difference between the lowest and second-lowest quartile remained statistically significant (odds ratio=1.50; 95% CI: 1.06–2.12). There was no evidence for a linear trend in all models.

Table 4. Unadjusted and adjusted associations between neighborhood prevalence of overweight at baseline and becoming overweight during follow-up

Dependent variable: overweight	Model 1				Model 2				Model 3			
N = 1674	OR	95% CI		P	OR	95% CI		P	OR	95% CI		P
Neighborhood prevalence of overweight												
1–Lowest	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
2	1.57	1.12	2.19	0.009	1.50	1.07	2.12	0.019	1.50	1.06	2.12	0.020
3	1.43	1.02	1.99	0.039	1.37	0.97	1.92	0.073	1.39	0.96	2.01	0.077
4–Highest	1.34	0.93	1.94	0.118	1.27	0.87	1.85	0.215	1.29	0.83	2.00	0.253
P- trend[a]				0.312				0.505				0.515
Age	-	-	-	-	1.00	0.99	1.00	0.272	1.00	0.99	1.00	0.261
Gender												
Male	-	-	-	-	1.00	-	-	-	1.00	-	-	-
Female	-	-	-	-	0.69	0.56	0.86	0.001	0.69	0.56	0.86	0.001
Education												
1–Lower	-	-	-	-	1.29	0.98	1.70	0.068	1.30	0.99	1.72	0.058
2	-	-	-	-	1.31	0.98	1.75	0.067	1.32	0.99	1.76	0.063
3–Higher	-	-	-	-	1.00	-	-	-	1.00	-	-	-
Neighborhood deprivation												
1–Least deprived	-	-	-	-	-	-	-	-	1.00	-	-	-
2	-	-	-	-	-	-	-	-	0.91	0.66	1.25	0.555
3	-	-	-	-	-	-	-	-	1.00	0.69	1.47	0.988
4–Most deprived	-	-	-	-	-	-	-	-	0.86	0.59	1.26	0.440

[a] *P-trend: neighborhood prevalence rate of overweight treated as continuous variable*

Discussion

To the best of knowledge, this study is the first to explore the prevalence of health-related behaviors and overweight as contextual determinants of behavioral change and becoming overweight. Results showed that living in a neighborhood with a high prevalence of non-smoking, no sports participation and overweight, increases the odds to subsequent changes in behavior. However, smoking cessation was only significantly related to neighborhood prevalence rates in the unadjusted model. After adjustments for age, gender and education only quitting sports participation and becoming overweight showed significant results. Further adjustment for neighborhood deprivation resulted only in a significant association between becoming overweight and neighborhood prevalence of overweight. It appears that for becoming overweight, the neighborhood prevalence rate is more important than for quitting smoking or quitting sports.

Unlike smoking and sports participation, overweight did not show statistically significant results for the p-trend test, indicating that the odds of becoming overweight did not increase linear with neighborhood prevalence rates of overweight. The pattern of the quartiles of neighborhood prevalence of overweight indicates that a threshold effect may be present. To test this, neighborhood prevalence of overweight was dichotomized using the first quartile (prevalence: 30%) as a cut-off point. Multilevel logistic regressions were conducted using the new variable with the first quartile as reference category (results not shown). Statistically significant results were obtained in the unadjusted model (odds ratio=1.46, 95% CI: 1.08–1.96), the model adjusted for age, gender and education (odds ratio=1.39, 95% CI: 1.03–1.89) and the model adjusted for age, gender, education and neighborhood deprivation (odds ratio=1.45, 95% CI: 1.05–2.00). These results suggest that starting from a 30% prevalence threshold, being exposed to additional overweight persons will not add to a shift in norm and hence does not increase the odds of becoming overweight.

Although evidence is weak, results support the hypothesis that people living in neighborhoods with high prevalence rates of health-related behaviors and overweight are more likely to change behaviors accordingly. This study is the first that examines this mechanism from a geographic perspective, disregarding

the social distance between people. Literature suggests that social distance or closeness of ties is more important than geographic distance in explaining behavioral change.[17,18] However, this study suggests that the prevalence in neighborhoods alone may already be sufficient to change health-related behaviors. Although neighborhood relationships are usually weaker than other active ties, and rarely socially close, these ties are a sizeable percentage of the people with whom they are frequently in contact.[27]

In order to find further support for the mechanism investigated, additional analyses also examined the association of healthy instead of unhealthy behaviors in a neighborhood context (results not shown). Analyses in which losing weight during follow-up was related to neighborhood prevalence of normal weight at baseline and in which initiation of sports was related to neighborhood prevalence of sports participation showed similar, but weaker associations. It suggests that neighborhood prevalence as a contextual determinant is perhaps more important for unhealthy behaviors than healthy behaviors, possibly because it may be easier to adopt unhealthy behaviors than healthy behavior.[28,29]

The current study can be extended by considering neighborhood prevalence rates of behavioral change instead of behaviors. Thus, hypothesizing that people living in a neighborhood with high prevalence rates of behavioral change will more likely copy this change in behavior. The current dataset only allows testing this hypothesis for the study smoking cessation. Using similar methods as in the original study, the association between neighborhood prevalence of ex-smoking (former smokers) at baseline and quitting smoking during follow-up was investigated (results not shown). Similar to study 1, results were only statistically significant in the unadjusted model (highest vs. lowest: odds ratio=1.58, 95% CI: 1.00–2.49). Although the relation is weak, there is some support for this extended hypothesis. This also weakly suggests that smoking cessation may be contagious at a neighborhood level, which has been suggested before in a paper.[30]

One potential limitation of this study is the use of self-reported data for smoking, BMI (overweight) and sports participation. It can be expected that reports are underestimated or overestimated.[31-33] In particular, self-reported data on body weight can result in an underestimation of the real body weight.[32] However, the

bias would affect the exposure and outcome in a similar way, which seems to make this less of a problem. A second limitation is that sports participation was measured differently at baseline and follow-up. At baseline, no validated questionnaire was available and at follow up the SQUASH questionnaire was used.[22] This may also have biased our results, but both questionnaires were quite similar and outcomes were categorized in identical categories.

A third potential limitation of this study concerns some low numbers of participants within neighborhoods, even up to only 1 person. In sensitivity analyses, all studies were conducted (1) excluding neighborhoods that were represented by only one participant, and (2) excluding neighborhoods that were represented by less than 5 participants. Results from these analyses did not differ much from the results that included all study participants (results not shown). Associations for no sports participation and overweight were even stronger than in the present study.

A fourth concern might be the inclusion of respondents above the age of 75 years at follow-up. This subgroup comprised around 10% of the total study population in each study. Excluding this subgroup resulted in patterns that were essentially similar in each study (results not shown). Only for smoking cessation results were not significant anymore, which is probably due to the lower power.

Another concern is that this study examines an association over a period of 13 years. Prevalence of health-related behaviors may change over time and residents may decide to move, which might affect prevalence rates and neighborhood deprivation. However, overall prevalence rates showed an increasing trend over time for smoking cessation and overweight, and a constant trend for no sports participation (results not shown). Thus, it can be expected that results for smoking cessation and becoming overweight might be underestimated, because actual prevalence rates in all neighborhoods may be higher. Also, movements of residents in the *GLOBE* study are limited. An earlier study showed that more than 70% did not move within the city of Eindhoven.[34]

A final concern is the use of neighborhood deprivation to adjust for contextual confounding. To the extent that the prevalence of health-related behaviors and overweight are the causal consequence of neighborhood deprivation, associations

may have been over adjusted. On the other hand, it remains possible that other underlying (contextual) factors may explain the associations.[19,35] This study corrects for neighborhood deprivation as a contextual confounder. However, it cannot be ruled out that perhaps other contextual determinants which are less related to neighborhood deprivation, such as neighborhood accessibility to facilities, may partially explain the relation. Nevertheless, it has been argued that adjusting for neighborhood deprivation, which represents neighborhood social economic status, is a convenient way to adjust for contextual confounding with minimal risk of over adjustments.[19]

In conclusion, neighborhood prevalence of health-related behaviors, and in particular overweight, appears to be associated with changes in health-related behaviors and overweight. If confirmed in other studies, it would imply that a culture of for example overweight may contribute to becoming overweight. This is a novel aspect that adds to the existing explanations of why people engage in healthy or unhealthy behaviors and why this clusters in neighborhoods. This understanding may help to develop and implement interventions aimed at promoting healthy behaviors or preventing unhealthy behaviors more efficiently. However, further longitudinal research is still needed to provide stronger evidence for this association.

Acknowledgments

The *GLOBE* study is carried out by the Department of Public Health of the Erasmus University Medical Centre in Rotterdam, in collaboration with Municipal Public Health Service in the study region (GGD Brabant-Zuidoost). The study has been and is sup- ported by grants of the Ministry of Public Health, Welfare and Sport, the Sick Fund Council, the Netherlands Organization for Advancement of Research, Erasmus University, and the Health Research and Development Council.

References

1. Chaix B. Geographic life environments and coronary heart disease: a literature review, theoretical contributions, methodological updates, and a research agenda. Annu Rev Public Health 2009; 30: 81-105.
2. Cummins S, Macintyre S. Food environments and obesity--neighbourhood or nation? Int J Epidemiol 2006; 35(1): 100-4.
3. MacIntyre S, Ellaway A. Neighborhoods and Health. In: Kawachi I, Berkman LF, eds. Neighborhoods and Health. New York: Oxford University Press; 2003.
4. Roux AVD. Investigating neighborhood and area effects on health. American Journal of Public Health 2001; 91(11): 1783-9.
5. Giskes K, van Lenthe FJ, Turrell G, Brug J, Mackenbach JP. Smokers living in deprived areas are less likely to quit: a longitudinal follow-up. Tob Control 2006; 15(6): 485-8.
6. van Lenthe FJ, Brug J, Mackenbach JP. Neighbourhood inequalities in physical inactivity: the role of neighbourhood attractiveness, proximity to local facilities and safety in the Netherlands. Soc Sci Med 2005; 60(4): 763-75.
7. van Lenthe FJ, Mackenbach JP. Neighbourhood deprivation and overweight: the GLOBE study. Int J Obes Relat Metab Disord 2002; 26(2): 234-40.
8. Wendel-Vos W, Droomers M, Kremers S, Brug J, van Lenthe F. Potential environmental determinants of physical activity in adults: a systematic review. Obes Rev 2007; 8(5): 425-40.
9. Cradock AL, Kawachi I, Colditz GA, Gortmaker SL, Buka SL. Neighborhood social cohesion and youth participation in physical activity in Chicago. Soc Sci Med 2009; 68(3): 427-35.
10. Echeverria S, Diez-Roux AV, Shea S, Borrell LN, Jackson S. Associations of neighborhood problems and neighborhood social cohesion with mental health and health behaviors: the Multi-Ethnic Study of Atherosclerosis. Health Place 2008; 14(4): 853-65.
11. McNeill LH, Kreuter MW, Subramanian SV. Social environment and physical activity: a review of concepts and evidence. Soc Sci Med 2006; 63(4): 1011-22.

12. Ross CE. Walking, exercising, and smoking: does neighborhood matter? Soc Sci Med 2000; 51(2): 265-74.
13. Go MH, Green HD, Jr., Kennedy DP, Pollard M, Tucker JS. Peer influence and selection effects on adolescent smoking. Drug Alcohol Depend 2010; 109(1-3): 239-42.
14. Lakin JL, Jefferis VE, Cheng CM, Chartrand TL. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. Journal of Nonverbal Behavior 2003; 27(3): 145-62.
15. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice-Hall; 1977.
16. Rogers EM. Diffussion of Innovations. 5th ed: New York: The Free Press; 2003.
17. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007; 357(4): 370-9.
18. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008; 358(21): 2249-58.
19. Chaix B, Leal C, Evans D. Neighborhood-level confounding in epidemiologic studies: unavoidable challenges, uncertain solutions. Epidemiology 2010; 21(1): 124-7.
20. van Lenthe FJ, Schrijvers CT, Droomers M, Joung IM, Louwman MJ, Mackenbach JP. Investigating explanations of socio-economic inequalities in health: the Dutch GLOBE study. Eur J Public Health 2004; 14(1): 63-70.
21. van Lenthe FJ, Kamphuis CB, Beenackers MA, et al. Cohort Profile: Understanding socioeconomic inequalities in health and health behaviours: the GLOBE study. Int J Epidemiol 2014; 43(3): 721-30.
22. Wendel-Vos GCW, Schuit AJ, Saris WHM, Kromhout D. Reproducibility and relative validity of the Short Questionnaire to Assess Health-enhancing physical activity. Journal of Clinical Epidemiology 2003; 56(12): 1163-9.
23. World Health Organization. Obesity and overweight fact sheet n°311. 2012. http://www.who.int/mediacentre/factsheets/fs311/en/index.html.
24. van Berkel-van Schaik AB, Tax B. Towards a standard operationalisation of socio-economic status for epidemiological and socio-medical research.: Ministerie van WVC, 1990.
25. Bates D, Maechler M, Bolker B. Lme4: Linear mixed-effects models using

S4 classes, 2012.
26. Team RDC. R: a language and environment for statistical computing. 2010. http://www.R-project.org.
27. Wellman B. Are personal communities local? A Dumptarian reconsideration. Social Networks 1996; 18(4): 347-54.
28. Hawkins J, Hollingworth W, Campbell R. Long-term smoking relapse: a study using the British household panel survey. Nic & Tob Res 2010; 12: 1228-35.
29. Piasecki TM. Relapse to smoking. Clin Psych Rev 2006; 26: 196-215.
30. Blok DJ, van Empelen P, van Lenthe FJ, Richardus JH, de Vlas SJ. Unhealthy behaviour is contagious: an invitation to exploit models for infectious diseases. Epidemiol Infect 2013; 141(3): 667-9.
31. Jago R, Baranowski T, Baranowski JC, Cullen KW, Thompson DI. Social desirability is associated with some physical activity, psychosocial variables and sedentary behavior but not self-reported physical activity among adolescent males. Health Educ Res 2007; 22(3): 438-49.
32. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: sensitivity and specificity using self-reported weight and height. Epidemiology 1990; 1(2): 146-52.
33. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. Am J Public Health 1994; 84(7): 1086-93.
34. van Lenthe FJ, Martikainen P, Mackenbach JP. Neighbourhood inequalities in health and health-related behaviour: results of selective migration? Health Place 2007; 13(1): 123-37.
35. Oakes JM. Commentary: advancing neighbourhood-effects research--selection, inferential support, and structural confounding. Int J Epidemiol 2006; 35(3): 643-7.

4

THE ROLE OF SMOKING IN SOCIAL NETWORKS ON SMOKING CESSATION AND RELAPSE AMONG ADULTS: A LONGITUDINAL STUDY

D.J. BLOK, S.J. DE VLAS, P. VAN EMPELEN, F.J. VAN LENTHE

Published in: Preventive Medicine (2017), Volume 99, Pages 105-110

Abstract

Understanding the spread of smoking cessation and relapse within social networks may offer new approaches to further curb the smoking epidemic. Whether smoking behavior among social network members determines smoking cessation and relapse of adults however, is less known.

For this study, longitudinal data of 4623 adults participating in the Dutch Longitudinal Internet Studies for the Social sciences (*LISS*) panel were collected in March 2013 with a follow-up in 2014. Logistic regression was used to examine the association between the proportion of smokers in social networks, and (1) smoking cessation (n=762) and (2) smoking relapse (n=1905). Analyses were adjusted for the size of the network, age, sex, and education.

Respondents with the largest proportion of smokers in their social network were less likely to quit smoking (OR=0.25; 95% CI=0.11–0.66) and more likely to experience a relapse (6.08; 3.01–12.00). Smoking cessation and relapse were most strongly associated with the proportion of smokers among household members and friends. The proportion of smokers in family outside the household was not related to smoking cessation and smoking relapse.

In conclusion, smoking behavior in social networks, especially among household members and friends, is strongly associated with smoking cessation and relapse. These findings further support the spread of smoking within social networks, and provide evidence for network-based interventions, particularly including household members and friends.

Introduction

Although smoking rates have declined in the past decade, smoking is still one of the leading preventable causes of death and the second highest contributing factor to the overall burden of disease globally.[1-3] In 2013, the smoking prevalence was around 23% in The Netherlands, which is comparable to the global smoking prevalence and that in the EU.[2,3] In The Netherlands, smoking causes approximately 19,000 deaths annually, and at 13% it is the highest contributing risk factor to the overall burden of disease.[2] Thus, there remains a need to design new effective interventions, also because of the growing interest in so-called endgame strategies.[4] The tobacco endgame suggests we should move beyond tobacco control toward an entirely tobacco free societies.

In 2008, Christakis and Fowler used unique data from the Framingham Study to show the spread of smoking via social network members, including spouses, siblings, friends, and neighbors.[5] Using logistic regression models, they assessed the relationship between contacts who quit smoking and smoking cessation of the subject in a dynamic social network over a 32-year period. Their findings suggested that spouses and friends, who quit smoking, were particularly relevant for smoking cessation. This may have important implications for the design of network-based interventions. Not only does it provide new targets for interventions, it also implies that interventions might be more effective since positive health behaviors might spread to others as well.[6]

Further support for contagiousness of smoking behavior within social networks however, remains scarce. This is probably due to the inherent complexity to collect longitudinal information about smoking behavior in dynamic social networks. Studies investigating smoking behaviors in networks often focused on smoking initiation among adolescents, demonstrating that smoking behavior of social contacts is strongly associated with, and perhaps even the cause of, smoking initiation.[7-10] The influence of smoking behaviors of social network members on smoking cessation and smoking relapse in adult populations, however, has received far less attention. Studies primarily related the number or proportion of smokers among either spouses, households or neighborhoods to smoking cessation or relapse.[11-14] For example, smoking cessation is shown to be

more likely in neighborhoods with a high prevalence of non-smokers.[11] Similarly, smoking relapse is more likely in households with a high number of smokers.[14] Empirical evidence on the importance of the nature of social ties (e.g. household member, friend, family) is scarce as well.

This study used the Longitudinal Internet Studies for the Social sciences (*LISS*) panel from The Netherlands, which is a large prospective study. Apart from measures about the respondents' smoking behavior, it also includes information about smoking behavior among social network members. This provides a unique opportunity to examine: (1) whether smokers with more smoking social network members are less likely to quit smoking during follow-up, and (2) whether former smokers with more smoking social network members are more likely to relapse. Unlike Christakis and Fowler who related quitting among social network members to smoking cessation, we hypothesized that having more smoking social network members makes it more difficult to quit smoking and also easier to relapse because of the continuous visual cues.[5,15] In addition, we assessed the dependence of these associations on specific types of social network ties, e.g. household members, close friends, and family members outside the household including parents and siblings.

Methods

Data

We used data from the Dutch *LISS* panel administered by *CentERdata* (Tilburg University, The Netherlands). It is a large Internet survey with almost 7000 individuals above the age of 15 years. The panel is based on a probability sample of households drawn from the population register and is therefore representative of the Dutch population.[16] The panel is operational since 2007. Every year, a longitudinal survey is set out, which covers a large variety of domains including education, income, work, housing, values, personality, time use and political views. More information about the *LISS* panel can be found in *Scherpenzeel et al. (2010)*[17] and at: www.lissdata.nl.

For this study, we added several questions on smoking behaviors of respondents and of members in their social network to the routine data collection by *CentERdata*.

IRB approval was therefore not necessary for this study. All participants gave consent to participate in the *LISS* panel. These prospective data were collected in March 2013 with a follow-up in April 2014. All data were anonymous and made publically available through the *LISS* panel's website.[18] Initially, 6562 respondents of the *LISS* panel were invited to fill out a questionnaire of which 5538 responded (84.4%). Of those, 5221 were approached for follow-up after one year, of which 4625 responded (88.6%). Only respondents who completed the questionnaire at both waves were included in our analysis. Subsequently, two subpopulations were identified: (1) smokers at baseline (n=762), and (2) former smokers at baseline (n=1905) (see Figure 1).

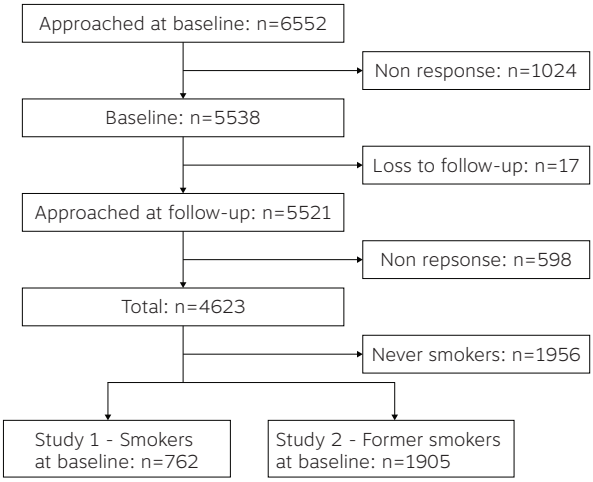


Figure 1. CONSORT flow diagram.

Measures

Self-reported smoking status was assessed by asking respondents whether they currently smoked (yes/no), and whether they had ever smoked before (yes/no). Smoking cessation was defined as smoking at baseline but not during follow-up. Similarly, smoking relapse was defined as smoking during follow-up by former smokers at baseline. Former smokers were those who did not smoke at baseline, but reported to have smoked in the past.

Among the respondents who smoked at baseline, additional questions were asked

regarding smoking frequency, smoking quantity and quit attempts. Every current smoker was asked whether he/she is a daily smoker or less than daily smoker, how many cigarettes per day (CPD) he/she smoked, and whether he/she had ever made a quit attempt (yes/no).

Respondents were asked to provide information about the composition of their social network and the smoking behaviors of all nominated social network ties. Specifically, each respondent was asked to name all household members, up to five closest friends, and all siblings and parents outside the household that were still alive at the time of the study. Close friends were defined as people with whom the respondent had close contact in the past six months. Siblings and parents outside the household together made up the group "family members outside the household". A respondent's total network size comprised all reported household members, friends, and family members outside the household.

Smoking behavior of social network members was assessed by asking the respondent to indicate for each social network tie whether he/she currently smoked, both at baseline and during follow-up. This information was used to calculate, the proportion of social network members that smoked at baseline and during follow-up (henceforth: proportion of smokers). This proportion was calculated for all social ties combined, as well as each type of social tie (i.e. household members, friends, and family members) separately.

Network size, age, sex, and educational level of the respondent were used as control variables. Adjustment for the size of the network was needed because the proportion of smokers is dependent on the size: a single smoker in a small network results in a higher proportion of smokers than in a large network. Also, the size itself might influence smoking cessation or relapse. It is known that smoking provides a way to cope with psychological mechanisms resulting from being socially isolated.[19] Educational level was defined as the respondent's highest attained level of education, selected from eight levels ranging from primary to academic education. For the analyses, these were further categorized as follows: lower (primary and lower secondary), middle (higher secondary), and higher (tertiary) education. Educational level has proven to be a good indicator of socio-economic status in The Netherlands.[20]

Analysis

Analyses were carried out separately for two subpopulations: smokers at baseline and former smokers at baseline. First the data was presented graphically as the proportion of (1) smoking cessation among the 762 smokers and (2) relapse among the 1905 former smokers for those with less than and those with >50% smokers in their social network.

Then, we specified logistic regression models where the respondent's smoking status was a function of the proportion of smokers in the social network. The model was adjusted for network size, age, sex and educational level. In separate logistic regressions, we examined the adjusted association of the proportion of smokers among social network members with (1) smoking cessation among smokers, and (2) smoking relapse among former smokers. In order to investigate the importance of the type of social tie, these analyses were performed separately for the proportion of smokers among (A) all social ties, (B) household members, (C) friends and (D) family members outside the household.

Since the relationship between smoking cessation and the proportion of smokers among social network members might differ with smoking frequency and smoking quantity of the respondent, we also repeated the analysis stratified by daily smokers, occasional (1–5 CPD), light (6–10 CPD), moderate (11–15 CPD), moderate-heavy (16–20 CPD), and heavy (N21 CPD) smokers. All analyses were conducted with the statistical package R (version 3.2.2).[21]

Results

Descriptive characteristics of the two subpopulations are provided in Table 1. At baseline, the average age was approximately 50 years among smokers and 58 years among former smokers. In both sub-populations, each respondent had on average five social ties. Among smokers, approximately 14% reported not to smoke at follow-up. Of all social ties in this subpopulation, 30% were smokers. Smoking was more prevalent among friends. Among former smokers, about 6% reported to smoke at follow-up. In this subpopulation, the proportion of smokers among all social ties was on average 15%. Again, the highest proportion was among friends.

Table 1. Characteristics of the study populations of smokers and former smokers.

Variables	Smokers at baseline (n=762)	Former smokers at baseline (n=1905)
Smoking cessation, *n (%)*	103 (13.5)	-
Smoking relapse, *n (%)*	-	121 (6.4)
Proportion of smokers among,[a] *mean (SD)*		
All social ties	0.30 (0.28)	0.15 (0.22)
Household members	0.25 (0.39)	0.09 (0.25)
Friends	0.26 (0.35)	0.13 (0.26)
Family members outside the household	0.13 (0.26)	0.10 (0.24)
Network size, *mean (SD)*		
All social ties	5.00 (3.00)	5.00 (3.15)
Household members	1.20 (1.22)	1.27 (1.14)
Friends	2.01 (1.83)	1.88 (1.81)
Family members outside the household	1.80 (1.91)	1.86 (1.92)
Age (years), *mean (SD)*	49.6 (15.40)	57.7 (14.70)
Female, *n (%)*	367 (48.2)	933 (49.0)
Education, *n (%)*		
Low	285 (37.4)	642 (33.7)
Middle	309 (40.6)	672 (35.3)
High	168 (22.0)	591 (31.0)

Note: Study conducted in the Netherlands (2013-2014)

[a] *Proportion of social network members who smoked at baseline (t=0) and during follow-up (t=1).*

Figure 2 illustrates the unadjusted estimates of the percentage of persons who stopped smoking and who relapsed during follow-up by the proportion of smokers among social network members. The percentage of respondents who quitted smoking was substantially lower with a high proportion of smokers among social network members. This relation was found for each type of social tie, except for the proportion of smokers among family members outside the household. Smoking relapse among former smokers was more likely when the proportion of smokers among network members was high.

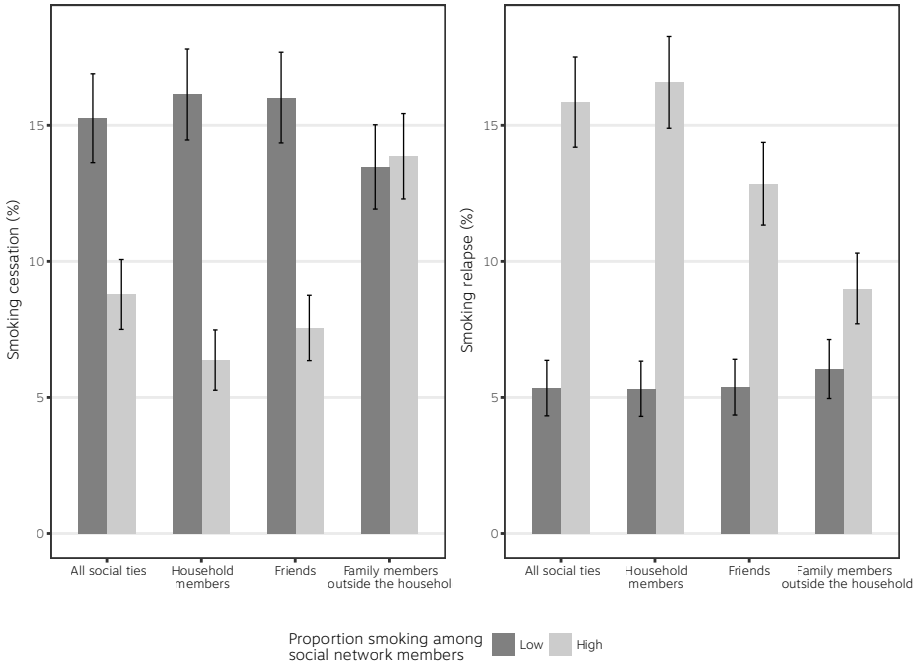


Figure 2. Unadjusted estimates of the percentage of persons who stopped smoking and who relapsed during follow-up by the proportion of smokers among social network members. The proportion of smokers among social networks members was dichotomized into: low (≤50%) and high (>50%). Study conducted in The Netherlands (2013–2014).

Table 2 provides the adjusted odds ratios for the association between the proportions of smokers in social networks and smoking cessation of participants during follow up. Smoking cessation was less likely when the respondent had higher proportions of smokers among all social ties (OR=0.25; 95% CI=0.10–0.66), household members (OR=0.27; 95% CI=0.12–0.56), and friends (OR=0.26; 95% CI=0.11–0.59). The proportion of smokers among family members outside the household did not show a relationship with smoking cessation.

Table 3 shows that among former smokers, higher proportions of smoking among all social ties increased the odds of smoking relapse (OR=6.08; 95% CI=3.01–12.00). Of all social ties, the proportion smokers among household members (OR=4.33; 95% CI=2.54–7.18) and friends (OR=2.68; 95% CI=1.44–4.83) showed a significant association with smoking relapse as well. Our results did not show a relation between smoking relapse and the proportion of smokers among family members outside the household.

Table 4 presents the stratified analysis for smoking cessation. When we stratified for daily smokers only, results did not differ from the complete analysis (see Table 2). Among occasional and light smokers, smoking cessation was less likely with high proportions of smokers among all social ties. Also, among moderate to heavy smokers, a high proportion of smokers among all social ties, and in particular among friends, made it less likely to quit smoking. Smoking cessation was not related to smoking behavior among moderate smokers.

Table 2. Adjusted associations between smoking cessation and the proportion among social network members

Variables	Smoking cessation (n=762)							
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Proportion of smokers among [a]								
All social ties	0.25	0.10-0.66	-	-	-	-	-	-
Household members	-	-	0.27	0.12-0.56	-	-	-	-
Friends	-	-	-	-	0.26	0.11-0.59	-	-
Family members outside the household	-	-	-	-	-	-	1.02	0.43-2.24
Network size	0.94	0.87-1.01	0.97	0.80-1.17	0.97	0.85-1.10	0.97	0.86-1.09
Age (10yrs)	0.94	0.82-1.08	0.96	0.83-1.10	0.94	0.82-1.08	0.98	0.85-1.12
Sex – Female	1.29	0.84-1.98	1.23	0.81-1.88	1.32	0.86-2.03	1.2	0.79-1.84
Education – Low (ref)	-	-	-	-	-	-	-	-
Education – Middle	1.49	0.90-2.49	1.45	0.87-2.42	1.54	0.93-2.58	1.46	0.89-2.44
Education – High	1.93	1.10-3.41	1.71	0.97-3.00	2.06	1.16-3.65	1.97	1.13-3.44

Note: Study conducted in the Netherlands (2013-2014)

[a] Proportion of social network members who smoked at baseline (t=0) and during follow-up (t=1).

Table 3. Adjusted associations between smoking relapse and the proportion smokers among social network members

Variables	Smoking relapse (n=1905)							
	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI
Proportion of smokers among[a]								
All social ties	6.08	3.01-12.00	-	-	-	-	-	-
Household members	-	-	4.33	2.54-7.18	-	-	-	-
Friends	-	-	-	-	2.68	1.44-4.83	-	-
Family members outside the household	-	-	-	-	-	-	1.04	0.48-2.06
Network size	0.99	0.93-1.06	0.89	0.74-1.05	1.02	0.91-1.14	0.95	0.84-1.05
Age (10yrs)	0.64	0.56-0.71	0.62	0.55-0.70	0.64	0.57-0.72	0.63	0.56-0.70
Sex – Female	0.80	0.54-1.18	0.77	0.52-1.14	0.80	0.54-1.18	0.81	0.55-1.18
Education – Low	-	-	-	-	-	-	-	-
Education – Middle	0.81	0.50-1.31	0.76	0.47-1.21	0.77	0.48-1.23	0.74	0.46-1.18
Education – High	1.17	0.73-1.88	1.05	0.66-1.69	1.07	0.67-1.72	1.01	0.64-1.61

Note: Study conducted in the Netherlands (2013-2014)

[a] *Proportion of social network members who smoked at baseline (t=0) and during follow-up (t=1).*

Table 4. Adjusted associations between smoking cessation and the proportion of smokers among social network members stratified by smoking frequency and quantity.

Variable	Smoking cessation											
	Frequency		Quantity[a]									
	Daily smokers (n=488)		Occasional (n=138)		Light (n=220)		Moderate (n=161)		Moderate-heavy (n=121)		Heavy (n=88)	
	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI
Proportion of smokers among[b]												
All social ties	0.26	0.10-0.66	0.48	0.08-2.33	0.02	0.001-0.25	1.59	0.25-9.02	0.02	0.0003-0.41	0.38	0.02-4.07
Household members	0.31	0.13-0.67	0.21	0.03-0.98	0.09	0.01-0.51	0.78	0.18-2.79	0.30	0.02-1.72	0.74	0.03-8.89
Friends	0.29	0.10-0.72	0.61	0.14-2.24	0.07	0.01-0.53	1.00	0.14-5.34	0.02	0.002-0.50	0.00	-Inf - Inf
Family outside the household	1.16	0.44-2.73	1.22	0.17-6.66	0.13	0.003-1.59	1.96	0.35-8.72	1.26	0.08-11.32	3.00	0.02-123

Note: All analyses are adjusted for network size, age, sex, and education. Study conducted in the Netherlands (2013-2014)[a] Occasional(1-5 cigarettes per day); Light (6-10 CPD); Moderate (11-15 CPD); Moderate-heavy (16-20 CPD); Heavy (>21 CPD).

[b] Proportion of social network members who smoked at baseline (t=0) and during follow-up (t=1).

Discussion

This study assessed the role of smoking in social networks on smoking cessation and smoking relapse in a Dutch adult population. Clearly, those with the largest proportion of smokers among their social network members were less likely to quit and more likely to experience a relapse. In addition, the type of social tie appeared to be important: smoking of household members and friends was strongly associated with smoking cessation and relapse, whereas the smoking behavior of family members outside the household was not.

Our findings support the significance of smoking behaviors in social networks for adult smoking cessation and relapse.[5,14,22] Smoking behaviors of others in social networks may have undesirable influences through peer pressure and modeling smoking behaviors.[9,23] It also supports the idea that smoking behaviors might be contagious.[24] Given the addictiveness of smoking, consistent visual cues of smoking by social network members may make it harder for a smoker to quit and easier for a former smoker to relapse smoking.[15]

With regard to the importance of social ties, our study is among the first to empirically demonstrate that smoking behavior of household members and friends are both important for smoking cessation and relapse in adults. Our results further indicate that smoking among friends is most important for moderate to heavy smokers, whereas smoking among household members is most important for occasional and light smokers. These findings may suggest the need for network-based interventions to target different people for occasional, light, and moderate to heavy smokers.

Surprisingly, smoking behavior of family members outside the household appeared to have no significant relation to smoking cessation or relapse, after adjusting for other factors. This seems contradictory to earlier findings by Christakis and Fowler, which suggest that socially close contacts are most important to explain smoking cessation.[5] However, as our estimates for the family member outside the household show very wide confidence intervals, these are consistent with both positive and negative effects on smoking cessation and relapse. It may reflect that some respondents have more frequent contact (i.e. closer contact) with their

family members than others. Since our data did not include a proxy for closeness of contacts, we were not able to test this hypothesis.

A limitation of this study was the use of self-reported smoking status that was measured fairly crude due to limited space in the survey. This might have resulted in an underestimation of smoking among respondents.[25] However, self-reported smoking status has shown to be reliable when measured under optimized conditions, such as assuring anonymity.[26] Also, smoking behaviors of social ties were measured based on self-reported perceptions of the respondent, which may be biased.[27] Former studies have shown that smokers tend to overestimate friends' smoking behavior, which then could result in an overestimation of the importance of smoking in social networks.[27,28] However, a recent study comparing egocentric and sociometric measures (i.e. self-reported by the social tie) showed that the level of agreement was high.[10]

Another limitation might be the short follow-up period. Generally, smoking cessation is a slow process and smokers may make multiple attempts before successfully quitting smoking.[29] It is very likely that smokers at baseline who reported not to smoke during follow-up, might still be in the process of quitting smoking, and that our observations of smoking cessation is merely just another quit attempt. Similarly, smoking relapse among former smokers at baseline might actually be a failed quit attempt. However, further adjusting our analysis for quit attempt (yes/no) did not alter our results.

A main concern when studying the role of social networks on health related behaviors or health in general, is homophily.[30] People tend to select others with similar behaviors as their friends, which may explain to some extent the change in behaviors. This phenomenon can also be observed in our data. At baseline, the proportion of smokers in social networks was twice as high among smokers as compared to former smokers. In our study however, selection seems to be less of a problem because family members are not selected, and household members are only to some extent selected.

We also did not account for any contextual effects.[31,32] People may be exposed to common environmental factors, such as availability of smoking, which may

to some extent cause them to quit or restart smoking. Our data did not contain any geographic or physical environmental measures, except for the degree of urbanization. Further adjusting our models for urbanization did not influence our results.

Conclusions

In conclusion, this study supports the finding that smoking cessation and smoking relapse are influenced by smoking behaviors of members in the social network. This association has now been demonstrated for a specific network consisting of household members, friends, and family members. Smoking behaviors of household members and friends are the most important to explain smoking cessation and relapse. Based on these findings, network-based interventions targeting household members or groups of friends may be most promising to establish a significant further reduction of the smoking prevalence. Collectively reducing the proportion of smokers in social networks might not only enforce others to quit smoking too, but might also make smoking relapse less likely.

Acknowledgement

The LISS panel data were collected by *CentERdata* (Tilburg University, The Netherlands) through its *MESS* project funded by The Netherlands Organization for Scientific Research. This work was supported by a grant from the Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.

References

1. Murray CJ, Lopez AD. Measuring the global burden of disease. N Engl J Med 2013; 369(5): 448-57.
2. RIVM. Volksgezondheid Toekomst Verkenning: Een gezonder Nederland. Bilthoven: RIVM, 2014.
3. World Health Organization. WHO report on the global tobacco epidemic, 2015.
4. McDaniel PA, Smith EA, Malone RE. The tobacco endgame: a qualitative review and synthesis. Tob Control 2016; 25(5): 594-604.
5. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008; 358(21): 2249-58.
6. Valente TW. Network interventions. Science 2012; 337(6090): 49-53.
7. Ennett ST, Faris R, Hipp J, et al. Peer smoking, other peer attributes, and adolescent cigarette smoking: a social network analysis. Prev Sci 2008; 9(2): 88-98.
8. Go MH, Green HD, Kennedy DP, Pollard M, Tucker JS. Peer influence and selection effects on adolescent smoking. Drug Alcohol Depen 2010; 109(1-3): 239-42.
9. Mercken L, Snijders TA, Steglich C, de Vries H. Dynamics of adolescent friendship networks and smoking behavior: social network analyses in six European countries. Soc Sci Med 2009; 69(10): 1506-14.
10. Valente TW, Fujimoto K, Soto D, Ritt-Olson A, Unger JB. A Comparison of Peer Influence Measures as Predictors of Smoking Among Predominately Hispanic/Latino High School Adolescents. J Adolescent Health 2013; 52(3): 358-64.
11. Blok DJ, de Vlas SJ, van Empelen P, Richardus JH, van Lenthe FJ. Changes in smoking, sports participation and overweight: does neighborhood prevalence matter? Health Place 2013; 23: 33-8.
12. Holahan CJ, North RJ, Holahan CK, Hayes RB, Powers DA, Ockene JK. Social influences on smoking in middle-aged and older women. Psychol Addict Behav 2012; 26(3): 519-26.
13. Hymowitz N, Sexton M, Ockene J, Grandits G. Baseline factors associated with smoking cessation and relapse. MRFIT Research Group. Prev Med

1991; 20(5): 590-601.
14. Moore S, Teixeira A, Stewart S. Effect of network social capital on the chances of smoking relapse: a two-year follow-up study of urban-dwelling adults. Am J Public Health 2014; 104(12): e72-6.
15. Caggiula AR, Donny EC, White AR, et al. Cue dependency of nicotine self-administration and smoking. Pharmacol Biochem Behav 2001; 70(4): 515-30.
16. de Vos K. Representativeness of the LISS panel 2008, 2009, 2010. Tilburg: CentERdata, 2010.
17. Scherpenzeel AC, Das M. "True" Longitudinal and Probability-Based Internet Panels: Evidence From the Netherlands. In: Das M, Ester P, Kaczmirek L, eds. Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies. Boca Raton: Taylor & Francis; 2010: 77-104.
18. CentERdata. LISS panel. 2016. http://www.lissdata.nl/dataarchive/ (accessed May 2015).
19. Choi HJ, Smith RA. Members, isolates, and liaisons: meta-analysis of adolescents' network positions and their smoking behavior. Subst Use Misuse 2013; 48(8): 612-22.
20. van Berkel-van Schaik AB, Tax B. Towards a standard operationalisation of socio-economic status for epidemiological and socio-medical research.: Ministerie van WVC, 1990.
21. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2016.
22. Zhou X, Nonnemaker J, Sherrill B, Gilsenan AW, Coste F, West R. Attempts to quit smoking and relapse: factors associated with success or failure from the ATTEMPT cohort study. Addict Behav 2009; 34(4): 365-73.
23. Lakin JL, Chartrand TL. Using nonconscious behavioral mimicry to create affiliation and rapport. Psychol Sci 2003; 14(4): 334-9.
24. Blok DJ, van Empelen P, van Lenthe FJ, Richardus JH, de Vlas SJ. Unhealthy behaviour is contagious: an invitation to exploit models for infectious diseases. Epidemiol Infect 2013; 141(3): 667-9.
25. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the

relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob Res 2009; 11(1): 12-24.

26. Dolcini MM, Adler NE, Ginsberg D. Factors influencing agreement between self-reports and biological measures of smoking among adolescents. J Res Adolescence 1996; 6(4): 515-42.
27. Henry DB, Kobus K, Schoeny ME. Accuracy and bias in adolescents' perceptions of friends' substance use. Psychol Addict Behav 2011; 25(1): 80-9.
28. Iannotti RJ, Bush PJ. Perceived vs. actual friends' use of alcohol, cigarettes, marijuana, and cocaine: Which has the most influence? J Youth Adolesc 1992; 21(3): 375-89.
29. Hymowitz N, Cummings KM, Hyland A, Lynn WR, Pechacek TF, Hartwell TD. Predictors of smoking cessation in a cohort of adult smokers followed for five years. Tob Control 1997; 6 Suppl 2: S57-62.
30. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Soc 2001; 27: 415-44.
31. Cohen-Cole E, Fletcher JM. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. J Health Econ 2008; 27(5): 1382-7.
32. Manski CF. Identification of Endogenous Social Effects - the Reflection Problem. Rev Econ Stud 1993; 60(3): 531-42.

5

REDUCING INCOME INEQUALITIES IN FOOD CONSUMPTION: EXPLORATIONS WITH AN AGENT-BASED MODEL

D.J. BLOK, S.J. DE VLAS, R. BAKKER, F.J. VAN LENTHE

Published in: American Journal of Preventive Medicine (2015), Volume 49, Pages 605-613

Abstract

Introduction

Individual and environmental factors dynamically interact in shaping income inequalities in healthy food consumption. The agent-based model, Health Behaviors Simulation (*HEBSIM*), was developed to describe income inequalities in healthy food consumption. It simulates interactions between households and their environment. *HEBSIM* was used to explore the impact of interventions aimed at reducing food consumption inequalities.

Methods

HEBSIM includes households and food outlets. Households are characterized by location, composition, income, and preference for food. Decisions about where to shop for food (fruit/vegetable stores, supermarkets, or discount supermarkets) and whether to visit fast food outlets are based on distance, price, and food preference. Food outlets can close and new food outlets can enter the system. Three interventions to reduce healthy food consumption inequalities were tested: (1) eliminating residential segregation; (2) lowering the prices of healthy food; and (3) providing health education. *HEBSIM* was quantified using data from Statistics Netherlands, Statistics Eindhoven, and the *GLOBE* study (2011).

Results

The model mimicked food consumption in Eindhoven. High-income households visited healthy food shops more often than low-income households. Eliminating residential segregation had the largest impact in reducing income inequalities in food consumption, but resulted partly from a worsening in healthy food consumption in high-income households. Lowering prices and health education could also substantially reduce inequalities. Most interventions took 5–10 years to reach their (almost) full effects.

Conclusions

HEBSIM is a promising tool for studying dynamic interactions between households and their environment and for assessing the impact of interventions on income inequalities in food consumption.

Introduction

Dietary behavior is a major contributor to socioeconomic inequalities in health.[1,2] Lower socioeconomic groups are less likely to consume fruit or vegetables and have higher fat-intake than higher socioeconomic groups.[3-5] Initial research attributed socioeconomic inequalities in food consumption to differences in food-related knowledge, beliefs, and attitudes between socioeconomic groups.[6] However, there is growing evidence of the importance of characteristics of the obesogenic environment, such as access to healthy foods and fast food outlets, and the price of healthy food.[7-10]

Decisions on food consumption result from interactions between individuals or households and their environment.[11-14] An environment with healthy food outlets not only influences the choice for healthy food but is also shaped by the preferences of residents because shops prefer to be close to their customers.[15,16] As such, differential access to healthy food outlets may be the outcome of feedback. In addition, the outcome may depend on differences between individuals or households in, for example, income or location, causing some people to be more influenced by their environments than others.[12,17] Thus, people and their environment form an interactive "system".[14,18,19]

Research aimed at identifying the best interventions to reduce income inequalities in healthy food consumption is subject to challenges. Evaluating or predicting the impact of interventions is difficult because it takes (a long) time before an intervention may show an effect and, in some cases, evaluation is not even feasible.[13] Moreover, current statistical approaches simplify the complex interactions by neglecting or fixing features of the system, such as changes in the food environment due to interventions.[13,19] Failing to properly account for such interactions may result in underestimated or overestimated effects of interventions.[13]

A "systems-thinking" method that is able to overcome these limitations is the use of agent-based models (ABMs).[13,20] ABMs are simulation models describing a system of heterogeneous agents that influence each other over time.[21,22] An agent can be any entity, such as a household or food outlet, and is characterized by

attributes (e.g., income) and follows behavioral rules, which together determine its behaviors. Behavioral rules, which are usually static, describe how an agent interacts with other agents and their environment. Agents can adapt their behaviors in response to changes in behaviors of other agents and to changes in their environment due to interventions, for example.[13,21,22] ABMs are therefore very suitable for exploring the impact of various interventions.[13,19,20]

ABMs are increasingly recognized in the fields of social epidemiology and health behavioral research.[23-26] *Auchincloss et al.*[23] were the first to use an ABM to understand income inequalities in diet. They modeled households' food shopping behavior in a hypothetical city and illustrated the importance of residential segregation, price, and food preference in shaping income inequalities in diet.

For this study, the agent-based model, *He*alth *B*ehaviors *Sim*ulation (*HEBSIM*), was developed by Erasmus MC. *HEBSIM* describes a real-world system explaining income inequalities in healthy food consumption in Eindhoven, The Netherlands. The model was quantified using empirical data and used to explore the impact of three (hypothetical) interventions aimed at reducing income inequalities in healthy food consumption: (1) eliminating residential segregation; (2) lowering prices of healthy food; and (3) health education.

Methods

Model

HEBSIM simulates the dynamic interaction between two types of agents: households and food outlets. It models households' food shopping behavior and the closures and openings of food outlets in a city. Time is represented in days. A simulation takes approximately 20 minutes to run. *HEBSIM* was programmed in Java using MASON and Magic-Tree, which are open-source Java libraries/tools for developing ABMs.[27,28] *HEBSIM* is available to researchers upon request.

In this study, the city of Eindhoven, The Netherlands was modeled. Initial conditions of the model were assigned using data about the population, obtained from Statistics Netherlands[29] and Eindhoven[30] and the 2011 wave of the Dutch prospective *GLOBE* study.[31] The *GLOBE* study includes a random sample of

respondents between the ages of 15 and 75 years living in the city of Eindhoven. The 2011 wave consisted of 3,863 participants, of whom 2,398 had valid measurements on food behaviors.[31]

Environment

The city of Eindhoven is approximately 88 km2 and includes 116 neighborhoods, 88 of which are residential neighborhoods (see Figure 1A). A realistic grid was constructed using a GIS file of Eindhoven as input.[29] Each grid cell is 10 m X 10 m in size and can be occupied by a household and a food or vacant outlet. Unoccupied grid cells are considered "other areas" (e.g., parks). Household and vacant outlets were randomly distributed on the grid based on the number of households and vacant (food and non-food) outlets per neighborhood.[30] Food outlets were located on actual locations based on a search in the Yellow Pages in July 2012. In 2012, Eindhoven had 97,523 households, 12 fruit/vegetable stores, 33 supermarkets, 22 discount supermarkets, 160 fast food outlets, and 279 vacant outlets (see Figure 1B and Figure 1C).[30]

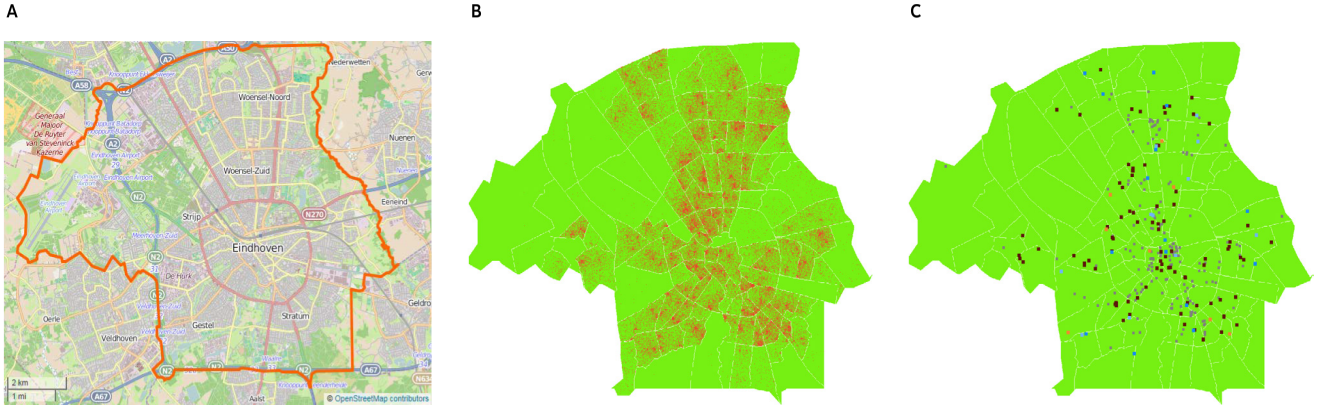


Figure 1. (A) Map of Eindhoven: the fifth largest city the Netherlands with 213,223 inhabitants and 97,523 households in 2012 (Source: OpenStreetMap.org); (B) Distribution of households on the grid at the start of the simulation: 97,523 households (red squares) in 88 neighborhoods; (C) Distribution of food outlets and vacant outlet on the grid at the start of the simulation: 12 fruit/vegetable stores (orange), 33 supermarkets (dark blue), 22 discount supermarket (light blue), 160 fast food outlets (brown) and 279 vacant outlets (grey).

Attributes

Besides a location, each household is characterized by three other attributes: (1) household composition; (2) income level; and (3) preference for healthy food. Household composition was assigned following an empirical distribution per neighborhood and included single-person, single-parent, and multi-person with and without children.[30] Income was dichotomized into high and low. Low-income households were defined as those with a net household income <$31,777 per year and was assigned following the distribution of low-income households per neighborhood.[29] The preference for healthy food represents the attitude toward healthy food, ranging from 0 (prefers unhealthy food) to 1 (prefers healthy food). It was randomly assigned following a beta distribution. Based on results of *Turrell and colleagues*[32], the mean preference for healthy food was 0.739 and 0.661 for high- and low-income households, respectively.

HEBSIM distinguishes four types of food outlets: fruit/vegetables stores, supermarkets, discount supermarkets, and fast food outlets. Besides a location, each type of food outlet has four other attributes: (1) quality of food; (2) price level; (3) monthly costs; and (4) capital. The quality of food refers to the level of healthy food sold in a food outlet, which is either mostly "healthy" or mostly "unhealthy." Fruit/vegetable stores and supermarkets were characterized as healthy, and discount supermarkets and fast food outlets as unhealthy. Studies have shown that people shopping at supermarkets eat and buy healthier food than those shopping at discount supermarkets.[33-35] Price levels were dichotomized into "cheap" and "expensive," because data only allowed quantification of households' behaviors for these categories. Fruit/vegetable stores and supermarkets were considered expensive, and discount supermarkets and fast food outlets were considered cheap.[33,34] Monthly costs, such as rent, remained unchanged during the simulation. Capital represents a food outlet's asset, which can increase or decrease during the simulation (see *"Agents and attributes"* in the Supplementary Appendix).

Behaviors

Behavioral rules of households were based on the literature and quantified using data. Two behaviors of households were modeled: (1) food shopping and (2) fast food visits. The multi-attribute utility theory was used to guide interactions between households and food outlets and to determine behaviors.[36,37] It was assumed that each household does food shopping three times a week.[34] At each food shopping moment, each household selects a food outlet (i.e., fruit/vegetable store, supermarket, discount supermarket) by assigning a utility to each food outlet using a utility function. The food outlet with the highest utility is selected. The utility function includes determinants of food behavior and a random noise to account for bounded rationality.[38] Determinants for food behavior were based on literature and include distance, price of a food outlet, and preference for healthy food.[39] Each household assigns a score to each determinant based on its own attributes and that of the food outlet. Scores for price were derived from the *GLOBE* study,[31] assuming these scores differed between high- and low-income households. Scores for preference were determined by matching the quality of a food outlet with the preference for healthy food, which already incorporated income differences.

Fast food visits were considered daily. Here, each household assigns a utility to visiting fast food and to eating at home. Again, a utility function was used, but now including two determinants only: preference and price level.[39] Additionally, it was assumed that the availability of fast food in the neighborhood increases the preference for fast food.[40-43] Scores of determinants were assigned similarly to food shopping. A household only visits a fast food outlet when the utility of fast food is higher than the utility of eating at home. Finally, a fast food outlet was chosen based on distance only because fast foods only differ from each other in location (see *"Behavioral rules of households"* in the Supplementary Appendix).

Food outlets respond to household decisions by either closing or opening a new food outlet. Closure is determined by a food outlet's capital, which increases with every customer (revenue) and decreases with monthly costs. If capital falls below zero, the food outlet closes and its location will become vacant. Every 30 days, new food outlets can be started at vacant locations. The number and type of new

food outlets is determined by the monthly average number of new food outlets obtained from the Chamber of Commerce.[44] A location is randomly selected in the neighborhood with the highest expected revenue (see *"Behavioral rules of food outlets"* in the Supplementary Appendix).

Calibration

The model was fitted to data and ran until model outcomes reached a steady state (15,000 days). Model outcomes were based on the average results of 100 iterations. Parameters for which no data were available were calibrated against data about food shopping, fast food visits, and food outlets obtained from the *GLOBE* study and literature.[31,45] These parameters included relative importance of distance, price and preference of healthy food, the importance of bounded rationality, and monthly costs. Calibration was performed through an iterative process until the proportion of households that visited food outlets in the past week and the number of food outlets matched the data (see Table 1 and *"Calibration"* in the Supplementary Appendix).

Table 1. Comparison of model outcomes with empirical data obtained from the *GLOBE* study

	Model outcomes[a]	(90% C.I.)[b]	Data
Proportion of households visited a fruit/vegetable store at least once in the past week:			
Fruit/vegetable store			
High-income	0.427	(0.354-0.488)	0.250
Low-income	0.369	(0.300-0.419)	0.331
Supermarket			
High-income	0.886	(0.857-0.914)	0.883
Low-income	0.813	(0.781-0.845)	0.842
Discount supermarket			
High-income	0.393	(0.344-0.431)	0.412
Low-income	0.554	(0.510-0.600)	0.584
Proportion of households that visited a fast food outlet in the past week:			
Never	0.252	(0.249-0.254)	0.249[c]
1-2 times	0.514	(0.512-0.517)	0.522[c]
>2 times	0.234	(0.230-0.236)	0.229[c]
Proportion of households that visited one type of food outlet	0.364	(0.329-0.405)	0.356
Average distance travelled for food shopping (meters)			
High-income	2052	(1951-2157)	-
Low-income	1775	(1695-1870)	-
Number of food outlets:			
Fruit/vegetable store	12	(9-14)	12
Supermarket	34	(30-37)	33
Discount supermarket	23	(20-26)	22
Fast food	161	(158-163)	160

[a] *Model outcomes in steady state (after 15,000 days)*

[b] *90% CI: 90% of the runs fall within this interval*

[c] *Obtained from French et al. (2001)*[45]

Outcome

Healthy food consumption was defined as the average proportion of times a household visited a healthy food outlet. Visits to fruit/vegetable stores and supermarkets counted as healthy food consumption, whereas visits to discount supermarkets and fast food outlets counted as unhealthy food consumption. Income inequalities in healthy food consumption were defined as the difference

between healthy food consumption of high- and low-income households.

Interventions

The impact of three (hypothetical) interventions, targeting potential upstream and downstream causes of income inequalities in food consumption, was explored. The first intervention assumed income inequalities in food consumption to be caused by differential access to healthy foods. The impact of eliminating residential segregation in neighborhoods on income inequalities in healthy food consumption was assessed. Households were relocated such that each neighborhood had the same proportion of high- and low-income households.

The second intervention assumed higher prices of healthy food to be a main contributor of income inequalities in food consumption. Therefore, the impact of reducing the prices of healthy food by subsidizing all healthy food outlets was assessed. Expensive healthy food outlets were changed to "cheap" and the subsidy was reflected in lowering costs.

The third intervention featured health education. In different hypothetical scenarios, it was assumed that health education (e.g., through mass media campaigns) would increase the preference for healthy food by 2%, 4%, 6%, and 8% in the population. It was further tested whether these scenarios would differ when: (1) all households were exposed to the intervention or (2) only low-income households were exposed. It was assumed that the intervention was fully effective and that the effect was similar across all exposed households.

Results

Table 1 shows that all model outcomes were close to the data, except the proportion of high-income households that visited a fruit/vegetable store. Also, high-income households traveled further on average than low-income households for food shopping.

Figure 2 summarizes the impact of all interventions on income inequalities in healthy food consumption compared to no intervention. In the baseline scenario, the difference in the proportion of healthy food consumption between high-

and low-income households was 14.8%. Eliminating residential segregation immediately reduced income inequalities in healthy food consumption to approximately 8.0%, which increased again to 10.4% in the long run because of adjustments of food outlets to the new household composition in neighborhoods. Lowering prices of healthy food decreased income inequalities in healthy food consumption to around 14.0% within 5 years and 13.4% in the long term. Health education targeting all households slightly widened income inequalities in healthy food consumption to 14.9%–15.4%, depending on the assumed level of increase in the preference for healthy food. By contrast, health education targeting only low-income households decreased inequalities to 12.2%–14.1%.

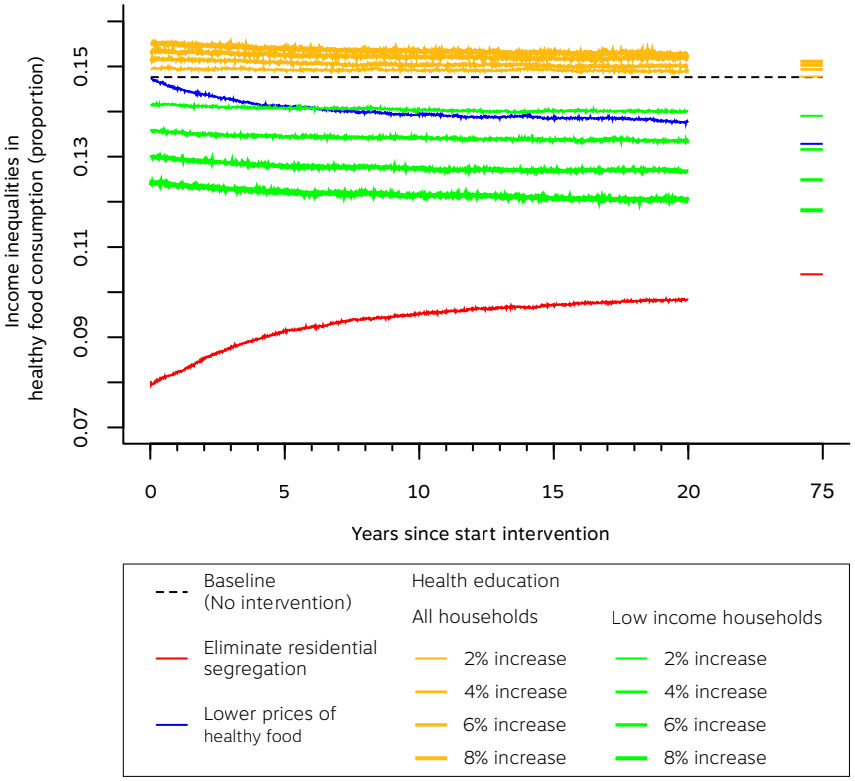


Figure 2. Impact of all interventions on socioeconomic inequalities in healthy food consumption. Model outcomes are the average of 100 runs. A new equilibrium was established after 75 years.

Figure 3 shows the impact of interventions on healthy food consumption of high- and low-income households separately. In the baseline scenario, healthy food consumption for high- and low-income households was 56.5% and 41.7%, respectively. Eliminating residential segregation increased healthy food consumption of low-income households to 44.9%, but decreased healthy food consumption of high-income households to 55.3% (see Figure 3A). This was because a portion of high-income households moved to former low-income neighborhoods with relatively more unhealthy food outlets, resulting in fewer visits to healthy food outlets because of distance. Lowering prices of healthy food only affected low-income households by increasing healthy food consumption to 43.2% (see Figure 3B). Health education targeting all households would increase healthy food consumption 57.3%–61.1% among high-income households and 42.5%–46.0% among low-income households (see Figure 3C). Health education targeting low-income households only increased healthy food consumption in the low-income group (42.0%–44.7%) (see Figure 3D). This increase is slightly lower than when all households are targeted, because overall more households visit healthy food outlets when all households are targeted, causing relatively more healthy food outlets to be opened. This change induced more low-income households to visit healthy food outlets.

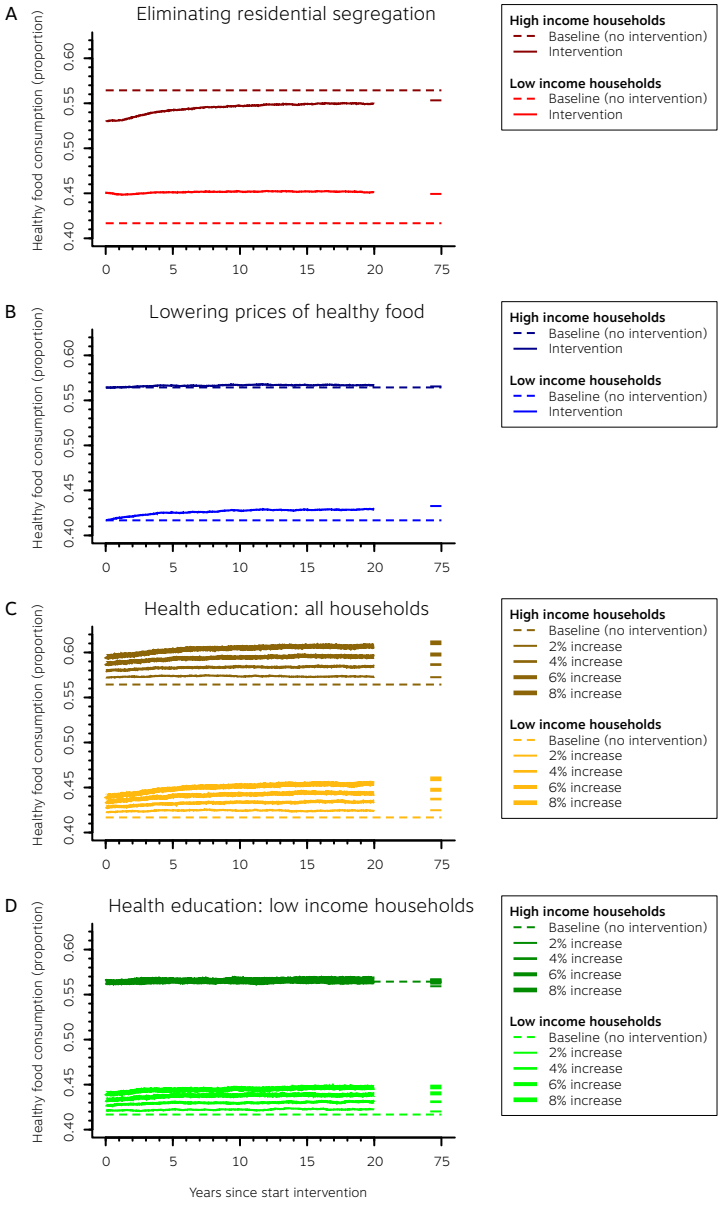


Figure 3. Impact of (A) eliminating residential segregation, (B) lowering prices of healthy food, (C) health education targeting all households and (D) health education targeting low-income households on healthy food consumption among high- and low-income households. Model outcomes are the average of 100 runs. A new equilibrium was established after 75 years.

Discussion

Eliminating residential segregation has the biggest impact on reducing income inequalities in healthy food consumption, but is partly the result of an unfavorable change in healthy food consumption among higher income groups. Although it offers the biggest benefit, it is obviously not very likely that this intervention will be implemented in reality. Lowering prices of healthy food and health education targeting only low-income households also reduces inequalities. Most interventions would take at least 5–10 years to approximate their full effect.

This is the second study to develop an ABM for income inequalities in food consumption. It uses a conceptual framework similar to *Auchincloss et al.*[23] Both studies modeled food shopping behaviors of households and dynamics of food outlets. The present study, however, included fast food visits as a contributor to unhealthy food consumption, and based closures and openings of food outlets more on underlying business principles. Also, habit was excluded as a determinant of food shopping, because it did not contribute to food behaviors in the model.[34] Furthermore, this study drew more on available data and used a realistic grid to represent a city to make it useful for policymakers. Nevertheless, the model still needs to be regarded as "basic" and the results can at most be considered preliminary.

Results of this study showed, similar to *Auchincloss and colleagues*[23], that residential segregation, prices, and preferences of food play an important role in shaping income inequalities in healthy diet. However, this study found a lower income inequality in healthy diet. This may be caused by the aforementioned differences between the models as well as the quantification of the model. In contrast to *Auchincloss et al.*,[23] *HEBSIM* was directly quantified from data on a real population.

Limitations

The application of ABM is still in its infancy and comes with several limitations. First, the lack of data might affect results. For example, fast food data were not available for Eindhoven; therefore, data from U.S. literature were used.[45] It is likely that households in Eindhoven eat fast food less often, meaning that healthy food

consumption might be underestimated.[10] The lack of data was also reflected in the model quantification. Household composition and age could not be quantified and therefore did not contribute to food behaviors, although they might play a role.[9,39] Secondly, households from surrounding villages were neglected. These households may occasionally shop for food in Eindhoven and to some extent influence the dynamics of food outlets.

Furthermore, the tested interventions are based on rather extreme assumptions and therefore not yet sufficient for policy recommendations. For example, it was assumed that each intervention had a perfect reach and was fully effective. In reality, interventions may only reach small or selective parts of the population, and their effectiveness may differ per household.[46,47] Also, an intervention such as eliminating residential segregation is a process that takes time and willingness of households, something the model did yet not take into account. The effect on healthy food consumption might therefore be overestimated in the short term. *HEBSIM*, however, will be able to accommodate such information in subsequent versions.

HEBSIM has substantial room for refinement. An important next step is to model food purchasing behaviors in more detail. This study determined healthy food consumption based on the type of food outlet visited. Obviously, households may still merely consume unhealthy food despite visiting food outlets that are considered healthy. It would be better to explicitly model the type of food households buy at each type of food outlet. This should, however, go hand in hand with data collection on food purchasing behaviors in different types of food outlets.

Another refinement is to allow household attributes, such as location, to change during the simulation. People may select a neighborhood based on safety, for example. Movements of people are associated with health inequalities and weakly with health behaviors.[48] Also, ethnicity that plays a role in explaining differences in food behaviors should be included.[49] Finally, it would be good to consider social environmental factors. Interactions with family or peers have an impact on food behaviors through, for example, social influence.[8,9,15]

Conclusions

This study illustrates that ABM is a promising method for studying dynamic interactions and assessing the impact of interventions to support decision makers. Despite the limitations, the findings of this study provide new insights on how various interventions affect healthy food consumption. With further refinements, *HEBSIM* will become more suitable for decision support.

Acknowledgements

This study was supported by a grant from the Netherlands Organization for Health Research and Development ZonMW (200400015). The *GLOBE* study is carried out by the Department of Public Health of the Erasmus University Medical Centre in Rotterdam, in collaboration with Municipal Public Health Service in the study region (GGD Brabant-Zuidoost). The study has been and is supported by grants of the Ministry of Public Health, Welfare and Sport, the Sick Fund Council, the Netherlands Organization for Advancement of Research, Erasmus University, and the Health Research and Development Council. No financial disclosures were reported by the authors of this paper.

References

1. James WP, Nelson M, Ralph A, Leather S. Socioeconomic determinants of health. The contribution of nutrition to inequalities in health. BMJ 1997; 314(7093): 1545-9.
2. Smith GD, Brunner E. Socio-economic differentials in health: the role of nutrition. Proc Nutr Soc 1997; 56(1A): 75-90.
3. Giskes K, Avendano M, Brug J, Kunst AE. A systematic review of studies on socioeconomic inequalities in dietary intakes associated with weight gain and overweight/obesity conducted among European adults. Obes Rev 2010; 11(6): 413-29.
4. Irala-Estevez JD, Groth M, Johansson L, Oltersdorf U, Prattala R, Martinez-Gonzalez MA. A systematic review of socio-economic differences in food habits in Europe: consumption of fruit and vegetables. Eur J Clin Nutr 2000; 54(9): 706-14.
5. Giskes K, van Lenthe FJ, Kamphuis CB, Huisman M, Brug J, Mackenbach JP. Household and food shopping environments: do they play a role in socioeconomic inequalities in fruit and vegetable consumption? A multilevel study among Dutch adults. J Epidemiol Community Health 2009; 63(2): 113-20.
6. Van Duyn MA, Kristal AR, Dodd K, et al. Association of awareness, intrapersonal and interpersonal factors, and stage of dietary change with fruit and vegetable consumption: a national survey. Am J Health Promot 2001; 16(2): 69-78.
7. Black C, Moon G, Baird J. Dietary inequalities: what is the evidence for the effect of the neighbourhood food environment? Health Place 2014; 27: 229-42.
8. Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? Obes Rev 2011; 12(5): e95-e106.
9. Kamphuis CB, Giskes K, de Bruijn GJ, Wendel-Vos W, Brug J, van Lenthe FJ. Environmental determinants of fruit and vegetable consumption among adults: a systematic review. Br J Nutr 2006; 96(4): 620-35.

10. Turrell G, Giskes K. Socioeconomic disadvantage and the purchase of takeaway food: a multilevel analysis. Appetite 2008; 51(1): 69-81.
11. Cummins S, Macintyre S. Food environments and obesity--neighbourhood or nation? Int J Epidemiol 2006; 35(1): 100-4.
12. Cummins S. Neighbourhood food environment and diet: time for improved conceptual models? Prev Med 2007; 44(3): 196-7.
13. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. Am J Epidemiol 2008; 168(1): 1-8.
14. Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health 2011; 101(9): 1627-34.
15. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating healthy food and eating environments: policy and environmental approaches. Annu Rev Public Health 2008; 29: 253-72.
16. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. Am J Public Health 2006; 96(2): 325-31.
17. Cummins S, Curtis S, Diez-Roux AV, Macintyre S. Understanding and representing 'place' in health research: a relational approach. Soc Sci Med 2007; 65(9): 1825-38.
18. Resnicow K, Page SE. Embracing chaos and complexity: a quantum change for public health. Am J Public Health 2008; 98(8): 1382-9.
19. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. Annu Rev Public Health 2012; 33: 357-76.
20. Blok DJ, van Empelen P, van Lenthe FJ, Richardus JH, de Vlas SJ. Unhealthy behaviour is contagious: an invitation to exploit models for infectious diseases. Epidemiol Infect 2013; 141(3): 667-9.
21. Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 2002; 99 Suppl 3: 7280-7.
22. Macal CM, North MJ. Tutorial on agent-based modelling and simulation. J Simulation 2010; 4(3): 151-62.
23. Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential

segregation. Am J Prev Med 2011; 40(3): 303-11.
24. El-Sayed AM, Seemann L, Scarborough P, Galea S. Are network-based interventions a useful antiobesity strategy? An application of simulation models for causal inference in epidemiology. Am J Epidemiol 2013; 178(2): 287-95.
25. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. A spatial agent-based model for the simulation of adults' daily walking within a city. Am J Prev Med 2011; 40(3): 353-61.
26. Yang Y, Diez-Roux AV. Using an agent-based model to simulate children's active travel to school. Int J Behav Nutr Phys Act 2013; 10: 67.
27. Luke S, Cioffi-Revilla C, Panait L, Sullivan K, Balan G. MASON: a multi-agent simulation environment. Simulation 2005; 82(7): 517-27.
28. Bakker R, Busker T, White RG. Rapid weighted random selection in agent-based models of infectious disease dynamics using augmented B-trees. The fifth international conference on advances in system simulation. Venice, Italy: IARIA; 2013.
29. Statistics Netherlands (CBS). http://statline.cbs.nl/statweb/ (accessed May 18 2015).
30. Statistics Eindhoven. http://eindhoven.buurtmonitor.nl/ (accessed May 18 2015).
31. van Lenthe FJ, Kamphuis CB, Beenackers MA, et al. Cohort Profile: Understanding socioeconomic inequalities in health and health behaviours: the GLOBE study. Int J Epidemiol 2014; 43(3): 721-30.
32. Turrell G. Socioeconomic differences in food preference and their influence on healthy food purchasing choices. J Hum Nutr Diet 1998; 11(2): 135-49.
33. Chaix B, Bean K, Daniel M, et al. Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). PLoS One 2012; 7(4): e32908.
34. EFMI. Consumententrends 2011. Leusden: Centraal Bureau Levensmiddelenhandel, 2011.
35. Morland K, Wing S, Diez Roux A. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. Am J Public Health 2002; 92(11): 1761-7.
36. Dyer J. Maut — Multiattribute Utility Theory. Multiple Criteria Decision

Analysis: State of the Art Surveys: Springer New York; 2005: 265-92.
37. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley; 1976.
38. Kennedy WG. Modelling human behaviour in agent-based models. In: Heppenstall AJ, Crooks AT, See LM, Batty M, eds. Agent-based models of geographical systems: Springer; 2012: 167-79.
39. Glanz K, Basil M, Maibach E, Goldberg J, Snyder D. Why Americans eat what they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. J Am Diet Assoc 1998; 98(10): 1118-26.
40. Reitzel LR, Regan SD, Nguyen N, et al. Density and proximity of fast food restaurants and body mass index among African Americans. Am J Public Health 2014; 104(1): 110-6.
41. Chou SY, Grossman M, Saffer H. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. J Health Econ 2004; 23(3): 565-87.
42. He M, Tucker P, Gilliland J, Irwin JD, Larsen K, Hess P. The influence of local food environments on adolescents' food purchasing behaviors. Int J Environ Res Public Health 2012; 9(4): 1458-71.
43. Dunn RA, Sharkey JR, Horel S. The effect of fast-food availability on fast-food consumption and obesity among rural residents: an analysis by race/ethnicity. Econ Hum Biol 2012; 10(1): 1-13.
44. Chamber of commerce. Brancheinformatie. http://www.ondernemersplein.nl/brancheinformatie/.
45. French SA, Story M, Neumark-Sztainer D, Fulkerson JA, Hannan P. Fast food restaurant use among adolescents: associations with nutrient intake, food choices and behavioral and psychosocial variables. Int J Obes Relat Metab Disord 2001; 25(12): 1823-33.
46. Pomerleau J, Lock K, Knai C, McKee M. Interventions designed to increase adult fruit and vegetable intake can be effective: a systematic review of the literature. J Nutr 2005; 135(10): 2486-95.
47. Ammerman AS, Lindquist CH, Lohr KN, Hersey J. The efficacy of behavioral interventions to modify dietary fat and fruit and vegetable intake: a review of the evidence. Prev Med 2002; 35(1): 25-41.

48. van Lenthe FJ, Martikainen P, Mackenbach JP. Neighbourhood inequalities in health and health-related behaviour: results of selective migration? Health Place 2007; 13(1): 123-37.
49. Brug J, van Stralen MM, Chinapaw MJ, et al. Differences in weight status and energy-balance related behaviours according to ethnic background among adolescents in seven countries in Europe: the ENERGY-project. Pediatr Obes 2012; 7(5): 399-411.

5

SUPPLEMENTARY APPENDIX

Agents and attributes

HEBSIM includes two types of agents: households and food outlets. These agents operate in an environment defined by neighborhoods within a city. Table S1 gives an overview of all attributes in the model.

Environment

The city of Eindhoven, situated in the South-Eastern part of the Netherlands, was modeled. The city is approximately 88 km^2 and includes 116 neighborhoods of which 88 are residential neighborhoods.[1] It is the fifth largest city of the Netherlands with 213,223 inhabitants in 2012. A grid was constructed using a GIS-file (raster-format) of the city of Eindhoven, which was obtained from Statistics Netherlands.[1] This file contains a rasterized representation of all neighborhoods of Eindhoven with actual sizes. Each grid cell is 10m x 10m in size and can be occupied by a household and a food outlet or a vacant (food and non-food) outlet. Unoccupied grid cells are considered "other areas", such as parks and roads.

Distribution of agents

Households, food outlets, and vacant outlets are distributed on the grid following empirical data. Locations of households and vacant outlets were determined by the number of households or vacant outlets per neighborhood obtained from Statistics Eindhoven[2] The exact location within each neighborhood was randomly determined assuming that household locations and vacant outlet locations were clustered (default: 3 clusters per neighborhood). Locations of food outlets were actual locations based on a search in the Yellow Pages in July 2012. A location is defined by x and y coordinates. In 2012, Eindhoven had 97,523 households, 12 fruit/vegetable stores, 33 supermarkets, 22 discount supermarkets, 160 fast foods, and 279 vacant outlets.

Household

Initial conditions of households' attributes are based on data on the population in the city of Eindhoven. Besides a location, each household has three other attributes: (1) household composition; (2) income level; and (3) preference for

healthy food. Household attributes do not change during the simulation.

Household composition

The model included four types of household compositions: single person, single parent, multi-person without children, and multi-person with children (Table S1). The household composition was assigned following an empirical distribution obtained from Statistics Eindhoven.[2] Household size was determined by the sum of the average number of adults and the average number of children per household. The average number of children in Eindhoven was on average 1.7 in 2012 (50,986 children in 29,384 households with children).[2] Household sizes were: 2.7 members in a single parent household, 2 members in a multi-person household without children, and 3.7 members in a multi-person household with children.

Income level

Households were categorized into low- or high-income. Each household was assigned an income based on the distribution of income per neighborhood, obtained from Statistics Netherlands.[1] Low-income was defined as households with a combined net household income below $31,777 per year.

Preference for healthy food

The preference for healthy food represents a household's attitude towards healthy food. The preference was expressed as a continuous number between 0 and 1, where 0 means prefers unhealthy food and 1 means prefers healthy food. We assumed that the average preference for healthy food differs between high- and low-income households. Based on results of *Turrell et al.*,[3] the mean preference for healthy food was set at 0.739 and 0.661 for high- and low-income households, respectively.

The preference for healthy food was randomly assigned to high- and low-income households following a Beta distribution with a mean that is equal to the mean preference for healthy food. The preference of high-income households was drawn from a Beta(4, 1.55) distribution (mean=0.739), assuming that 5.5% have

a score less than 0.4 and 16% a score higher than 0.9. The preference for low-income households was drawn from a Beta (4, 2.06) distribution (mean=0.661), assuming that around 10% of the households have a score less than 0.4 and 7.5% a score higher than 0.9.

Food outlet

The model distinguishes four types of food outlets: (1) fruit/vegetable stores; (2) supermarkets; (3) discount supermarkets; and (4) fast food outlets. Besides a location, each type of food outlet has four other attributes: (1) quality of food; (2) price level; (3) monthly costs; and (4) capital. The model allows the composition of food outlets in a city to change during the simulation. Dynamic processes cause food outlets to close and to start new food outlets (see *"Behavioral rules of food outlets"*). Each simulation run starts with a different number of food outlets allowing the system to arrive at equilibrium, which should be equal to the observed number of food outlets. The starting number of food outlets was randomly drawn from a Uniform distribution: fruit/vegetable stores U(7,17); supermarkets U(28,38); discount supermarkets U(17,27); fast food outlets U(130,190).

Quality of food

The quality of food refers to the level of healthy food sold in a food outlet, which is either mostly "healthy" or mostly "unhealthy". Healthy food outlets mainly sell healthy, while unhealthy food outlets mainly sell unhealthy food. In the model, we chose to categorize fruit/vegetable stores and supermarkets as "healthy", and discount supermarkets and fast foods as "unhealthy". This (crude) assumption was based on findings in the literature[4,5] and a Dutch consumer report.[6] These have shown that people shopping at supermarkets eat and buy healthier food compared to those shopping at discount supermarkets.

Price level

The price level is defined as the average price level of a food outlet. Food outlets were classified into "cheap" or "expensive". Following a Dutch consumer report,[6] the price level of fruit/vegetable stores and supermarkets was set to "expensive" and the price level of discount supermarkets and fast food outlets was set to

"cheap".[6]

Monthly costs

Monthly costs cover all sorts of costs of a food outlet, such as rent. The monthly costs remained unchanged during the simulation and only differed per type of food outlet (i.e., fruit/vegetable stores, supermarket, discount supermarket, and fast food). Since actual costs were unknown, this parameter was calibrated such that the number of food outlets in the steady state matches the observed number of food outlets in Eindhoven.

Capital

Each food outlet has capital that increases with revenue and decreases with monthly costs. At the start of the simulation or when a new food outlet is started, each food outlet is also assigned a starting capital to protect a food outlet from directly closing down during times when it does not have enough customers. We assumed that food outlets may survive up to five months without customers. To account for heterogeneity, starting capital was randomly determined by a random draw between 0 and 5 times the monthly costs (Uniform distribution).

Table S1. Overview of agents' attributes

Attributes	
Household	
Location	Fixed location on grid (x and y coordinates)
Household composition	"Single-person" or "Single parent" or "Multi-person without children" or "Multi-person with children"
Income level	"Low-income" or "High-income"
Preference for healthy food	Range: 0 to 1 (0 = prefer unhealthy food; 1= prefer healthy food); Random draw from Beta distribution
Food outlet	
Location	Location may change during the simulation (x and y coordinates)
Quality of food	
Fruit/vegetable store	"Healthy"
Supermarket	"Healthy"
Discount supermarket	"Unhealthy"
Fast food outlet	"Unhealthy"
Price level	
Fruit/vegetable store	"Expensive"
Supermarket	"Expensive"
Discount supermarket	"Cheap"
Fast food outlet	"Cheap"
Monthly costs	Monthly costs are calibrated
Capital	Starting capital was drawn from a Uniform (0, maximum capital) distribution. Maximum starting capital is equal to 5 times the monthly costs;

Behavioral rules of households

In HEBSIM, two behaviors of households were modeled: food shopping and fast food visits. Behavioral rules of households were based on literature and quantified using empirical data. The multi-attribute utility theory was used to guide interactions between households and food outlets and to determine behaviors.[7,8]

Food shopping

Based on a Dutch consumer report,[6] we assumed that households shop for food three times a week. At each food shopping event, each household selects a food outlet. Households can only shop for food at fruit/vegetable stores, supermarkets, or discount supermarkets. In the Netherlands, fruit/vegetable stores are considered complementary to supermarkets or discount supermarkets.[9] We therefore assumed that each household visits a supermarket or discount supermarket at least once a week. Decisions on where to shop for food are based on three determinants: distance to a food outlet, price of a food outlet, and a household's preference for healthy food. These determinants have been pointed out to be important factors for food behaviors or the decision what to eat.[10,11]

In the model, selection of a food outlet for food shopping by household i is determined by assigning a utility to all eligible food outlets j using a utility function (see Equation 1). The food outlet with the highest utility at that time is selected.

$$\begin{aligned} Utility_{i,j} = {} & w_{distance} \cdot distance\ score_{i,j} + w_{price} \cdot price\ score_{i,j} \\ & + w_{preference} \cdot preference\ score_{i,j} + \varepsilon_{i,j} \end{aligned} \quad (1)$$

The utility function includes determinants of food behavior and a random noise ($\varepsilon_{i,j}$). Each determinant was assigned a score and a weight. Here, the utility of food outlet j is determined by the distance to a food outlet, the price level of a food outlet, and the preference. All scores were assigned based on available data. The weights (i.e. $w_{distance}$, w_{price} and $w_{preference}$) indicate the relative importance of the determinants in the decision-making. Random noise ($\varepsilon_{i,j}$) represents bounded rationality, which causes people to deviate from the model-derived

optimal choice.[12] Random noise was drawn from a Normal distribution with a mean equal to zero and a user-defined standard deviation. Weights and user-defined standard deviation were calibrated such that the proportion of households that visited fruit/vegetable stores, supermarkets, and discount supermarkets matched the data obtained from the GLOBE study (see *"Calibration"*).

Scores

Scores for food shopping were based on the attributes of the household itself and the attributes of the food outlet. All scores were scaled such that the average score is 1.

Distance score

Each household i calculates the Euclidian distance to a food outlet j. The distance score of food outlet j is defined as the inverse of the distance to food outlet:

$$distance\ score_{i,j} = \frac{1}{distance_{i,j}} \quad (2)$$

Price score

The price score depends on a food outlet's price level and household's income level. We assumed that high- and low-income groups value cheap and expensive food outlets differently. Table S2 shows the price scores as used in the model. These scores were obtained from the 2011 wave of the Dutch GLOBE study.[13] Respondents could indicate the importance of different aspects of food on a scale of 1-4, ranging from "not important at all" to "very important". Price scores were based on two of those aspects: importance of cheap food and importance of non-expensive food. A high score on the importance of cheap food indicates that cheap food is important. This score was used as a proxy to value cheap food outlets. A high score on the importance of non-expensive food indicates that non-expensive food is not important. This score was used as a proxy to value expensive food outlets.

Table S2. Price score by income level

	High-income	Low-income
Price		
Cheap	2.21	2.84
Expensive	2.43	1.84

**Price scores are calculated from Dutch GLOBE study.[13]*

Preference score

The preference score depends on the quality of a food outlet and the household's preference for healthy food. We assumed that households with a high preference for healthy food value "healthy" food outlets more than "unhealthy" food outlets, while households with a low preference for healthy food value "unhealthy" food outlets more than "healthy" food outlets. The preference score of food outlet was determined as follows:

$$preference\ score_{i,j} = \begin{cases} preference\ for\ healthy\ food_i, & if\ healthy\ food\ outlet \\ 1 - preference\ for\ healthy\ food_i, & if\ unhealthy\ food\ outlet \end{cases} \quad (3)$$

Fast food visits

Fast food visits were considered daily. The same approach was used to determine fast food visits. Here, each household i assigns a utility to visiting fast food and to a utility to eating or cooking at home. The same utility function was used, but it only includes two determinants: (1) price level of fast food and eating at home, and (2) preference for fast food and eating at home (Equation 4). These have been pointed out to be important determinants of fast food.[10] We further assumed that the preference for fast food was also influenced by the availability of fast food in the neighborhood.[14-17]

$$Utility_{i,j} = w_{preference} \cdot preference\ score_{i,j} + w_{price} \cdot price\ score_{i,j} + \varepsilon_{i,j} \quad (4)$$

The utility of fast food is determined by the preference for fast food and the price score of fast food. Similarly, the utility of eating at home is determined by the

preference score of eating at home and the price score of eating at home. Weights and the random noise were calibrated such that the proportion of households that visits fast food matched empirical data (see *"Calibration"*).

The option with the highest utility was chosen. If visiting a fast food outlet was chosen, a household will also choose which fast food outlet j to visit. Since all fast food outlets only differ from each other in location, this decision was only based on distance. Using the same utility function, a utility is assigned to each fast food outlet only based on distance:

$$Utility_{i,j} = w_{distance} \cdot distance\ score_{i,j} + \varepsilon_{i,j} \quad (5)$$

The fast food outlet with the highest utility was chosen. The score for distance is defined as the inverse Euclidian distance (see Equation 2).

Scores

Scores for fast food visits were based on the attributes of the household itself and the attributes of the fast food visits and eating at home. All scores were scaled such that the average score was 1. These scores were assigned as follows:

Preference score

We assumed that eating or cooking at home is a healthier than visiting a fast food outlet. So, the preference score for eating at home is equal to the households' preference for healthy food:

$$preference\ score_{i,eating\ at\ home} = preference\ for\ healthy\ food_i \quad (6)$$

We assumed that visiting fast food is unhealthy and is influenced by the availability of fast foods in neighborhood n. Households living in a neighborhood with a lot of fast food outlets have a higher preference for fast food compared to households living in neighborhoods with no fast food outlet. Several studies have shown that the availability of fast food outlets is associated with food purchasing behavior and BMI.[14-17] The preference for fast food is therefore calculated as:

$$preference\ score_{i,n,\,fast\ food} = (1 - preference\ for\ healthy\ food_i) + availability\ score_n \quad (7)$$

$$availability\ score_n = \frac{number\ of\ fastfood\ outlets_n}{total\ number\ of\ fastfood\ outlets} \quad (8)$$

Price score

The price score depends on the price level of cooking or eating at home and fast food visit. A fast food visit was considered "cheap". Eating or cooking at home could be either "cheap" or "expensive" depending where the household did their latest food shopping. For example, if a household shopped at a discount supermarket the price level of cooking at home was also considered "cheap". Price score were assigned following Table S2.

Behavioral rules of food outlets

Food outlets can respond to decisions of households in two ways: closing a food outlet or starting a new food outlet.

Closure

Every 30 days, a food outlet can close. The decision to close is determined by a food outlet's capital, which increases with every customer (revenue) and decreases with monthly costs. A food outlet's capital accumulates if revenue exceeds monthly costs and depletes if monthly costs exceed revenue. If capital falls below zero, the food outlet closes. The location of the closed food outlet becomes vacant.

Revenue is determined by the number of customers taking into account household sizes. We assumed that each household member has a weekly food budget of 1. Thus, the total food budget of a household is equal to the household size. The food budget is spent on food shopping and fast food visits. In the model, households shopped for food three times a week and could visit a fast food outlet up to six times a week. The minimal amount that household i could spend at food outlet j was therefore defined as the food budget of household i divided

by nine:

$$Min.\ amount\ spent_{i,j} = \frac{1}{9} \cdot\ food\ budget_i \tag{9}$$

We assumed that households spend less at fast food outlets and fruit/vegetable stores compared to supermarkets and discount supermarkets.[6] Also, households that visit fast food outlets frequently may spend less on food shopping. The revenue R of food outlet j obtained by a visit of household i is therefore calculated as:

$$R_{i,j} = \begin{cases} Min.\ amount\ spent_{i,j} & ,\ if\ fast\ food\ or\ fruit\ /\ vegetable\ store \\ \dfrac{Food\ budget_i - \left[Min.\ amount\ spent_{i,j} \cdot \left(fast\ food\ \&\ fvstore\ frequency_i \right) \right]}{\left(supermarket\ \&\ disc.supermarket\ frequency_i \right)} & ,\ if\ supermarket\ or\ disc.supermarket \end{cases} \tag{10}$$

New food outlet

Every 30 days, new food outlets can be started at a vacant location. The decision to start a food outlet is determined by an empirical probability distribution per type of food outlet. The monthly probability to start a new food outlet per type of food outlet was calculated using data of the Chamber of Commerce.[18] The data include the fraction of total food outlets that started a new food outlet and the number of new food outlets per month per type of food outlet from April 2011 to May 2013 in the province of Noord-Brabant, in which Eindhoven is situated. The number of food outlet (per type) in Eindhoven was multiplied with the fraction of total food outlets that started a new food outlet to obtain the number of new food outlets in the city of Eindhoven. Using this number, we calculated the monthly probability to open 0 new stores, 1 new store or 2 new stores per type of food outlet (Table S3).

Table S3. Monthly probability of starting a new food outlet for fruit/vegetables stores, supermarkets, discount supermarkets, and fast food outlets

Number of new food outlets	Monthly probability			
	Fruit/vegetable store	Supermarket	Discount supermarket	Fast food outlet
0	0.96	0.86	0.91	0.00
1	0.04	0.14	0.09	0.91
2	-	-	-	0.09

The location of a new food outlet will be selected in the neighborhood with the highest expected revenue. The expected revenue of a food outlet of type k (i.e. fruit/vegetable store, supermarket, discount supermarket, fast food outlet) in neighborhood n is based on the total revenue in the past month and the total number of food outlets in the neighborhood:

$$Expected\ revenue_{k,n} = \frac{Total\ revenue_{k,n}}{Number\ of\ food\ outlets_{n} + 1} \quad (11)$$

The exact location of the new food outlet is randomly determined at one of the vacant locations in the selected neighborhood.

Calibration

Calibration was performed through an iterative process until all model outcomes matched the data obtained from the Dutch GLOBE study and literature. Model outcomes include:

1. The proportion of households visited a
    A. Fruit/vegetable store
    B. Supermarket
    C. Discount supermarket at least once in the past week
2. The proportion of households visited a fast food outlet
    A. 0 times
    B. 1-2 times
    C. >2 times in the past week
3. The proportion of households that shopped at one type of food outlet in

the past week

4. The number of
    A. Fruit/vegetable stores
    B. Supermarkets
    C. Discount supermarkets
    D. Fast foods in the city of Eindhoven (see Table 1 in Chapter 5).

Parameters for which no empirical data were available were calibrated such that the model outcomes were consistent with the empirical data. Table S4 presents all calibrated parameters of the model and their optimal values.

The weights of food shopping and fast food behaviors and random noise were calibrated such that the proportion of households that visited food outlets matched the data from the GLOBE study (Table 1 in Chapter 5). From the calibration process it turned out that preference was the most important determinant followed by price and distance. This finding is in line with our knowledge from data and literature.[11,19] A recent discrete choice experiment among participant of the GLOBE study also showed that preference or taste is the most important determinant followed by price, and finally distance (results not shown). Also, literature suggests that preference or taste is the most important determinant, followed by price or cost.[11,19]

The optimal values of the random noise after calibration were 0.8 and 0.6 for food shopping and fast food visiting, respectively. Since the scores of the determinants were on average 1, this indicates that a substantial part is explained by bounded rationality (approximately 40%). However, other determinants, such as lack of time, mood, social support, or social influences, may also play a role.[20,21]

Monthly costs per type of food outlet were calibrated to arrive at the observed number of food outlets accounting for the rate of new start-ups. Table S4 shows that supermarkets have the highest monthly costs, followed by discount supermarket and fruit/vegetable stores. Fast food outlets have the lowest monthly costs. It is very reasonable to accept that this is also the pattern that is present in the real world. Supermarkets and discount supermarkets are in general larger and have more personnel and thus have higher monthly costs compared to fruit/

vegetable stores and fast food outlets.

Table S4. Values of Calibrated Parameters

Calibrated parameters	Values
Food shopping	
Weight of distance	0.2
Weight of price	0.35
Weight of preference for healthy	0.45
Fast food visits	
Weight of preference for fast food	0.6
Weight of price	0.4
Random noise	
Food shopping	0.8
Fast food visits	0.6
Monthly costs	
Fruit/vegetable store	4800
Supermarket	15000
Discount supermarket	8300
Fast food outlet	1025

References

1. Statistics Netherlands (CBS). http://statline.cbs.nl/statweb/ (accessed May 18 2015).
2. Statistics Eindhoven. http://eindhoven.buurtmonitor.nl/ (accessed May 18 2015).
3. Turrell G. Socioeconomic differences in food preference and their influence on healthy food purchasing choices. J Hum Nutr Diet 1998; 11(2): 135-49.
4. Chaix B, Bean K, Daniel M, et al. Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). PLoS One 2012; 7(4): e32908.
5. Morland K, Wing S, Diez Roux A. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. Am J Public Health 2002; 92(11): 1761-7.
6. EFMI. Consumententrends 2011. Leusden: Centraal Bureau Levensmiddelenhandel, 2011.
7. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley; 1976.
8. Dyer J. Maut — Multiattribute Utility Theory. Multiple Criteria Decision Analysis: State of the Art Surveys: Springer New York; 2005: 265-92.
9. Deloitte Branchegroep Retail. Consumentenonderzoek 2013: Deloitte, 2013.
10. Story M, Neumark-Sztainer D, French S. Individual and environmental influences on adolescent eating behaviors. J Am Diet Assoc 2002; 102(3 Suppl): S40-51.
11. Glanz K, Basil M, Maibach E, Goldberg J, Snyder D. Why Americans eat what they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. J Am Diet Assoc 1998; 98(10): 1118-26.
12. Kennedy WG. Modelling human behaviour in agent-based models. In: Heppenstall AJ, Crooks AT, See LM, Batty M, eds. Agent-based models of geographical systems: Springer; 2012: 167-79.
13. van Lenthe FJ, Kamphuis CB, Beenackers MA, et al. Cohort Profile:

Understanding socioeconomic inequalities in health and health behaviours: the GLOBE study. Int J Epidemiol 2014; 43(3): 721-30.
14. Reitzel LR, Regan SD, Nguyen N, et al. Density and proximity of fast food restaurants and body mass index among African Americans. Am J Public Health 2014; 104(1): 110-6.
15. Chou SY, Grossman M, Saffer H. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. J Health Econ 2004; 23(3): 565-87.
16. He M, Tucker P, Gilliland J, Irwin JD, Larsen K, Hess P. The influence of local food environments on adolescents' food purchasing behaviors. Int J Environ Res Public Health 2012; 9(4): 1458-71.
17. Dunn RA, Sharkey JR, Horel S. The effect of fast-food availability on fast-food consumption and obesity among rural residents: an analysis by race/ethnicity. Econ Hum Biol 2012; 10(1): 1-13.
18. Chamber of commerce. Brancheinformatie. http://www.ondernemersplein.nl/brancheinformatie.
19. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating healthy food and eating environments: policy and environmental approaches. Annu Rev Public Health 2008; 29: 253-72.
20. Steenhuis IH, Waterlander WE, de Mul A. Consumer food choices: the role of price and pricing strategies. Public Health Nutr 2011; 14(12): 2220-6.
21. Kamphuis CB, van Lenthe FJ, Giskes K, Brug J, Mackenbach JP. Perceived environmental determinants of physical activity and fruit and vegetable consumption among high and low socioeconomic groups in the Netherlands. Health Place 2007; 13(2): 493-503.

6

THE IMPACT OF INDIVIDUAL AND ENVIRONMENTAL INTERVENTIONS ON INCOME INEQUALITIES IN SPORTS PARTICIPATION: EXPLORATIONS WITH AN AGENT-BASED MODEL

D.J. BLOK, F.J. VAN LENTHE, S.J. DE VLAS

Submitted for Publication

Abstract

Background

Income inequalities in sports participation are shaped by a system in which individuals and the environment interact. We developed an agent-based model (ABM) that could represent this system and used it to explore the impact of individual and environmental interventions on reducing inequalities in sports participation.

Methods

Our ABM simulates sports participation of individuals in the Dutch city of Eindhoven. In the model, sports participation is determined by an individual's intention to start sports (at a fitness center, sports club or self-organized), which is influenced by attributes of individuals (i.e. age, sex, income), sports facilities (i.e. price, accessibility) and the social environment (i.e. social cohesion, social influence). Sports facilities can adapt to changes in the demand by closures or startups, which in turn influence the intention of individuals to sport. We explored the impact of five interventions scenarios.

Results

Providing health education, lowering prices of sports facilities, increasing availability of sports facilities and increasing safety levels could increase sports participation from 63.1% to 65.7%, 63.5%, 64.1% and 63.5%, respectively. Combining all interventions could increase sports participation with 4.1% points (from 63.1% to 67.2%) and reduces absolute income inequalities in sports participation from 11.3% to 10.1%. Marked effects are only achieved after five to ten years.

Conclusions

Our findings highlight that increasing sports participation and reducing income inequalities in sports participation requires sustained effort with the generally modest population-level effects only visible in the long-term. Our study illustrates the potential of ABMs for testing the long-term effect of interventions of complex public health problems.

Background

Physical activity, including sports participation, prevents obesity, diabetes, cardiovascular diseases, and several cancers.[1,2] Consistent evidence shows that adults with a lower socioeconomic position (SEP) are less likely to participate in physical activity and sports than their counterparts with a higher SEP.[3] The reduction of socioeconomic inequalities in health behaviors through the promotion of sport among those in lower socioeconomic groups is a major challenge in public health.

Socioeconomic inequalities in sports participation are shaped by individual and environmental factors.[4-6] According to theories of behavioral change, the intention to participate in sport is the most proximal individual determinant of sports participation, which varies by age, sex and SEP.[7] More distal environmental characteristics of the built environment, such as the proximity to and price levels of sports facilities, and (the perception of) social safety contribute to the explanation of socioeconomic inequalities in sports participation.[7-9] Furthermore, social environmental factors, such as social cohesion and social influence, are relevant determinants of sports participation.[9,10] Lower socioeconomic groups are more likely to have a small social network and low social cohesion.[8]

It is increasingly recognized that many of these determinants interact with and feedback on each other, creating a complex causal web.[4,6] For example, people are to some extent sorted in neighborhoods based on characteristics, such as age, income, causing spatial clustering.[11] Characteristics of the individual and the residential neighborhood influence sports participation behaviors of individuals. As a feedback to individual behavior, the availability of sports facilities in neighborhood may change, which subsequently influences sports participation behavior, thereby changing social influences (or norms) that in turn may affect sports participation. To be able to identify optimal ways to promote sports participation and to decrease inequalities in sports participation, there is a need to account for this complex non-linear system.

Agent-based models (ABMs) can simulate a dynamic system in which individuals interact with each other and their environment.[4,12,13] An ABM contains heterogeneous agents (here: individuals and sports facilities) with specific

characteristics (e.g. income level) and behaviors (e.g. sports participation) that can be followed over time. Behavioral rules describe how individuals interact with each other and the environment. An ABM captures feedback loops and adaptations of agents, e.g. a behavioral change of agents based on changing environments (such as more sports facilities).[14] In addition, it can test the (long term) effects of intervention scenarios and compare them to a counterfactual.[4,15] Recent ABM studies in social epidemiology have focused on dietary behaviors,[16-18] social networks and obesity[19,20] and daily walking.[21,22] Thus far, no ABM has been developed to model sports participation.

In this paper we present a new agent-based model within the Health Behaviors Simulation (*HEBSIM*) suite.[23] Our aim is to represent a system that simulates sports participation among adults with different income levels to study income inequalities in sports participation, as emerging from interactions between individuals and sports facilities in neighborhoods of a city. The model is used to explore the long term impact of interventions targeting the intention of individuals.

Methods

General modeling approach

We modeled the city of Eindhoven in the Netherlands with its 88 residential neighborhoods using GIS data obtained from Statistics Netherlands.[24] The time step in the model represents one month. During the simulation, individuals become older, can die, and move out and into the city, based on patterns observed in empirical data.[24,25] Individuals entering the simulation do not participate in sport. During their life course, all individuals can start, quit and restart sports participation in three categories of sports as illustrated in Figure 1: fitness, sports club (e.g. football, tennis), and self-organized (e.g. running).[26] Whether, when and how often (i.e. monthly or weekly) an individual engages in sports participation is determined by the intention to do sports (the main determinant of sport according to the Theory of Planned Behavior[27]). The intention to do sports results from interactions between attributes of individuals (i.e. age, sex, income), sports facilities (i.e. price, accessibility), and the (social) environment (i.e. safety, social cohesion, social influence), as proposed in a social

ecological approach.[28,29] Older age groups, females and lower income groups may have lower intention to do sports.[30] Expensive and remote sports facilities, lower neighborhood safety levels and lower social cohesion levels are modeled as barriers of sports participation.[7-9,31,32] The model also includes the social influence of direct neighbors, which may discourage or encourage sports participation.[10] In response to sports participation behaviors of individuals, sports facilities can open or close over time, to which in turn individuals change sports participation behaviors. The text below provides an overview of the agent's attributes, sports participation and behaviors of sports facilities. A detailed model description can be found in the Supplementary Appendix.

Attributes

The modeled population consists of 173,567 individuals between the age of 18 and 85 years, which were distributed according to the observed number per neighborhood in Eindhoven in 2014.[25] Each individual is characterized by age, sex, income level and intention to do sports. Age, sex, and income were assigned to each individual based on the distribution per neighborhood: the average age of the total population was 46 years, 49% was female, and 41%, 40% and 19% of the population had a low-, middle-, and high-income, respectively.[25] Each individual is assigned an intention to do sports at creation. To account for heterogeneity in intention, an individual's intention was drawn from a Gamma distribution with a mean of one (see *"Attributes of individuals"* in the Supplementary Appendix). The individual's intention to do sports is then multiplied with the average number of times people participate in sports per year in the city. As a result, those with low intentions are less likely to do sports than those with higher intentions. The average number of times people participate in sports was calibrated against observed data on the overall sports participation in Eindhoven.[33] This product is called intention henceforth.

Figure 1. Example of twelve individuals as they may exist in the model in a selected period. Each bar represents the life of one individual, who can decide to start sports, quit sports or change the frequency of sports in the category fitness, sports club or self-organized. Individuals are grouped by income level, i.e. high-, middle- and low-income. As individuals get older, they can die or move out of the city. Such individuals are replaced by new individuals to keep the population constant. The vertical dotted line represents the moment of the start of four interventions: providing health education, lowering prices of sports facilities, increasing availability of sports facilities and improving neighborhood safety.

Sports facilities consist of fitness centers and sports club facilities, and can only be placed at a designated location for fitness centers and sports club facilities. These locations were assigned on the grid based on the actual number per neighborhood: i.e. 305 fitness center locations and 98 sports club locations in total. At the start of the simulation, fitness centers and sports club facilities are created based on the actual number of existing sports facilities per neighborhood in 2016: 30 fitness centers and 158 sport club facilities.[34,35] The price level of a facility (dichotomized as "cheap" or "expensive") was assigned based on the fraction of expensive facilities (0.37 for fitness centers and 0.40 for sports club facilities). This fraction was determined by either the average monthly contribution-fee of a sports facility (cut-off: €20/month) or by the nature of the facility (see *"Attributes of sports facilities"* in the Supplementary Appendix).[36]

Modeling sports participation

For each category of sports, the individual's intention was multiplied by an age group, sex, income, price, accessibility, safety, social cohesion and social influence score (see *"Starting sports participation"* in the Supplementary Appendix). Age group scores were assigned to the young (18-35yrs), middle (35-55yrs) and old (55-85yrs) age group, sex scores to males and females, and income scores to the high-, middle- and low-income group. These scores were determined by calibrating the model against observed sports participation by age group, sex, and income level.

The price and accessibility scores depend on the selected fitness center or sports club facility. We assumed that sports participation in the category self-organized is not influenced by the price and accessibility score. To choose which sports facility to go to, individuals rank all fitness centers or sports club facilities based on a preference score and select the best fitness center or sports club facility (see *"Selection of a sports facility"* in the Supplementary Appendix). The price score is set to 1.0, if the price level of the facility is cheap, and 0.85 if it is expensive. The latter is estimated using data from the *GLOBE* study.[37] The accessibility score is measured as , where is the distance to a sports facility, and is the distance decay. The distance decay of sports participation in the categories fitness and sports club were calibrated such that model outcomes match sports participation

by categories of sports.

The perceived safety and social cohesion scores of an individual's neighborhood were both derived from data.[25] Lastly, the intention was assumed to increase with the social influence score, which is measured as the proportion of direct neighbors that engage in sports. Direct neighbors are defined as all those that live within a 50-meter radius of the individual.

The time between entering the model and starting sports in a category of sports was modeled as exponential random variables with the category-specific intention as rate. For example, an individual with a very low intention may never start sports in a lifetime, while an individual with a very high intention is very likely to start or restart sports soon. At the time of starting sports participation, the individual is either categorized into monthly or weekly sports participation (see *"Frequency of sports participation"* in the Supplementary Appendix).

Quitting sports participation is determined by annual quitting probabilities. These were chosen at 0.28, 0.12 and 0.27 for the category fitness, sports club, and self-organized, respectively, based on national Dutch data.[38] Individuals who quit sports can restart sports after some time. The time between quitting sports and restarting sports is determined in a similar way to starting sports. In addition, individuals can also change the frequency of sports at the end of every year after starting sports. The probability of changing the frequency was derived from data about the intention to increase (i.e. from monthly to weekly) or decrease (i.e. from weekly to monthly) the frequency: 0.21 and 0.09, respectively.[38] An individual can also decrease or increase the frequency, whenever he/she starts multiple sports or quits sports when engaged in multiple sports. This occurs with an assumed probability of 0.5 (see *"Quitting sports participation"* and *"Changing frequency of sports participation"* in the Supplementary Appendix).

Modeling behaviors of sports facilities

During the simulation, new sports facilities can open and existing sports facilities can close in the city, and thus altering the composition of sports facilities. We assumed that on average one fitness center and one sports club facility close every year, and that on average also one fitness center and one sports club

facility open in the city every year. In our model, always the facility with the fewest members closes, upon which the location becomes vacant. A new facility is opened at a vacant location of a neighborhood with the highest demand for sports (see *"Sports facility closures and startups"* in the Supplementary Appendix).

Model calibration and outcome

The calibration process was performed using a grid search in which eight unknown parameters were calibrated under three assumptions of variation in intention (see *"Model calibration"* in the Supplementary Appendix). The model was run for 50 years to make sure it reached equilibrium. Model outcomes in equilibrium were matched to the observed overall sports participation and sports participation by age group, sex, income and category of sports in Eindhoven.[33] Results in this paper are based on the best fitted model, i.e. assuming a Gamma(1.0, 0.5) for intention scores (see Table 1, Table S2 and Figure S4 in the Supplementary Appendix).

Table 1. Calibrated model parameters

Parameters	Value	(95% CI [a])
Overall intention, mean	153.5	(146.2 – 160.9)
Age group score		
18-35yrs	1.0	-
35-54yrs	0.157	(0.139 – 0.175)
>55yrs	0.148	(0.136 – 0.160)
Sex score		
Male	1.0	-
Female	0.659	(0.608 – 0.710)
Income score		
High-income	1.0	-
Middle-income	0.471	(0.428 – 0.514)
Low-income	0.428	(0.387 – 0.470)
Distance delay score		
Fitness center	0.029	(0.028 – 0.031)
Sports club facility	0.027	(0.025 – 0.028)

[a] *95% confidence interval*

The outcome of interest is the annual overall sports participation and sports participation by income level. Absolute income inequality was calculated as the difference between sports participation in the high- and low-income group. Final model outcomes were the result of the average of 80 simulation runs. Intervals reflecting 95% uncertainty ranges were constructed by discarding the two highest and lowest outcome values.

Interventions

Five intervention scenarios were explored: (1) providing health education, (2) lowering prices of sports facilities, (3) increasing the availability of sports facilities, (4) improving neighborhood safety and (5) combining all previous interventions simultaneously (i.e. multilevel intervention).[29] Results of intervention scenarios were compared to a scenario with no intervention for a period of 25 years, assuming everything remains unchanged. All interventions were assumed to be of immediate effect.

1. Health education was modeled as increasing the intention to start sports participation. As we wanted to examine reasonable changes in intention, an increase of an individual's original intention by a factor of 1.5 was assumed. This factor was informed by the difference between the intention of high- and low-income individuals, which crudely differ a factor of 2 (see Table 1). The intervention was provided to 15% of the individuals who do not participate in sports. In separate scenarios, we further examined more optimistic scenarios in which the effect factor was gradually increased from 1.5 to 3 and the reach from 15% to 50%.
2. Lowering prices of sports facilities was modeled by changing the price level of "expensive" fitness centers and sports club facilities to "cheap". As sports facilities are cheap from then on, it takes away the barrier of price level.
3. Increasing the availability of sports facilities in neighborhoods with low sports participation is hypothesized to increase sports participation, because it reduces distance.[31] In this scenario, one fitness center and one sports club facility were added to the five neighborhoods with the lowest sports participation. The newly created facilities are protected from closure in the next 10 years. In separate scenarios, we also explored the option of

increasing the number of neighborhoods with new sports facilities.

4. Improving safety levels reduces a barrier to start sports participation.[7] In this scenario, the perceived neighborhood safety score was increased to the mean safety score of the entire city. This intervention only applies to neighborhoods with a safety score that is below the mean at the time of the start of the intervention. We also explored the impact of increasing the safety score to that of the neighborhood with the highest safety.

Results

Figure 2 presents the modeled impact of five intervention scenarios on sports participation. Providing health education, lowering the prices of sports facilities, increasing the availability of sports facilities and improving safety could increase sports participation from 63.1% to 65.7%, 63.5%, 64.1% and 63.5%, respectively. The large uncertainty around these predictions indicates that there is a chance that lowering prices, increasing availability of facilities and safety may have no effect on sports participation at a population level, but the effect may also increase to 66.1%, 66.5%, and 66.1%, respectively. Combined interventions could yield an increase of 4.1% points (from 63.1% to 67.2%) in sports participation, which equals an additional 7100 individuals starting sports. The effects of interventions gradually increase over time with any marked effects usually only achieved after 5 to 10 years.

Figure 2B shows the impact after 25 years specified by the category of sports. Interventions aimed at sports facilities only increase sports participation in the category fitness and sports club. In a combined approach, sports participation in the category fitness increased from 17.5% to 22.1%, sports club from 34.0% to 40.3% and self-organized from 52.4% to 54.9%.

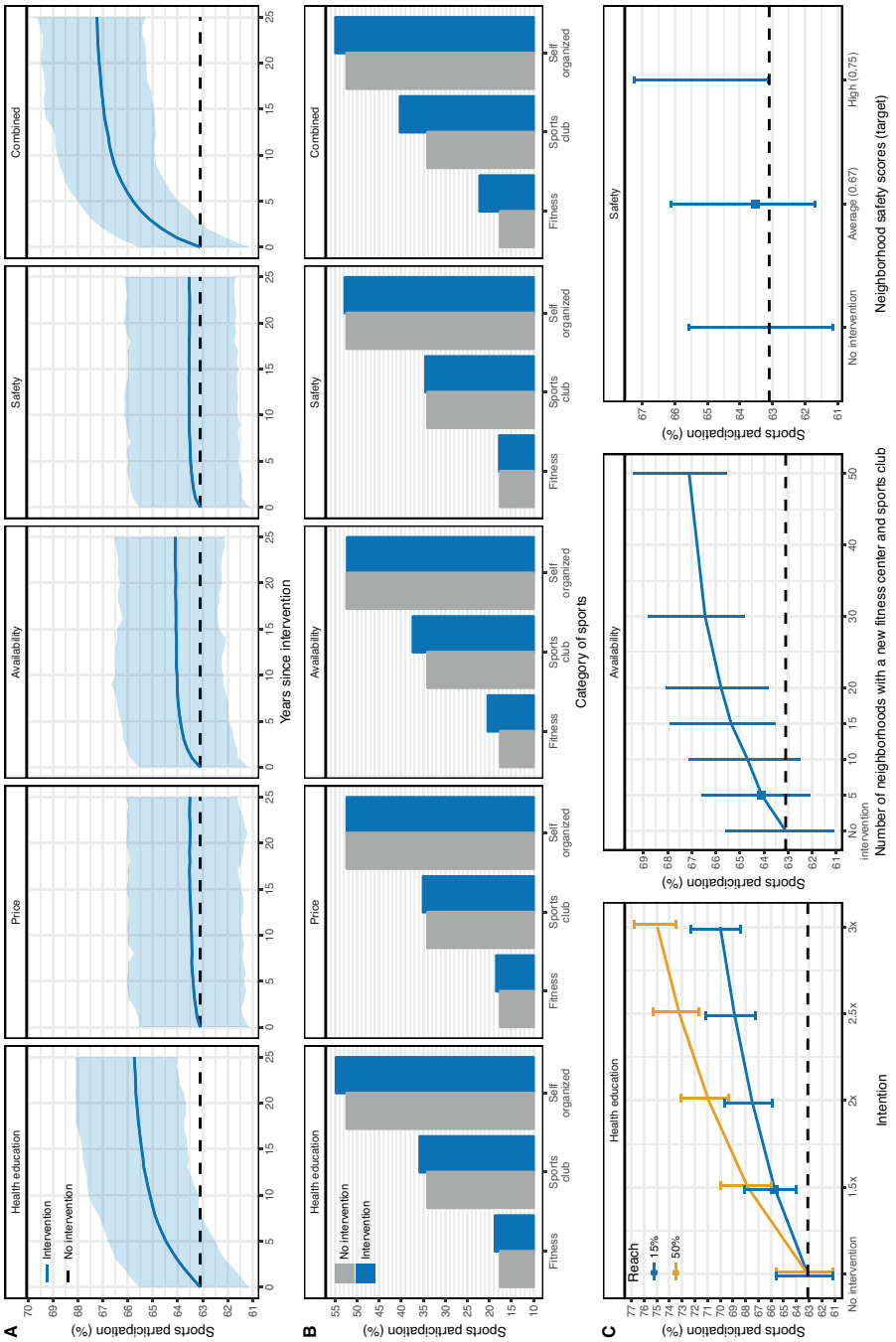

A
B
C
Health education
Price
Availability
Safety
Combined
Intervention
No intervention
Sports participation (%)
Years since intervention
Category of sports
Fitness
Sports club
Self organized
Reach
15%
50%
Intention
No intervention
1.5x
2x
2.5x
3x
Number of neighborhoods with a new fitness center and sports club
No intervention
Average (0.67)
High (0.75)
Neighborhood safety scores (target)

Figure 2. Predicted impact of five intervention scenarios on sports participation. (A) Impact on the proportion of total sports participation over time; (B) Impact on the proportion of sports participation by the category of sports after 25 years compared; (C) Impact of alternative intervention scenarios of health education, availability of sports facilities, and safety levels on sports participation after 25 years. Interventions scenarios include: (1) providing health education (effectiveness: 1.5x original intention; reach: 15%), (2) lowering price level of expensive sports facilities to cheap; (3) increasing availability of sports facilities in five neighborhoods; (4) improving safety (target: average perceived safety score); (5) combining all previous interventions. The shaded area represents the 95% uncertainty interval due to parameter uncertainty and stochastic variation. Alternative intervention scenarios include: (1) varying effect of health education on an individual's intention (1.5x, 2x, 2.5x, 3x original intention) and its reach (15 and 50%); (2) increasing the number of neighborhoods with new sports facilities (10, 15, 20, 30, 50); (3) increasing the target level of safety (high level). The squares represent the outcomes of the intervention scenarios, and circles represent the outcomes of alternative intervention scenarios. The error bars represent the 95% interval due to parameter uncertainty and stochastic variation.

Doubling the effect of health education on intention would increase sports participation to 70% (see Figure 2C). Also, increasing the reach to 50% would more than double the impact. Building sports facilities in more neighborhoods would increase sports participation up to 67.1%. However, the additional gain for doing this in an additional neighborhood diminishes after 15 neighborhoods. Further increasing the perceived safety score to 0.75 in all neighborhoods would increase sports participation to 64.9%.

Figure 3 shows the impact on income inequalities in sports participation. At baseline the modeled sports participation is 71.6%, 61.8% and 60.3% for the high-, middle- and low- income group, respectively. This corresponds to an absolute inequality between the high and low income groups of 11.3%. All interventions show a larger increase in lower income groups compared to the high income group, indicating a decrease in absolute income inequalities (see Figure 3B). Again, combining all interventions yields the largest impact: absolute income inequalities in sports participation then drop to 10.1% after 25 years.

Absolute income inequalities in sports participation within neighborhoods are predicted to decrease in almost all of the neighborhoods (see Figure 4). However, neighborhoods with large inequalities at baseline will continue to have the largest inequalities after the interventions. These neighborhoods are primarily situated at the periphery of the city.

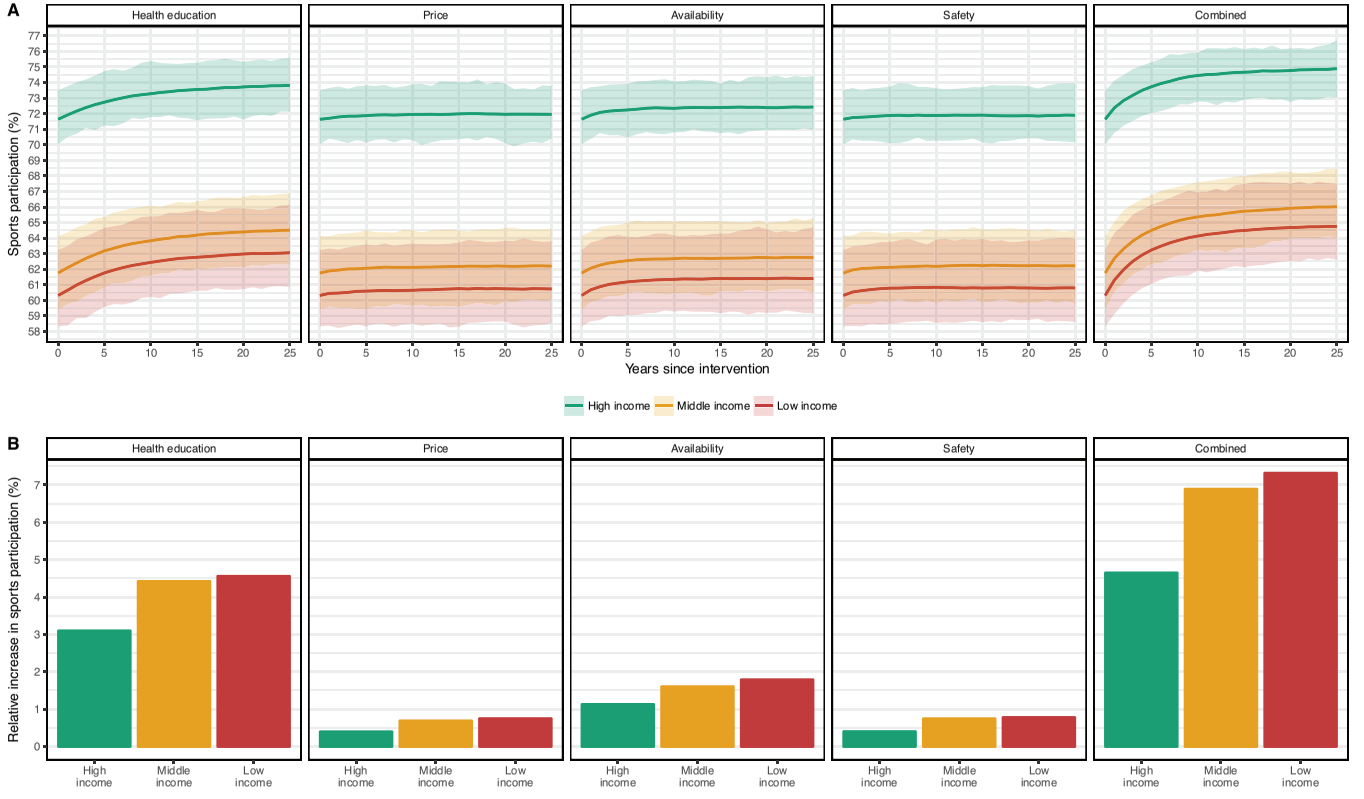


Figure 3. Predicted impact of interventions on sports participation rates by income group. (A) Impact on the proportion of total sports participation by income group over time; (B) Relative increase in sports participation after 25 years by income group; Interventions scenarios include: (1) providing health education (effectiveness: 1.5x original intention; reach: 15%), (2) lowering price level of expensive sports facilities to cheap; (3) increasing availability in five neighborhoods; (4) improving safety (target: average perceived safety score); (5) combining all previous interventions. The shaded area represents the 95% uncertainty interval, which reflects parameter uncertainty and stochastic variation.

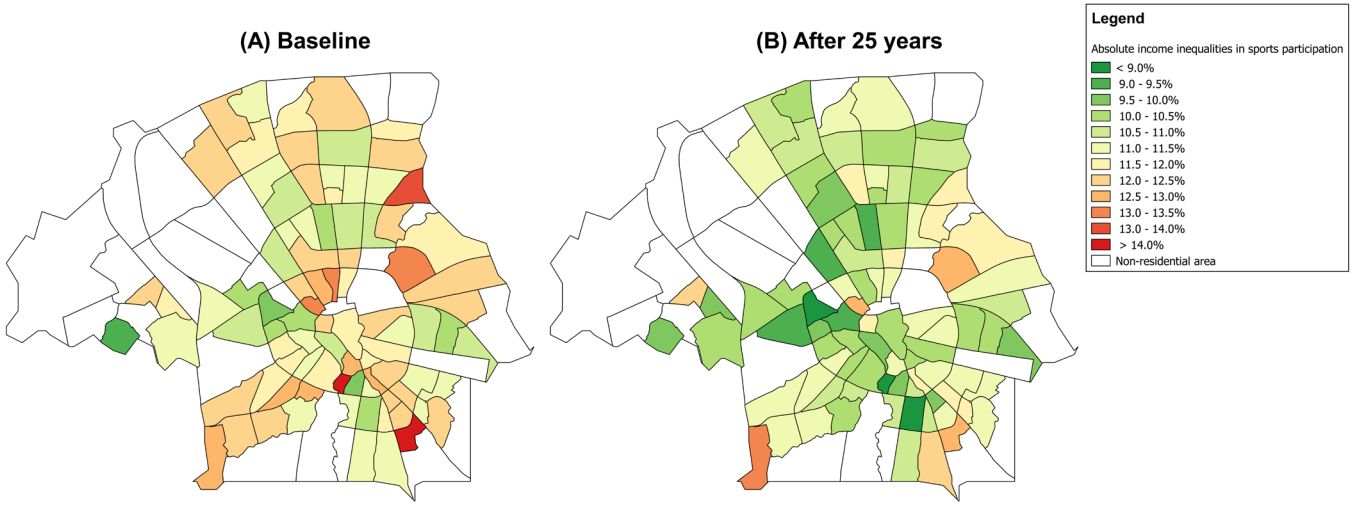


Figure 4. Predicted impact of interventions on absolute income inequalities in sports participation rates. (A) Absolute income inequalities of sports participation at baseline; (B) absolute income inequalities of sports participation after 25 years with combined interventions. Absolute income inequalities are the difference between the proportion of sports participation in the high and low income group.

Discussion

This study illustrates the potential of agent-based modeling to explore the population-level effects of various interventions in the context of a system. Our model predicts that providing health education, increasing the availability of sports facilities, lowering prices of facilities and improving safety levels can increase sports participation. Also, they modestly reduce absolute income inequalities in sports participation. The model suggests that the largest gain can be attained through health education, if the effect and reach is sufficiently large. Interventions targeting prices of sports facilities, the availability of sports facilities and safety levels solely generally have a modest impact.

The relatively large impact of health education may be surprising but can be explained because we assumed a large effect size, i.e. the intention increases by a factor of 1.5. An effect of this size, however, may not be feasible in practice yet, indicating that much work still needs to be done in this area. Our findings also illustrate that interventions may not have large effects on its own, and that an approach in which health education is combined with environmental interventions is more favorable. A single intervention generally does not sufficiently increase individuals' intention to start sports participation, whereas a multilevel intervention approach increases it to a level high enough to start sports. This result is in line with the key premises of social ecological models, which suggest that a multilevel approach is more beneficial compared to a single intervention approach.[29] As empirical evidence on the effect of multilevel interventions remains limited, mainly due to difficulties to design and conduct controlled experiments, agent-based modeling can aid in testing the added value of multilevel interventions.[29,39]

Significant population-level effects are usually obtained after 5 or 10 years, which supports the general idea that policies and interventions may affect population health slowly.[40] It should be noted that we even assumed the interventions to have an immediate impact, so that the real impact likely appears after an even longer time. Still, we found that combined interventions could eventually increase sports participation by as much as 4.1% points, which equals around 7,100 individuals in this city of 173,000 residents. It should be realized that it concerns a population level effect, and as such cannot be compared with the results of

efficacy trials. These explorative findings suggest that much work still needs to be done to increase sports participation at population-level, in particular through making health education more effective and increasing the reach (see Figure 2C). As in most cases direct evidence for the long-term population-level effect of an intervention is unknown, ABMs can be useful to predict population-level effects of interventions studies.

In terms of reducing absolute income inequalities in sports participation, the impact of interventions can be seen as modest at most. The reason for this is that interventions increase sports participation in all income groups, and only slightly more in the low-income group. It highlights that strategies to increase the overall sports participation may have not the desired impact on income inequalities in sports participation. If the aim is to decrease income inequalities in sports participation, a shift to tailored interventions might be needed: i.e. more directly targeting interventions to this group.

Health education is known to be more effective in higher than lower socioeconomic groups. Nevertheless, our results show a slightly larger impact in the low-income group than in the high-income group. This can be explained by the fact that in our model relatively more individuals from the low-income group are reached by the intervention. This is because the intervention is only provided to individuals who do not engage in sports and the number of such individuals is higher in the low-income group than the high-income group. This illustrates that a population-wide approach of an intervention that is known to have a greater impact on high-income individuals may still reduce income inequalities in sports participation as long as it mainly affects individuals from the low-income group.

As our model is intended as a proof-of-concept, there is still substantial room for refinements. First, better data are needed to validate modeled processes. Although our model was able to replicate the observed sports participation by age group, sex, income and category of sports in Eindhoven, the data used were cross-sectional and therefore did not inform us about how accurate starting and quitting sports over time were modeled. This would require longitudinal data about how often individuals start and quit sports over time. Also, our model predictions of baseline absolute income inequalities in sports participation at a neighborhood

level were not validated, because of lack of data on sports participation by income and category of sports at a neighborhood level. Furthermore, population-level effects of our interventions were not validated to data, not even in the short-term, as such data are not available. Information on interventions might become available through natural experiments that evaluates population health interventions.[41]

Second, several aspects of our model are extremely simplified. For example, starting sports in each of the three categories of sports (i.e. fitness, sports club, self-organized) is modeled completely independently. Also, changing frequency and quitting sports does not depend on an individual's intentions. To model these processes more accurately, data are needed that follow individuals' sports participation behavior in different categories of sports over time and include information about key drivers of starting, quitting and changing frequency of sports participation. Another simplification is that price levels of sports facilities are dichotomized. This may to some extent explain why the impact of lowering prices of sports facilities on sports participation turned out to be very modest. It currently primarily affects individuals living in proximity to expensive sports facilities with the majority of sports facilities in the city being cheap. A continuous price level might have been a better approach. However, this would require data about the relationship between price levels and sports participation. Another simplification is the assumption that social influence is only exerted by direct neighbors. Although previous studies showed the importance of neighbors for health behaviors, this may not fully reflect the impact of an individual's social network.[10,42] Close contacts, such as friends, are not considered in this model, even though this has been argued to be most influential with regard to health behaviors.[42] In order to model close contacts, network data is required which also include sports participation behaviors of the respondents.

Third, our model does not account for all relevant processes that might influence of sports participation among adults. For example, it is known that sports participation in adulthood is partly determined by the sports participation during childhood or adolescence. Moreover, these decisions in childhood are largely influenced by the parents' SEP, which partly explains income inequalities at later stages of life. Further extending the model by incorporating a life course perspective might therefore be an important next step.[43,44]

Fourth, the interventions explored in this study have strong assumptions. For example, we assumed that all interventions have an immediate effect, while in reality this likely is a much slower process which may even take years (e.g. building new facilities). Our predicted time frame for interventions to reach large impacts may therefore be underestimated. Also, we assumed that the composition of sports facilities is optimal at baseline. Our model was run for 50 years to reach equilibrium allowing the locations of sports facilities to be optimized through closures and startups. The impact of interventions targeting sports facilities may have been slightly larger if we would have started from a situation in which the composition of sports facilities is suboptimal. Furthermore, long term projections of interventions should always be interpreted with care, as populations may change over time in terms of size, age structure and income distribution per neighborhood. Therefore, our (long-term) findings cannot be interpreted for policy-purposes yet, and should at most be considered indicative.

We conclude that ABMs have potential for developing and testing the population-level effects of various interventions in the context of a system. Our findings highlight that increasing sports participation and reducing income inequalities in sports participation requires sustained effort with population-level effects only being visible in the long-term. Our study also illustrates the level of complexity of an ABM and highlights gaps in empirical data. With further refinements, ABMs may eventually become useful tools to support decision-makers in answering questions in public health arising from complex interactions.

Acknowledgements

This study was supported by a grant from the Netherlands Organization for Health Research and Development ZonMW (200400015).

References

1. Wannamethee SG, Shaper AG. Physical activity in the prevention of cardiovascular disease: an epidemiological perspective. Sports Med 2001; 31(2): 101-14.
2. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006; 174(6): 801-9.
3. Mackenbach JP, Stirbu I, Roskam AJ, et al. Socioeconomic inequalities in health in 22 European countries. N Engl J Med 2008; 358(23): 2468-81.
4. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. Am J Epidemiol 2008; 168(1): 1-8.
5. Spence JC, Lee RE. Toward a comprehensive model of physical activity. Psychol Sport Exerc 2003; 4: 7-24.
6. Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health 2011; 101(9): 1627-34.
7. Beenackers MA, Kamphuis CB, Burdorf A, Mackenbach JP, van Lenthe FJ. Sports participation, perceived neighborhood safety, and individual cognitions: how do they interact? Int J Behav Nutr Phys Act 2011; 8: 76.
8. Kamphuis CB, Van Lenthe FJ, Giskes K, Huisman M, Brug J, Mackenbach JP. Socioeconomic status, environmental and individual factors, and sports participation. Med Sci Sports Exerc 2008; 40(1): 71-81.
9. Wendel-Vos W, Droomers M, Kremers S, Brug J, van Lenthe F. Potential environmental determinants of physical activity in adults: a systematic review. Obes Rev 2007; 8(5): 425-40.
10. Blok DJ, de Vlas SJ, van Empelen P, Richardus JH, van Lenthe FJ. Changes in smoking, sports participation and overweight: does neighborhood prevalence matter? Health Place 2013; 23: 33-8.
11. van Lenthe FJ, Martikainen P, Mackenbach JP. Neighbourhood inequalities in health and health-related behaviour: results of selective migration? Health Place 2007; 13(1): 123-37.
12. Blok DJ, P vE, FJ vL, Richardus JH, SJ dV. Unhealthy behaviour is contagious: an invitation to exploit models for infectious diseases. Epidemiol Infect 2013; 141(3): 667-9.

13. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. Annu Rev Public Health 2012; 33: 357-76.
14. Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 2002; 99 Suppl 3: 7280-7.
15. Nianogo RA, Arah OA. Agent-based modeling of noncommunicable diseases: a systematic review. Am J Public Health 2015; 105(3): e20-31.
16. Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential segregation. Am J Prev Med 2011; 40(3): 303-11.
17. Zhang D, Giabbanelli PJ, Arah OA, Zimmerman FJ. Impact of different policies on unhealthy dietary behaviors in an urban adult population: an agent-based simulation model. Am J Public Health 2014; 104(7): 1217-22.
18. Blok DJ, de Vlas SJ, Bakker R, van Lenthe FJ. Reducing Income Inequalities in Food Consumption: Explorations With an Agent-Based Model. Am J Prev Med 2015; 49(4): 605-13.
19. El-Sayed AM, Seemann L, Scarborough P, Galea S. Are network-based interventions a useful antiobesity strategy? An application of simulation models for causal inference in epidemiology. Am J Epidemiol 2013; 178(2): 287-95.
20. Zhang J, Tong L, Lamberson PJ, Durazo-Arvizu RA, Luke A, Shoham DA. Leveraging social influence to address overweight and obesity using agent-based models: the role of adolescent social networks. Soc Sci Med 2015; 125: 203-13.
21. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. A spatial agent-based model for the simulation of adults' daily walking within a city. Am J Prev Med 2011; 40(3): 353-61.
22. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. Exploring walking differences by socioeconomic status using a spatial agent-based model. Health Place 2012; 18(1): 96-9.
23. Blok DJ, de Vlas SJ, Bakker R, van lenthe FJ. Reducing income inequalities in food consumption: Explorations with an agent-based model. AM J Prev Med 2015.
24. Statistics Netherlands. http://statline.cbs.nl/statweb/.

25. Statistics Eindhoven. http://eindhoven.buurtmonitor.nl.
26. Directorate-General for Education and Culture. Special Eurobarometer 412 'Sport and physical activity'. Brussels: European Commission, 2014.
27. Azjen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991; 50(2): 179-211.
28. Stokols D. Translating social ecological theory into guidelines for community health promotion. Am J Health Promot 1996; 10(4): 282-98.
29. Sallis J, Owen N, Fisher E. Ecological models of health behavior. In: Glanz K, Rimer B, Viswanath K, eds. Health behavior and health education. 4th ed. San Francisco: Jossey-Bas; 2008.
30. Breuer C, Hallmann K, Wicker P, Feiler S. Socio-economic patterns of sport demand and ageing. Eur Rev Aging Phys A 2010; 7(2): 61-70.
31. Karusisi N, Thomas F, Meline J, Chaix B. Spatial accessibility to specific sport facilities and corresponding sport practice: the RECORD Study. Int J Behav Nutr Phys Act 2013; 10(1): 48.
32. Prins RG, van Empelen P, Te Velde SJ, et al. Availability of sports facilities as moderator of the intention-sports participation relationship among adolescents. Health Educ Res 2010; 25(3): 489-97.
33. BIO Eindhoven. Sportparticipatie (2014), 2015.
34. Comensis. Fitness register. http://fitnessregister.nl/ (accessed January 2016).
35. Municipality of Eindhoven. Eindhoven sport. https://eindhovensport.nl/ (accessed January 2016).
36. Van Spronsen & Partners. Het fitnesscentrum in beeld: Van Spronsen & Partners, 2013.
37. van Lenthe FJ, Kamphuis CB, Beenackers MA, et al. Cohort Profile: Understanding socioeconomic inequalities in health and health behaviours: the GLOBE study. Int J Epidemiol 2014; 43(3): 721-30.
38. NOC*NSF. Sportersmonitor 2012: NOC*NSF, 2013.
39. Glass TA, McAtee MJ. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. Soc Sci Med 2006; 62(7): 1650-71.
40. Sterman JD. Learning from evidence in a complex world. Am J Public Health 2006; 96(3): 505-14.

41. Craig P, Cooper C, Gunnell D, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health 2012; 66(12): 1182-6.
42. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007; 357(4): 370-9.
43. Tammelin T, Nayha S, Hills AP, Jarvelin MR. Adolescent participation in sports and adult physical activity. Am J Prev Med 2003; 24(1): 22-8.
44. Telama R, Yang X, Viikari J, Valimaki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood: a 21-year tracking study. Am J Prev Med 2005; 28(3): 267-73.

6

SUPPLEMENTARY APPENDIX

General approach

Individuals and sports facilities were modelled as agents that interact with each other within neighborhoods of a city. We simulated the city of Eindhoven, situated in the South-Eastern part of the Netherlands. A grid was constructed based on a GIS map of Eindhoven, obtained from Statistics Netherlands.[1] The size of the city is 88km^2 with 116 neighborhoods of which 88 are residential neighborhoods. The map was rasterized such that each grid cell represents 10 m x 10m in size. Each grid cell can be occupied by an individual and/or a sports facility.

The simulated population includes individuals between the age of 18 and 85 years. The population of Eindhoven consist of 173,567 persons of age 18-85 years in 2014, living across 88 residential neighborhoods.[2] The numbers per neighborhood range from 138 to 4,965 (see Figure S1A). The actual number of individuals was randomly placed in a neighborhood. Each individual is characterized by attributes including age, sex, income, intention and location.

The model considers two types of sports facilities: fitness centers and sports club facilities (e.g. football, tennis). Fitness centers and sports club facilities can only be placed at a designated location for fitness centers and sports club facilities, respectively. These locations are assigned on the grid based on the actual number per neighborhood: i.e. 305 fitness center locations and 98 sports club locations in total (see Figure S1B and Figure S1C).[2,3] At the start of the simulation fitness centers and sports club facilities are created based on the actual number of existing sports facilities per neighborhood: 30 fitness centers and 158 sports club facilities. These were identified by accessing the national fitness register and sports club database of Eindhoven in February 2016. Only one fitness center is allowed to be placed at one designated fitness center location, but multiple sports clubs are allowed to be placed at one designated sports club location, as is the case in the city of Eindhoven. Sports facilities are characterized by price level and location.

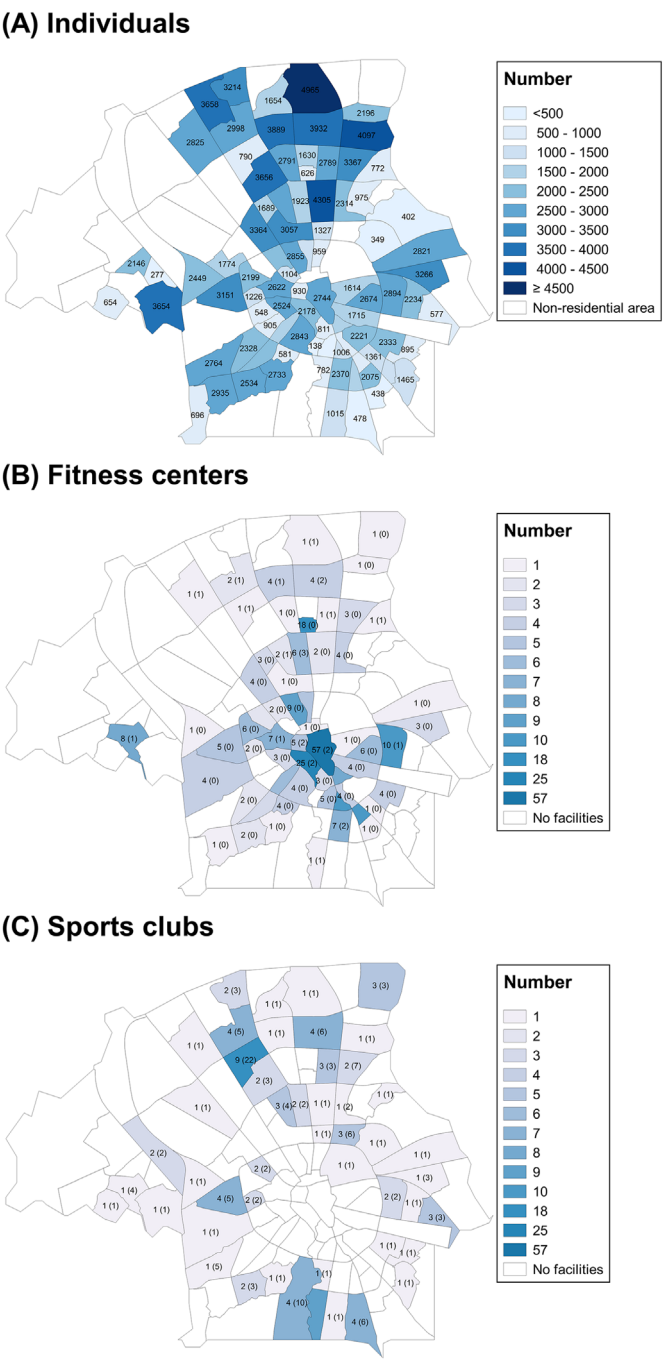


Figure S1. Map of Eindhoven with (A) the number of individuals between the age of 18-85 years, (B) the number of fitness center locations, (C) the number of sports club locations per neighborhood. The map contains in total 116 neighborhoods. Blue neighborhoods in (A) represent residential neighborhoods (in total 88). The blue areas in (B) and (C) represent neighborhoods with fitness center locations and sports club locations, respectively. The numbers between brackets represent the number of fitness centers and sports club facilities at the start of the simulation. Facilities can only be located in neighborhoods with known locations. Darker areas represent higher numbers per neighborhood. Source: Municipality of Eindhoven,[2] national fitness register[5] and sports club database of Eindhoven[3]

In the model, individuals get older, die, or move out of the city, and they can engage in sports participation. During their life course, these individuals can start, quit and restart sports participation in three categories of sports: fitness, sports club (e.g. football, tennis), and self-organized (e.g. running).[4] Whether, when and how often (i.e. monthly or weekly) an individual engages in sports participation is determined by the intention to do sports. In response to sports participation behaviors of individuals, sports facilities can open or close over time, to which in turn individuals may change their sports participation behaviors.

Attributes of individuals

Each individual is characterized by individual-level attributes, including age, sex, income level and intention to do sports. Age, sex, and income were assigned to each individual based on the observed distribution per neighborhood (see Table S1).[1,2]

Table S1. Characteristics of the modelled population

Parameter	City level	Neighborhood level (Range[a]: min – max)	Source
Age (mean=46 years)			2
18-35yrs	33%	4 – 64%	
35-55yrs	35%	20 – 61%	
55-85yrs	32%	7 – 67%	
Sex			
Female	49%	39 – 57%	2
Income			1
Low	41%	16 – 54%	
Modal	40%	12 – 50%	
High	19%	5 – 55%	

[a] *Range between 88 neighborhoods*

The intention to sports of individual i was calculated as:

$$intention_i = nSports \cdot intentionS_i \quad (1)$$

With:

$nSports$ is the mean number of times per year individuals participate in sports in the city. This parameter was calibrated such that model outcomes match the observed sports participation in Eindhoven.

$intentionS_i$ represents the individual intention score. Each individual has an intention score, which follows a Gamma distribution with mean 1.0 and shape equal to k, i.e. Gamma(1.0, k). Lower values of k represent more individual variation. We calibrated the model under three assumptions of the shape parameter k: 0.5, 1.0, 3.0 (see *"Model calibration"*).

Attributes of sports facilities

The price level, either "cheap" or "expensive", of a sports facility was assigned based on the fraction of expensive fitness centers and sports club facilities. For fitness centers, this fraction was determined by the average monthly contribution-fee of a fitness center. All fitness centers with a price-level above €20 per month were considered expensive.[6] The fraction of expensive sports club facilities was determined by the type of sports: golf, tennis and equestrian sports were considered relatively expensive, while football, swimming, athletics were considered relatively cheap. The fractions of expensive fitness centers and sports club facilities were 0.37 and 0.40, respectively.

Demographic processes

During the simulation, individuals become older, and can die, move out of the city and new individuals can move into the city. The age of death of an individual was determined upon entrance to the city based on the Dutch survival curve of 2014 (see Figure S2).[1] At the age of death, the individual is removed from the simulation. Also, all individuals above the age of 85 years are removed from the modelled population, as the model only includes individuals between the age of 18 and 85 years.

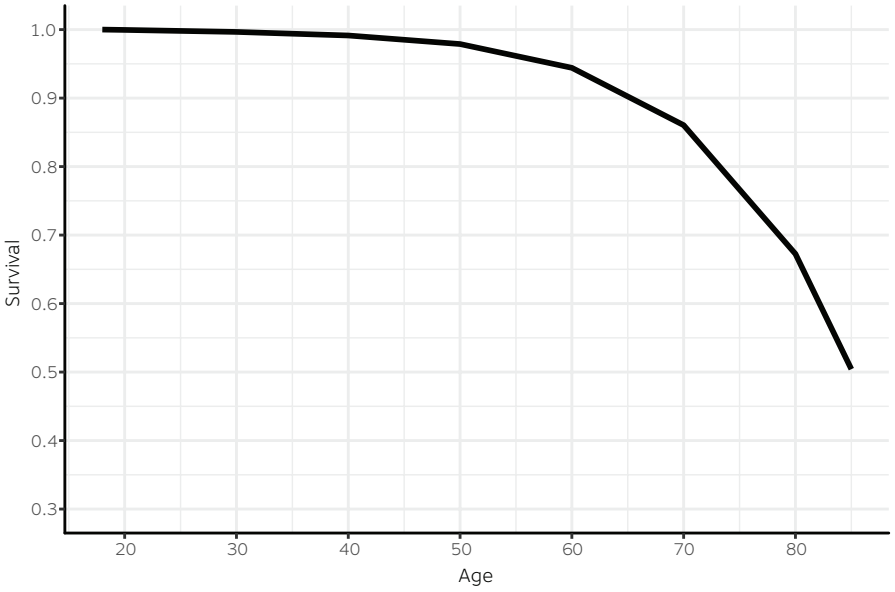


Figure S2. Survival rate of people at the age of 18 years. Source: Statistics Netherlands (CBS)[1]

Every year a fraction of the individuals migrates to another city. This fraction was derived from the observed annual numbers of out-migration per neighborhood to other cities. Around 7.5% of the population migrates to another city per year (range between neighborhoods: 2 – 18%).[2]

To keep the population per neighborhood at a constant size throughout the simulation runs, individuals that are removed (either due to death, old age or migration) are replaced by new individuals. The age and income level of the new individual adheres to the observed age, sex and income distribution of the neighborhood. This assures that also the neighborhood age, sex and income distribution remain unchanged.

Starting sports participation

In our model, individuals can start sports in one or more categories c:

1. Fitness
2. Sports club (e.g. football, tennis)

3. Self-organized (e.g. running)

As a result, each individual has three intentions: intention of doing sports in the category (1) fitness, (2) sports club and (3) self-organized. The intentions are influenced by attributes of individuals (i.e. age, sex, income), and the (social) environment (i.e. safety, social cohesion, social influence). Additionally, the intention of doing sports in the category fitness and sports club depend on the characteristics of fitness centers and sports club facilities (i.e. price level, accessibility), respectively. Therefore, to calculate the intention of doing sports in the category fitness and sports club, each individual selects a fitness center and sports club facility first (see *"Selection of a sports facility"*).

To determine whether and when an individual starts sports, a duration until starting sports participation is drawn following an exponential distribution based on an individual's intention for each of the three categories of sports separately (see *"Intentions"* and *"Time of starting sports"*). Each intention of an individual is in this way translated into a waiting time until starting sports. As the duration until sports participation is drawn for each of the three categories of sports separately, an individual can engage in sports in one or more categories of sports simultaneously. The duration until starting sports for each category of sports c is determined in the following situations:

1. At the time of entering the simulation. An individual enters the simulation either at the start of the model or whenever he/she migrates into the city. Each individual starts without engaging in any sports participation. A duration until starting sports is then determined for each of the three categories of sports separately.
2. At the time an individual quits sports participation in the category of sports c. Upon quitting sports, a duration until restarting sports of that same category c is calculated.
3. At the time the fitness center or sports club facility, of which the individual is a member, closes. A closure forces an individual to quit sports in the category of sports to which the facility belongs to. Upon quitting sports, a new duration until restarting sports in that same category of sports is calculated.

4. At the time a new fitness center or sports club enters the simulation and the individual is not engaged in sports in the category fitness or sports club yet. The individual determines whether the new fitness center or sports club facility is preferred over the currently selected fitness center or sports club facility. If that is the case, a new duration until starting sports is drawn.

Selection of a sports facility

A preference score is assigned to all fitness centers or sports club facilities in the city using Equation (2). The fitness center or sports club facility with the highest preference score is selected.

$$Preference_{i,f,c} = \frac{1}{d(l_i, l_{f,c})} \cdot priceS(p_{f,c}) + \varepsilon_f \quad (2)$$

With:

$d(l_i, l_{f,c})$ is the Euclidian distance between the location l of individual i and the location l of sports facility f of category of sports c (i.e. fitness or sports club).

$priceS(p_{f,c})$ is the price score, which is determined by the price level p of sports facility f of category of sports c. The price score was pre-set at 1.0, if the price level p is cheap. The price score of expensive sports facilities was derived from the GLOBE study (wave 2004).[7] Respondents indicated whether an expensive facility was considered a barrier to start physical activity. In total, 15% of the respondents indicated this was case. Based on this result, we set the price score of expensive facilities to 0.85, which can be considered as a crude proxy.

ε_f is a random variable following a Normal distribution with μ =0 and σ =0.05. This "random noise" was added to represent bounded rationality. Individuals do not always choose a sports facility with perfect rationality.[8]

Intentions

The intention of doing sports of individual i at time t is calculated for each of the three categories of sports c separately. If the category of sports c is fitness or sports club, it is calculated as:

$$\begin{aligned} Intention(t)_{i,c,f} = {} & intention_i \cdot a\_groupS(a(t)_i) \cdot sexS(sx_i) \cdot incomeS(il_i) \\ & \cdot accessS(l_i, l_{f,c}, b_c) \cdot priceS(p_{f,c}) \cdot safetyS(n_i) \cdot s\_cohesionS(n_i) \\ & \cdot s\_influenceS(t)_i \end{aligned} \quad (3a)$$

The intention of doing sports in the category self-organized does not depend on characteristics of a sports facility, and is therefore calculated as:

$$\begin{aligned} Intention(t)_{i,c} = {} & intention_i \cdot a\_groupS(a(t)_i) \cdot sexS(sx_i) \cdot incomeS(il_i) \\ & \cdot safetyS(n_i) \cdot s\_cohesionS(n_i) \cdot s\_influenceS(t)_i \end{aligned} \quad (3b)$$

With:

$intention_i$ is the intention to do sports of individual i. It reflects the (unadjusted) number of times an individual is planning to participate in sports per year. See *"Attributes of individuals"* for more information about how individual intention is assigned.

$a\_groupS(a(t)_i)$ is the age group score of an individual i with age a at time t. Age was categorized here into three groups: young (18-35yrs), middle (35-55yrs) and old (55-85yrs). The score of the age group young was pre-fixed at 1.0. The remaining scores were calibrated to match sports participation by age group.

$sexS(sx_i)$ is the sex score of an individual i with sex sx. The score of males was pre-fixed at 1.0. The score of females was calibrated to match sports participation by sex.

$incomeS(il_i)$ is the income score of an individual i with income level il. Income was categorized into three groups: high, middle and low income. The score of the high income group was pre-set at 1.0. The scores of the remaining income categories were calibrated to match sports participation by income level.

$accessS(l_i, l_{f,c}, b_c)$ is the accessibility score, which is measured as:

$$accessS(l_i, l_{f,c}, b_c) = e^{-\beta_c \cdot d(l_i, l_{f,c})} \quad (4)$$

With:

$d(l_i, l_{f,c})$ is the distance between the location l of individual i and the location l of sports facility f (i.e. fitness center, when c =fitness, or sports club facility, when c =sports club).

β_c is the distance decay of category of sports c (i.e. fitness or sports club). The distance decay of the category fitness and sports club were calibrated to match the observed proportion of people doing sports in the category fitness and sports club.

Accessibility was not considered for sports in the category self-organized (e.g. running), because it can be started from home.

$priceS(p_{f,c})$ is the price score, which is based on the price level p of the selected sports facility f (i.e. fitness center, when c = fitness, or sports club facility, when c =sports club). The price score was pre-set at 1.0, if the price level p is cheap. The price score of expensive sports facilities was set to 0.85, based on the GLOBE study (see *"Selection of a sports facility"*).[7] As the intention of doing sports in the category self-organized is not influenced by price score, we implicitly assume that the price level is "cheap".

$safetyS(n_i)$ is the safety score of the residential neighborhood n of individual i. This score is based on the perceived neighborhood safety as derived from the data from the Municipality of Eindhoven.[2] The perceived safety was assessed by a survey among residents of Eindhoven. The score ranges from 0.0 to 1.0, where 1.0 reflects perfect safety. Figure S3A shows the perceived safety score per neighborhood. The mean perceived safety score of Eindhoven is 0.67.

$s\_cohesionS(n_i)$ is the social cohesion score of the residential neighborhood n of individual i. This score is based on the perceived neighborhood social cohesion level as derived from the data from the Municipality of Eindhoven.[2] The perceived social cohesion score was assessed by a survey among residents of Eindhoven and included questions about social connections, trust, feeling at home.[2] The score ranges from 0.0 to 1.0, where 1.0 is perfect social cohesion. Figure S3B shows the social cohesion score per neighborhood.

$s\_influenceS(t)_i$ is the social influence score at time t. The intention increases proportionally with the number of direct neighbors engaging in sports. The social influence score is measured as:

$$s\_influenceS(t)_i = 1 + \alpha \quad \frac{DN(t)_i}{N(t)_i} \tag{5}$$

With:

α indicates the strength of social influence. In this study, we set α to 1.0, which implies that social influence can increase the intention by a factor of two. This is reasonable as the difference in intention between individuals with a high income is twice as high as those with a low income (see Table 1 in Chapter 6).

$DN(t)_i$ is the total number of direct neighbors of individual i that engage in sports at time t. Direct neighbors are those that are living within a 50-meter radius of the individual.

$N(t)_i$ is the total number of direct neighbors of individual i at time t.

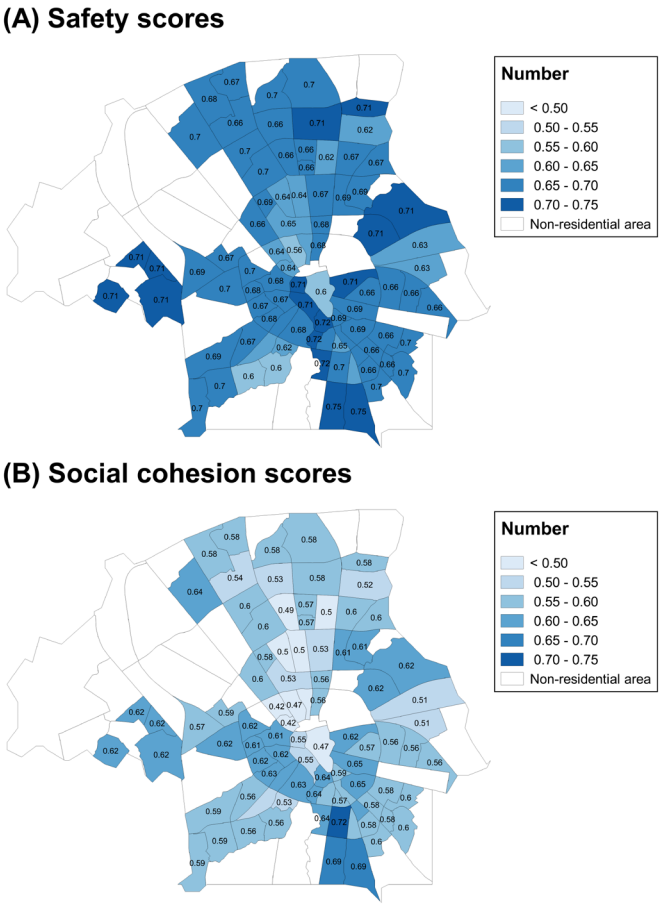


Figure S3. Perceived neighborhood (A) safety scores and (B) social cohesion scores. Source: Municipality of Eindhoven.[2]

Time of starting sports

The time of starting sports participation of individual i in category of sports c (i.e. fitness, sports club, self-organized) is determined as follows:

$$time\_sport_{i,c} = t + duration_{i,c} \quad (6)$$

With:

t	is the current time.
$duration_{i,c} \sim Exponential(\lambda_{i,c})$	is the duration in years until starting sports by individual i in category of sports c. The duration until starting sports follows an exponential distribution with rate equal to: $1/Intention_{i,c,f}$ for the category fitness and sports club, and $1/Intention_{i,c}$ for the category self-organized.

Frequency of sports participation

At the time of starting sports in a category of sports c, the frequency of sports participation is determined for that particular category of sports. The individual is categorized into either: (1) "monthly sports participation", or (2) "weekly sports participation". Frequency was determined by the mean number of times per year an individual does sports which is based on an individual's intention of doing sports in that particular category of sports. The individual is categorized into "monthly sports participation", if this number is between 12 and 40, and "weekly sports participation", if above 40. If this number is smaller than 12, we assumed that the individual does not participate in sports in that particular category of sports, and a new duration until starting sports in that particular category of sports is calculated.

Quitting sports participation

Quitting sports can occur during the following events:

1. At the end of every year since the start of sports participation
2. At the time a sports facility closes (only for sports participation in the category fitness or sports club)
3. At the time of starting sports in another category of sports

At the end of every year since the start of sports participation, an individual quits sports participation with a probability of 0.28, 0.12, and 0.27 in the category fitness, sports club, and self-organized, respectively. These probabilities were

based on a survey that included questions about the intention to quit sports in the next year.[9] A closure of a sports facility forces an individual to quit sports in either the category fitness or sports club. Upon quitting sports in a particular category of sports, a new duration until restarting sports for that particular category is calculated (see *"Starting sports participation"*).

Starting sports in another category of sports can cause an individual to quit current sports, if the frequency of current sports is "monthly". The reasoning behind this is that due to possible time constraints an individual may decide to quit sports he/she currently engages in. The probability of quitting current sports as a result of starting sports in another category was arbitrarily chosen to be 0.5.

Changing frequency of sports participation

Changing frequency of sports participation (i.e. "monthly" or "weekly") can occur during the following events:

1. At the end of every year since the start of sports participation
2. At the time of starting sports in another category of sports
3. At the time of quitting sports while an individual still does sports in another category of sports

At the end of every year since the start of sports participation in any category of sports, an individual can increase frequency from "monthly" to "weekly" and decrease frequency from "weekly" to "monthly" with a probability of 0.21 and 0.08, respectively. These probabilities were derived from a Dutch survey asking respondents whether they would increase, decrease their current frequency of sports, or are happy with their current frequency of sports.[9]

Starting sports at another category of sports can decrease the frequency of current sports from "weekly" to "monthly". The probability of decreasing the frequency of current sports when starting sports in another category of sports was arbitrarily chosen to be 0.5.

Quitting sports in a particular category of sports (either fitness, sports club or self-organized) while still engaging in sports of another category may trigger an

increase in frequency of sports from “monthly” to “weekly”. The probability of increasing frequency during this event was arbitrarily chosen to be 0.5.

Sports facility closures and startups

In the model, sports facilities can be closed and new sports facilities can be opened in the city. The number of fitness center and sports club facility startups in the province of Noord-Brabant were 14 and 8, respectively, in 2013.[10] As Eindhoven has approximately 10% of the population of Noord-Brabant, we crudely assumed that on average one fitness center and one sports club facility are opened every year. We also assumed that on average one fitness center and one sports club facility close every year. This is reasonable as the composition of sports facility in Eindhoven has fairly stable been stable in recent years.[10]

The time of closures and startups of a fitness center and sports club facility are drawn from an exponential distribution with a rate of 1.0 per year. After each startup or closure, the time of the next startup or closure is determined.

Closure

At the time of closure of a fitness center or sports club facility, the fitness center or sports club facility with the lowest number of members is closed. The location of the sports facility becomes vacant.

Startup

At the time of startup of a fitness center or sports club facility, a new fitness center or sports club facility is opened. The location of the new facility is determined by the demand for sports. We assume that sports facilities prefer establishment in neighborhoods where there is a large demand. The demand in neighborhoods with vacant facilities is assessed using Equation (7). The neighborhood with the highest demand is selected. The new fitness center or sports club is then randomly located at one of the vacant locations in that neighborhood.

$$Demand_{c,n} = \frac{SP(t)_{c,n}}{1 + Fac(t)_{c,n}} \quad (7)$$

With:

$SP(t)_{c,n}$ is the total number of individuals that engage in sports in category of sports c (i.e. fitness or sports club) in neighborhood n at time t.

$Fac(t)_{c,n}$ is the total number of existing sports facilities of category of sports c (i.e. fitness or sports club) in neighborhood n at time t.

Model calibration

The calibration process was performed using a grid search in which parameter sets of all unknown parameters were sampled from a uniform distribution. In total, eight parameters were calibrated (see Table S2). The model was run for 50 years to make sure it reached equilibrium. Model outcomes in equilibrium were matched to the observed overall sports participation and sports participation by age group, sex, income, and category of sports. The goodness-of-fit was assessed by maximizing the log-likelihood (LL) assuming a normal distribution. Optimal parameter values and the range of parameter uncertainty (95% confidence interval) were derived through a polynomial regression model. The model was calibrated under three assumptions of intention scores: (A) Gamma(1.0, 0.5), (B) Gamma(1.0, 1.0), and (C) Gamma(1.0, 3.0) (see also *"Attributes of individuals"*). The model assuming a shape parameter of 0.5 provided the best overall fit.

Population-level effects of interventions are the result of the average of 80 simulation runs. Uncertainty intervals (95%) reflecting parameter uncertainty were constructed by discarding the two highest and lowest outcome values of 80 simulation runs. Figure S4 compares the observed data with the model outcomes in equilibrium (i.e. after 50 years). Results in the manuscript were all based on the best fitted model, i.e. assuming a Gamma(1.0, 0.5) for variation of intention.

Table S2. Calibrated parameters.

Parameters	(A) Gamma(1.0, 0.5)[a]			(B) Gamma(1.0, 1.0)[a]			(C) Gamma(1.0, 3.0)[a]		
	Value	Low	High	Value	Low	High	Value	Low	High
Mean intention	153.5	146.2	160.9	50.3	48.5	52.1	20.9	20.4	21.5
a_groupS									
18-35yrs	1.0	-	-	1.0	-	-	1.0	-	-
35-54yrs	0.157	0.139	0.175	0.200	0.182	0.217	0.284	0.264	0.305
>55yrs	0.148	0.136	0.160	0.180	0.169	0.190	0.242	0.230	0.254
sexS									
Male	1.0	-	-	1.0			1.0	-	-
Female	0.659	0.608	0.710	0.808	0.754	0.862	0.773	0.732	0.813
incomeS									
High-income	1.0	-	-	1.0	-	-	1.0	-	-
Middle-income	0.471	0.428	0.514	0.587	0.545	0.629	0.640	0.605	0.675
Low-income	0.428	0.387	0.470	0.465	0.421	0.509	0.547	0.511	0.582
Distance decay (β_c)									
Fitness	0.029	0.028	0.031	0.022	0.020	0.025	0.013	0.011	0.015
Sports club	0.027	0.025	0.028	0.018	0.015	0.020	0.010	0.006	0.013
LL	35.3			20.0			-82.7		

[a] *Variation in individual intention follows a Gamma distribution with mean 1.0 and equal shape and rate parameter.*

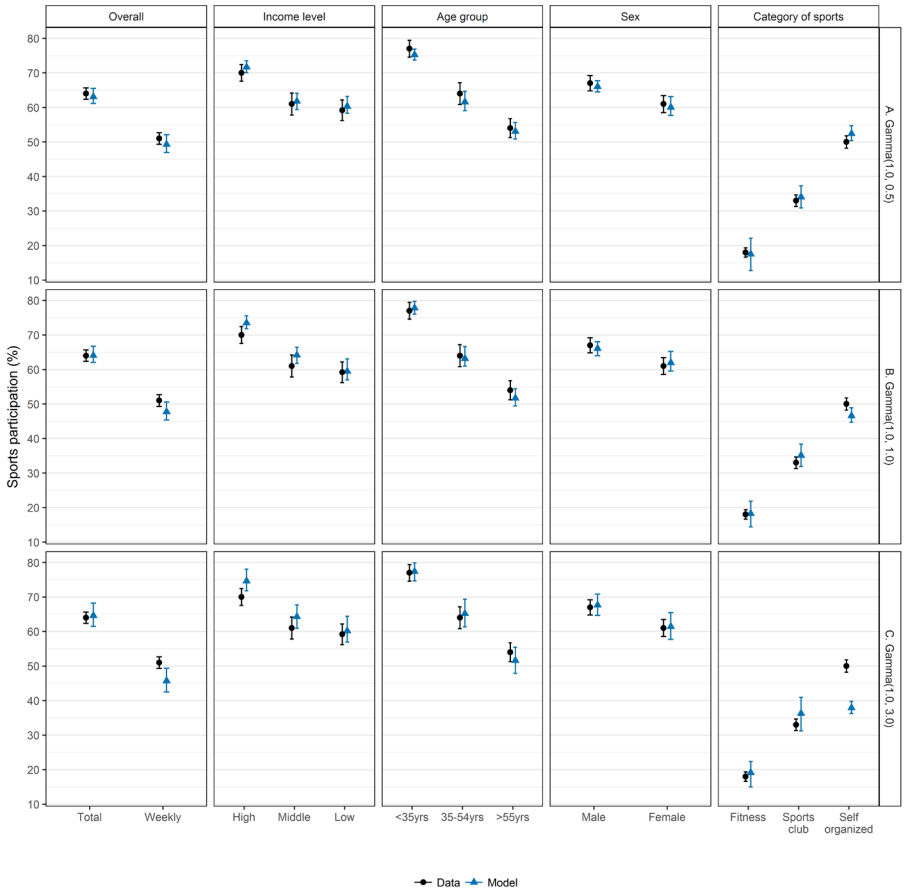


Figure S4. Comparison of modelled sports participation to the observed rates in Eindhoven in 2014. Model outcomes of overall sports participation, weekly sports participation, and overall sports participation by age group, sex, income level and category of sports (i.e. fitness, sports club, self-organized) are compared with data. Overall weekly sports participation included individuals who engage in (1) weekly sports in one of the categories of sports or (2) monthly sports at all three of the categories of sports (i.e. fitness, sports club, and self-organized). Model parameters were calibrated under three assumptions of intention between individuals: (A) Gamma(1.0, 0.5), (B) Gamma(1.0, 1.0), (C) Gamma(1.0, 3.0). Error bars represent the 95% uncertainty intervals reflecting parameter uncertainty and stochastic variation. Source data: Municipality of Eindhoven.

References

1. Statistics Netherlands. http://statline.cbs.nl/statweb/.
2. Statistics Eindhoven. http://eindhoven.buurtmonitor.nl.
3. Municipality of Eindhoven. Eindhoven sport. https://eindhovensport.nl/ (accessed January 2016).
4. Directorate-General for Education and Culture. Special Eurobarometer 412 'Sport and physical activity'. Brussels: European Commission, 2014.
5. Comensis. Fitness register. http://fitnessregister.nl/ (accessed January 2016).
6. Van Spronsen & Partners. Het fitnesscentrum in beeld: Van Spronsen & Partners, 2013.
7. van Lenthe FJ, Kamphuis CB, Beenackers MA, et al. Cohort Profile: Understanding socioeconomic inequalities in health and health behaviours: the GLOBE study. Int J Epidemiol 2014; 43(3): 721-30.
8. Simon HA. Rationality in Psychology and Economics. J Bus 1986; 59(4): S209-S24.
9. NOC*NSF. Sportersmonitor 2012: NOC*NSF, 2013.
10. Chamber of commerce. Brancheinformatie. www.ondernemersplein.nl/brancheinformatie.

7

GENERAL DISCUSSION

The overall aim of this thesis was to explore and quantify the importance of social networks as a determinant of health behaviors, and to investigate the usefulness of agent-based models (ABMs) as a tool for assessing the impact of interventions to reduce socioeconomic inequalities in health behaviors. This chapter starts with a critical appraisal of the main findings of this thesis, followed by experiences with developing ABMs, and potential future ABM research areas. It ends with an overview of the main conclusions and recommendations.

Critical appraisal of the main findings

Objective 1: To investigate to which extent the spread of unhealthy behaviors and infectious diseases share similarities and how infectious disease modeling could be applied for health behavioral research.

The suggestion that health behaviors, such as smoking and obesity, might be contagious already exists for some time.[1] However, empirical evidence to support this suggestion was limited. Christakis and Fowler were among the first to empirically demonstrate that health behaviors spread from person-to-person and may be contagious. Specifically, they showed that becoming obese or quitting smoking is more likely when social network members have become obese or quitted smoking.[2,3] The aim of *Chapter 2* was to find further support for this idea. Besides the contagiousness of health behaviors, we were interested to which extent the spread of unhealthy behaviors shows other traits similar with infectious diseases. Comparing conceptual frameworks of health behaviors with infectious disease, we found six additional similarities: (1) the presence of super-spreaders, (2) heterogeneity in susceptibility, (3) heterogeneity in infectiousness, (4) clustering of infected individuals, (5) the influence of the physical environment, and (6) vaccination. More specifically, super-spreaders in infectious diseases greatly resemble opinion leaders or role models in the spread of unhealthy behaviors. Heterogeneity in susceptibility shows remarkable similarities with variation between individuals in adopting unhealthy behavior, while heterogeneity in infectiousness is similar to variation in strength of social influence due to for example the type of contact (e.g. close vs weak contacts). Both infectious diseases and health behaviors can cluster in social networks. Also, the physical environment may facilitate or inhibit acquiring an infection or adopting a behavior.

Finally, we argued that vaccination, which reduces the risk of infectious disease, resembles social inoculation, whereby persons develop resistance to persuasive cues or messages about unhealthy behaviors.[4]

An important implication of these similarities is that modeling techniques to study infectious disease epidemiology and control could be exploited in health behavioral research. In public health, the field of infectious diseases was arguably the first to adopt individual-based modeling or agent-based modeling. Scholars in infectious disease epidemiology have long understood that the course of disease transmission in a population results from complex interactions between biology, social contacts, and environments. Their history of developing and applying ABMs goes back decades and resulted in many scientific papers.[5,6] Moreover, these models have been used to inform policy-makers.[7] Existing infectious disease modeling concepts could be very useful for model developments in health behavioral research, in particular with respect to social networks and social contagion. Relevant infectious disease models include, among others, models for influenza,[8-10] sexual transmitted infections,[11-14] and leprosy.[5,15] Since these models include social interactions, elements can be translated or used as starting point for health behavioral models to avoid redundant preparations.

An illustrative example from infectious diseases epidemiology is the transmission of leprosy, which has different risks for household and other contacts. In the individual-based model *SIMCOLEP*, the formation, dissolution and changes of households has been explicitly modeled as stochastic processes.[5,15] This was included because the risk to acquire the infection from household contacts is higher compared to people in the general population. Such an approach might be useful for health behavioral research as well, because close contacts are more influential than contacts in the general population.[2,3]

Moreover, infectious disease modeling provides useful concepts, such as the basic reproduction number (R_0), i.e. the average number of successful transmissions per infectious person in a fully susceptible population. An increase in the prevalence of smoking only occurs if $R_0>1$, which means that each smoker on average triggers at least one other individual to start smoking. The intervention goal could be to reduce R_0 to below 1, to stop further spreading of smoking. Another useful

concept is herd immunity, which occurs if enough people (i.e. more than $1-1/R_0$) in a population have become immune to an infection. Herd immunity provides protection to people who are not immune and prevents further spread indirectly. If enough people do not engage in unhealthy behaviors, this may in a similar way prevent others to start unhealthy behaviors. Models can be useful to explore whether this would also be true for unhealthy behaviors and to determine how many cases of unhealthy behaviors could have been prevented as a result of interventions.

Although both fields share similarities, there are also some fundamental differences between infectious diseases and unhealthy behaviors, which were not highlighted so much in *Chapter 2*. Firstly and probably the most important difference is that infectious diseases have clear etiologies and causes. Health behaviors have complex interacting causes, which are hard to pin down. For example, obesity is the result of a reciprocal relationship with physical activity and dietary behaviors, further determined by genetic and environmental influences, such as access to food and food types and social networks.[16] Secondly, although unhealthy behaviors may spread from person-to-person, this does not mean that they are transmitted as an infectious disease. An important difference, of course, is that health behaviors lack an infectious agent. Thirdly, for most infectious diseases, the pattern of transmission between individuals is "one-to-one". Thus, one susceptible individual acquires an infection from one infectious individual. In contrast, health behaviors are in most cases caused by the collective influence of many individuals. Besides the influence between individuals, other factors such as social norms may also play an important role.

Vaccination as a known means in the control of infectious diseases is obviously not exactly the same as social inoculation in health behavioral research, although we argued that the process does show similarities (i.e. both reduce risk for individuals and the population as a whole). Vaccination exerts a purely biological response, while social inoculation only exerts a psychological response.[4] Moreover, they also differ in the route of administration. Vaccination is administered by injection or orally.[17] Social inoculation is usually delivered as program aimed to prevent the behavior.[4] Although the means might be different from infectious diseases, their strategy and purpose are not.

In conclusion, the spread of unhealthy behaviors shows similarities as well as some fundamental differences with the spread of infectious diseases. Yet, the conceptual similarities are sufficiently strong for health behavioral scientists to consider exploiting infectious disease modeling approaches, especially when these concern social interactions. Close collaboration with infectious disease modelers is therefore highly recommended, to benefit from their decades of experience.

Objective 2: To quantify the associations between social networks and smoking, sports participation and overweight, and whether these associations vary by type of social network tie.

Chapters 3 and *4* show that health behaviors of social networks members are associated with the adoption of both healthy as well as unhealthy behaviors. In *Chapter 3*, we showed that living in neighborhoods with a high prevalence of non-smoking, no sports participation and overweight increased the odds of smoking cessation, quitting sports and becoming overweight. Similar but weaker associations were found for initiating smoking, starting sports, and losing weight. It implies that a culture of, for example, overweight in neighborhoods may contribute to becoming overweight. A recent study showed similar associations between the prevalence of smoking in neighborhoods and changes in smoking behaviors of neighborhood residents.[18]

In *Chapter 4*, we assessed the role of smoking in social networks on smoking cessation and smoking relapse in a Dutch adult population. Smokers with the largest proportion of smokers among social network members were less likely to quit and more likely to experience a relapse. In addition, the type of social tie appeared to be important: smoking of household members and friends were strongly associated with smoking cessation and relapse, whereas the smoking behavior of family members outside the household was not. Furthermore, smoking cessation among moderate-to-heavy smokers was mainly associated with smoking among friends. This supports earlier findings showing that behaviors of socially close contacts are most important for changes in health behaviors.[2,3]

In both chapters, we found stronger associations of the influence of social networks

with unhealthy behaviors than healthy behaviors, suggesting that the influence of social networks is more important for unhealthy behaviors. The prevalence of no sports participation and overweight in neighborhoods showed a stronger relationship with quitting sports and becoming overweight than the prevalence of sports participation and normal weight with starting sports and losing weight (see *Chapter 3*). Similarly, the associations between smoking in social networks and smoking relapse was stronger compared to the associations between social networks and smoking cessation (see *Chapter 4*). This underscores that it may be easier to adopt unhealthy behaviors than healthy behavior.[19,20]

Although our findings support the thought that health behaviors are influenced by behavior of others, the role of social networks can also be explained by two other pathways. First, individuals could choose their friends based on smoking or weight-related behaviors or factors associated with these behaviors, a phenomenon called homophily or selection.[21] For example, sensation seekers may tend to connect with each other and are more likely to start smoking.[22] This could also play a role in the study of *Chapter 4*. Second, members of a social network might be exposed to a common environmental factor, which causes them to jointly change behaviors.[23,24] Although the study of *Chapter 3* corrects for neighborhood deprivation as a contextual confounder, other contextual determinants, such as neighborhood accessibility to facilities, could also partially explain the relation. In *Chapter 4*, we did not account for any contextual effects. Also here, people may be exposed to, for example, neighborhood deprivation, which may to some extent cause them to quit or relapse smoking.[18] Failing to fully adjust for homophily and contextual effects remains a problem in social network studies in general, especially if we want to move from mere associations to causality.

Studies involving social networks generally tend to simplify interpersonal relationships. The studies presented in *Chapters 3* and *4* also reflect a simplified social environment/network, i.e. neighborhood contacts or close social ties, such as household members, close friends, and family members. It is likely that close social ties are more likely to influence health behaviors than weaker ties (i.e. neighborhood contacts). However, it is also known that weaker ties can serve as bridges between groups to spread health behaviors.[25] In this thesis, we were not able to look into weak and close ties simultaneously, nor how weak ties

might affect the relationship of close ties and changes in health behaviors. A related challenge is how to measure close ties. In *Chapter 4*, we did not find any association between smoking among close family members outside the household and smoking cessation and relapse. This seems contradictory to earlier findings, which suggest that socially close contacts are most important to explain smoking cessation.[3] The question that arises is whether these socially close ties are truly strong and how to measure tie strength optimally for future research. One may elucidate tie strength by adding questions about frequency of contact and time spend with a contact. Alternatively, it has been argued that people named first in questionnaires are generally closer than those named later, which could also help to get better insight into tie strength.[26]

In *Chapter 4*, social network data were collected using an egocentric approach, where a respondent's perception of behaviors of others was used. On the one hand, this approach seems less reliable because respondents are generally not very accurate in their estimates of behaviors of others. On the other hand, however, people are more likely to be influenced by their perception of behaviors in a social network than actual behaviors. This discrepancy may be associated with other factors, such as the behavior and the boundary of the social network that is being studied, social distance and tie strength. Also, egocentric data do not allow linking respondents or contacts of a respondent to others in the data set. Data on links may provide essential information about the tie strengths: mutual ties are stronger than unidirectional ties. Alternative data collection approaches that could be considered are sequenced data, where respondents nominate others to participate in the study, or census data, which include (all) members of a community.[27] These approaches have the advantage of being able to link people in a dataset and to utilize other sociometric network measures.

In conclusion, our findings indicate that healthy and unhealthy behaviors in personal social networks and neighborhoods are important for changes in health behaviors. They also support the idea that both healthy and unhealthy behaviors are contagious, as discussed in *Chapter 2*. Future research assessing the relevance of personal networks and the broader context of networks (e.g. weaker ties) simultaneously can help us understand how they might interact with each other. Based on our findings, smoking interventions targeting household

members or groups of friends may be most promising to establish a significant further reduction of the smoking prevalence. In addition, as health behaviors might spread to others, the effect of various interventions might be more effective than initially thought. However, further research on this is needed.

Objective 3: To develop two agent-based models (ABMs) to explore the potential impact of interventions aimed at reducing socioeconomic inequalities in food consumption and sports participation.

Chapters 5 and *6* present two ABMs: a model for income inequalities in food consumption (hereafter: food model) and a model for income inequalities in sports participation (hereafter: sports model). Both models were intended as proof of concepts to assess their potential usefulness in health behavioral research. Our food model captured the dynamics between households and fruit and vegetable stores, supermarkets, discount supermarkets and fast-food outlets (see *Chapter 5*). Food shopping behavior of households was determined by the distance, price of a food outlet, preference for healthy food, in the context of a dynamic environment in which food outlets could close and start. *Chapter 6* presents our sports model, which incorporated the interactions between individuals and the social and physical environment. Sports participation was determined by characteristics of individuals (e.g. intention to start sports), characteristics of sports facilities (price and distance), neighborhood safety, social cohesion and the social influence of neighbors. Both studies illustrate that agent-based modeling can be used to encompass relevant individual and environmental factors of income inequalities of food consumption and sports participation, and to account for complex system properties including emergence, feedback and adaptation.

Apart from accounting for interaction and feedbacks, we illustrated the usefulness of the models to explore the impact of various interventions ranging from environmental ('upstream') to individual ('downstream') interventions.[28-30] This also included interventions for which randomization in real-life is impossible: e.g. eliminating residential segregation in *Chapter 5*. The food model was used to assess three interventions: (1) eliminating residential segregation, (2) lowering prices of healthy food, and (3) health education. Explorative results showed that eliminating residential segregation yields the highest impact on reducing absolute

income inequalities in healthy food consumption (4.4% reduction in the long run), but is partly the result of an unfavorable change in healthy food consumption among higher income groups. Lowering prices and providing health education to low-income households could also significantly reduce income inequalities in food consumption providing sustained effort.

The sports model was used to test the impact of: (1) providing health education, (2) lowering prices of sports facilities, (3) increasing the availability of sports facilities, (4) improving neighborhood safety, and (5) combining all previous interventions simultaneously (i.e. multilevel intervention). Our model predicts that providing all interventions can increase sports participation and modestly reduce absolute income inequalities in sports participation over time. The largest gain can be attained through health education, if its effect and reach are sufficient large. Combining interventions in a multilevel approach showed the largest impact on sports participation and income inequalities in sports participation: 3.8%-point increase and 1.2%-point reduction, respectively. These explorative findings suggest that much work still needs to be done to increase sports participation at a population-level. If the aim is to decrease income inequalities in sports participation, a shift to tailored interventions might be needed.

Our findings support the hypothesis that on a short term the population-level effect of an intervention may not represent the full potential impact of the intervention.[31,32] In *Chapters 5* and *6*, the impact of interventions increases over time. The full potential effects of these interventions are only approximated after five or ten years. This may imply that studies investigating the population-level effects of an intervention should at least have a follow-up of five years. This, however, is most likely not feasible in practice. As in most cases direct evidence for the long-term population-level effect of an intervention is unknown, ABMs can be useful to explore such population-level effects of interventions studies.

As our models were intended as proof-of-concept, they remain a simple abstract representation of what drives income inequalities in food consumption and sports participation. Our models have much room for improvements.

Firstly, it should be noted that our current models do not support the person-to-

person influence of health behaviors in social networks as suggested in *Chapter 2* yet. Although in our sports model we made a first step by allowing sports participation to be influenced by direct neighbors, it still lacks the influence from (personal) social networks on behaviors. A future version of our models should include dynamics of relevant social networks and model the social contagion of behaviors. This would also have implications for the impact of interventions, as interventions might also spread through social networks.

Secondly, the scope of our models was relatively narrow. For example, our sports model did not account for the influence of the presence of parks and neighborhoods aesthetics on (self-organized) sports participation. Also, in both models, possible relevant individual-level characteristics were excluded such as educational level or ethnicity.[33]

Thirdly, model predictions are subjected to model assumptions. Here, some extreme assumptions may limit the practical implication of results. For example, in both studies, we assumed that interventions would have an immediate effect. However, eliminating residential segregation or improving safety is a slow process in real-life, something the models did not yet take into account. Also, the population-level effects of the interventions studied have not been validated to data, because of a lack of data. Furthermore, model predictions do not account for future changes in the population in terms of age and income distribution over time. It is therefore important that models are regularly updated using the latest information. Also, the evolving and improving literature in these areas may require updating or even restructuring the model and its assumptions.

Fourthly, the lack of readily available data and data granularity remain an issue for model parametrization and validation. For example, since no data were available to link food outlet visiting behavior to actual food intake, we chose the type of food outlet to be a proxy for healthy or unhealthy food consumption. Hence, our food model is actually more a model of food shopping than food consumption. Generally, it is known that developing models unveils gaps in knowledge and the need for new data, which until now were perhaps not considered. Model development and data collection should therefore preferably go hand in hand. Other examples of relevant missing data in our studies include, among others,

data on preferences for healthy and unhealthy food; price preferences and willingness to travel (distance) for food consumption or sport participation in relation to age, sex and income; data on movement and to what extent this relates to availability of food outlets and/or sports facilities. Discrete choice experiments could be useful to elucidate preferences and decision rules.[34] The need for new data does not only improve models, but could potentially also bring the field of health behavioral research forward.[28]

In conclusion, our findings show that agent-based modeling holds great promise to study health behaviors in the context of a complex system in which individuals interact with each other and the environment. Our models should be considered as proof-of-concept, highlighting the level of complexity and specification required to develop a model, as well as the gaps in available data to inform the model. The models also illustrate how interventions might impact health behaviors while accounting for interactions between individuals and their environment. However, these findings should not be interpreted too literally, but are for now only mainly illustrative. Our ABMs are still very much in their infancy and need further refinement in order to be useful to understand income inequalities in health behaviors, and how these may be reduced effectively.

Experiences with ABM development

The two ABMs presented in this thesis are among the first developed for the field of health promotion. These models were both developed from scratch, and decisions had to be made about the model design including agents and environment, model implementation, data and model evaluation. A general rule in agent-based modeling is that the model should be simple enough to provide useful insight and at the same time complex enough to not misrepresent the real world. Decisions to ensure this do not always follow science and may be in certain cases argued as arbitrary. *Bonabeau (2002)* calls this process more an art than a science.[35] Below, we discuss some of our key choices regarding agent-based model design.

Agents and environment

The first step is to define agents. Individuals are the obvious type of agents to

represent a population, which was also the case in our sports model (see *Chapter 6*). However, in some cases another entity might fit the desired outcome better. For example, in the food model households were chosen as agents, because food decisions are typically made at a household level (see *Chapter 5*). The second step is to determine the number of agents to be modeled. In both of our models, we chose to model the actual number of agents as derived from data. This approach may not be very efficient, because large numbers of agents increase the computation time of the model. In general, there is little guidance on determining the number of agents in the model. At least, the model should be sufficiently large to reflect heterogeneity and interaction between agents. Scaling down the population size by, say, a factor of ten would not work in our models, because we modeled the population per neighborhood, with several of them only harboring 200-300 residents. In retrospect, we could have focused on the highly populated neighborhoods only to make our model more efficient. Generally, for most purposes, large numbers of agents are unnecessary.[36]

Besides individuals and households, food outlets and sports facilities were also modeled as agents. This allowed for interactions of outlets and facilities with individuals or households at agent-level, i.e. closure and startups depending on the number of customers or members. This, however, required additional data to model processes of food outlets and sports facilities. An alternative approach would have been to model food outlets or sports facilities as an attribute of neighborhoods, which indicates the number of healthy or unhealthy food outlets or sport facilities. This number can change as a result of changes in the demand for food or sports in a neighborhood. This approach would maintain the influence of available of food outlets and sports facilities on food and sports decisions, but at the cost of losing heterogeneity in outlets/facilities (e.g. location), accuracy (e.g. distance) and losing the interactions between individuals/households and outlets/facilities at agent level. Such an approach is therefore only preferred in public health problems in which exact locations and distances do not play an essential role in determining the outcome.

The physical environment represented a particular city (Eindhoven) with all its neighborhoods. It was modeled using a realistic grid with relatively small cell sizes (i.e. 10m x 10m). The advantage of a fine realistic grid is that it represents

distances accurately and is more intuitive for users or policy-makers to understand. However, such a grid would only be advised in models where actual distances are crucial, for example, walking behaviors along actual routes in neighborhoods. In other cases, a crude or even hypothetical grid might already be sufficient to reflect all complexities of relevance. This would also make models more efficient in terms of computation power. In retrospect, a crude (hypothetical) grid would have been a more efficient choice for our purpose.

Model implementation

Models should as much as possible be embedded in existing theories and available knowledge. The selection of relevant individual and environmental factors of food consumption and sports participation was based on: (1) their importance to explain health behaviors after reviewing literature, and (2) their relevance for potential intervention targets. Obviously, not all relevant factors might have been included, making the scope of the model purposely narrow. The paradox is that if a model has a large number of factors (i.e. more complicated), it will be just as difficult to interpret and validate.[37]

In both of our ABM models we applied a social ecological perspective, meaning that both individual and environmental factors interact with each other, and from which income inequalities in food consumption or sports participation emerge. However, the implementation of this perspective was model specific. In the food model, households' food consumption decisions were modeled following the multi attribute utility theory, which originated from economic theory. The reason for this approach was that food shopping is to a large extent guided by economic considerations. In the sports model, individuals' decisions were to a large extent based on the principles of the Theory of Planned Behavior with intention as the main driver of sports participation. This illustrates that models can be developed with a focus on different aspects of the phenomenon, which would make the model only relevant in the context of its aim.

Data and model evaluation

To assure some realism in our models, we based parameter values and decision rules on empirical data where possible. Since finding detailed data to quantify

the models was a challenge, we have used a variety of data sources including Statistics Netherlands, Statistics of the Municipality of Eindhoven, existing cohort studies (GLOBE) and scientific literature and reports. Using data from various sources is common practice to overcome the challenge to locate data with sufficient detail.[38] However, as described earlier, data to quantify preferences and to inform decision rules are often lacking. For that reason, the development of models should preferably go hand in hand with new data collection.

ABMs as any kind of mathematical models require thorough evaluation, which includes verification, calibration and validation. Verification determines whether the model behaves according to its intention, which is primarily tested during the programming phase. The next step is calibration, which is the process of estimating values of parameters that could not be quantified directly using data such that the system replicates the real-world data. The calibration of the food model was performed through an iterative process until model outcomes matched the data (see *Chapter 5* for more details). The optimal values of the calibrated parameters were used to predict the impact of interventions. These results, however, do not reflect uncertainty in calibrated parameter values. This can be regarded as a limitation of this study. The calibration of the sport model was performed using a grid search. Optimal parameter values and the range of parameter uncertainty were derived through a polynomial regression meta-model. Final model outcomes and predictions of interventions included an uncertainty interval (see *Chapter 6* for more details). Such an approach is highly recommended in ABM research, as it may also provide insight into the variability of model outcomes and parameter estimates.

The last step is validation that assesses to which extent predicted results matches reality. It would be good to compare the projected trend and predicted impact of interventions from our models to independent data, whenever these become available in the future. Validation of models would increase reliability of modeled results and is crucial if we want these models to become an aid for policy.

In order to trust the results, transparency and making assumptions explicit is an important part of the development. Here, each model was accompanied by an appendix which provides considerable transparency about model design,

structure, input data, and calibration (see Supplementary Appendix of *Chapters 5* and *6*). Transparency does not only enable the scientific community to evaluate model face validity, but also to assure reproducibility. Recently, a number of protocols have been written to guide future ABM research.[39-41] Such protocols typically comprise model overview, design concepts, and other details (e.g. input data and calibration methods).

The future of agent-based modeling in health behavioral research

As researchers recognize the need of incorporating the system from which health behaviors emerge, there has been growing interest in the application of agent-based modeling.[39,42] In the past years, a number of studies have been published, focusing on smoking,[43] dietary behaviors,[44-49] walking behaviors,[50-57] and obesity.[56,58-60] Most of these studies developed a model to assess the impact of various scenarios, similar to the ABMs presented in this thesis. Of all available studies only three account for health disparities: one on income inequalities in dietary behaviors,[44] and two on racial disparities in food consumption and obesity.[45,56]

ABM in public health, social epidemiology and health promotion research is only in its initial stages. Further development and updates of the models are necessary to keep up with the evolving knowledge about interactions and feedbacks relevant to health behaviors. As mentioned, ABM is hardly applied in studies on socioeconomic inequalities in health. It has, however, been argued that ABM research in this area would be especially relevant, because it may provide better understanding about why health disparities are persistent and robust over time, and how interventions might impact health disparities over time, as good counterfactuals are lacking.[28] In order to answer such questions, models should be able to adequately replicate processes from which inequalities arise. This remains a challenge for future ABM research.

Existing ABM research focused either on the physical environment, the social environment or both. The modeled physical environment varied widely in level of detail, ranging from hypothetical grids to realistic fine-grained grids.[44-48,50-57] The latter was primarily used for ABM studies of walking behaviors. In all models

the environment was active with inherent characteristics and rules. The social environment in ABM research was often limited or simplified.[45,46,56,58-62] Generally, changes in social networks were modeled using fixed annual probabilities and based on simple mechanisms, such as matching the average behavior.[60] All published studies have implemented one of the established mathematical network models: (1) random, (2) lattice, (3) small-world, (4) spatial, and (5) scale-free networks.[63-67] The random network is the most popular.[45,46,56,59,68] Three studies based their network on empirical network data.[60,62,69] A limitation of all established network models is that they do not (fully) account for the possibility of people mixing with similar people (i.e. homophily), as well as possible changes in network structure and ties over time.

An alternative way of incorporating social networks in ABMs is to explicitly model the processes. This means that the formation, changes and dissolution of social networks are (fully) described by behavioral rules. Processes such as homophily and network properties (e.g. cluster and degree coefficients) emerge as a result of these rules. To quantify these rules and validate the network dynamics, longitudinal data on social networks are needed. This approach is widely used in ecological systems and in infectious disease modeling,[6,70] but has to our knowledge not yet been applied in health behavioral research.

Modeling the influence of social networks on health behaviors also remains a challenge. This should at minimum encompass the variation between individuals in adopting health behaviors and the variation in the strength of social influence (see *Chapter 2*). For example, close friends may be more influential than family members (see *Chapter 4*). To quantify social contagion of health behaviors, longitudinal data on both social networks and health behaviors are required. Incorporating social contagion in models is relevant for designing and testing network interventions.[71] Changing behaviors of key players in a network might be sufficient to change behaviors in others. The success of network interventions at population level, however, depends on concepts, such as tipping points (i.e. number of people needed that change behavior in order for others to follow) and feedbacks (i.e. changes in one person influences another person which in turn influence others).[27,28] Such concepts have hardly entered the field of behavioral research, because they require the use of even more complex models than those

presented in this thesis.[28] Designing policy relevant models that realistically incorporate social networks and their influence on health behaviors remain a challenge for future ABM research in health behaviors.

To further evolve agent-based modeling in public health, there is a need to reshape current public health research, policy and practice, which is still skewed towards traditional methods. This will require building capacity, restructuring research funds, and increasing acceptance by the wider scientific community and stakeholders.[42] The involvement of stakeholders at a very early stage of model development is important: not only to formulate policy-relevant targets or research questions, but also to think about relevant processes, knowledge gaps, and possible collection of new data. With the growing scientific literature and increasing number of complex systems workshops, complex systems approaches are slowly gaining ground within the current paradigm.

Conclusions and recommendations

Conclusions

- The spread of unhealthy behaviors shows similarities with the spread of infectious diseases, despite some fundamental differences.
- The local presence of health behaviors and overweight is associated with subsequent changes in health behaviors and overweight. This suggests that norms or culture in a neighborhood in favor of, for example, overweight may contribute to becoming overweight.
- Smokers and former-smokers with a high proportion of smokers in their social networks are less likely to quit smoking and more likely to relapse smoking, respectively. Smoking by household members and friends is the most important in this respect.
- ABM has much potential to study health behaviors in the context of a system and to analyze the population-level impact of various interventions aimed at promoting healthy behaviors and reducing health disparities.
- ABM in health behavioral research is still in its infancy. There is much room for refinements, especially concerning the role of the social environment.

Recommendations

- ABM research on the spread of unhealthy behaviors should exploit relevant concepts of infectious disease modeling, in particular concerning social interactions.
- Research assessing networks of family, friends and neighbors simultaneously can help us to better understand which contacts are most important for changes in personal health behavior.
- Smoking interventions that also consider household members and close friends may be most promising to establish a significant further reduction of the smoking prevalence.
- ABM in health behavioral research should go hand in hand in hand with data collection and should preferably be carried out in cooperation with stakeholders, including policy-makers, at an early stage of the project.

References

1. Glanz K, Bishop DB. The Role of Behavioral Science Theory in Development and Implementation of Public Health Interventions. Annual Review of Public Health, Vol 31 2010; 31: 399-418.
2. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007; 357(4): 370-9.
3. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008; 358(21): 2249-58.
4. Evans RI, Raines BE, Hanselka L. Developing data-based communications in social psychological research: adolescent smoking prevention. J Appl Soc Psych 1984; 14: 289-95.
5. Blok DJ, de Vlas SJ, Fischer EA, Richardus JH. Mathematical modelling of leprosy and its control. Adv Parasitol 2015; 87: 33-51.
6. Vynnycky E, White RG. An introduction to infectious disease modelling. New York: Oxford University Press; 2010.
7. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. Annu Rev Public Health 2012; 33: 357-76.
8. Longini IM, Halloran ME, Nizam A, Yang Y. Containing pandemic influenza with antiviral agents. American Journal of Epidemiology 2004; 159(7): 623-33.
9. Longini IM, Nizam A, Xu SF, et al. Containing pandemic influenza at the source. Science 2005; 309(5737): 1083-7.
10. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature 2006; 442(7101): 448-52.
11. Korenromp EL, Van Vliet C, Bakker R, De Vlas SJ, Habbema JDF. HIV spread and partnership reduction for different patterns of sexual behaviour - a study with the microsimulation model STDSIM. Math Pop Studies 2000; 8: 135-73.
12. Hontelez JA, Lurie MN, Barnighausen T, et al. Elimination of HIV in South Africa through expanded access to antiretroviral therapy: a model comparison study. PLoS Med 2013; 10(10): e1001534.

13. Kretzschmar M, vanDuynhoven YTHP, Severijnen AJ. Modeling prevention strategies for gonorrhea and Chlamydia using stochastic network simulations. American Journal of Epidemiology 1996; 144(3): 306-17.
14. Ghani AC, Garnett GP. Risks of acquiring and transmitting sexually transmitted diseases in sexual partner networks. Sex Transm Dis 2000; 27(10): 579-87.
15. Fischer E, De Vlas S, Meima A, Habbema D, Richardus J. Different mechanisms for heterogeneity in leprosy susceptibility can explain disease clustering within households. PLoS One 2010; 5(11): e14061.
16. Kadouh HC, Acosta A. Current paradigms in the etiology of obesity. Tech Gastrointest Endosc 2016.
17. World Health Organization. Vaccines. http://www.who.int/topics/vaccines/en/ (accessed December 2016).
18. Ivory VC, Blakely T, Richardson K, Thomson G, Carter K. Do changes in neighborhood and household levels of smoking and deprivation result in changes in individual smoking behavior? A large-scale longitudinal study of New Zealand adults. Am J Epidemiol 2015; 182(5): 431-40.
19. Hawkins J, Hollingworth W, Campbell R. Long-term smoking relapse: a study using the British household panel survey. Nic & Tob Res 2010; 12: 1228-35.
20. Piasecki TM. Relapse to smoking. Clin Psych Rev 2006; 26: 196-215.
21. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Soc 2001; 27: 415-44.
22. Banerjee SC, Greene K, Bagdasarov Z, Campo S. 'My friends love to tan': examining sensation seeking and the mediating role of association with friends who use tanning beds on tanning bed use intentions. Health Educ Res 2009; 24(6): 989-98.
23. Chaix B, Leal C, Evans D. Neighborhood-level confounding in epidemiologic studies: unavoidable challenges, uncertain solutions. Epidemiology 2010; 21(1): 124-7.
24. Cohen-Cole E, Fletcher JM. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. J Health Econ 2008; 27(5): 1382-7.
25. Granovetter M. The strength of weak ties. Am J Sociol 1973; 78: 1360-80.

26. Valente TW, Vlahov D. Selective risk taking among needle exchange participants: Implications for supplemental interventions. American Journal of Public Health 2001; 91(3): 406-11.
27. Valente TW. Social networks and health. New York: Oxford University Press, Inc.; 2010.
28. Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health 2011; 101(9): 1627-34.
29. Galea S, Hall C, Kaplan GA. Social epidemiology and complex system dynamic modelling as applied to health behaviour and drug use research. Int J Drug Policy 2009; 20(3): 209-16.
30. Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol 2010; 39(1): 97-106.
31. Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? Ann Behav Med 2004; 27(1): 3-12.
32. Glass TA, McAtee MJ. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. Soc Sci Med 2006; 62(7): 1650-71.
33. Brug J, van Stralen MM, Chinapaw MJ, et al. Differences in weight status and energy-balance related behaviours according to ethnic background among adolescents in seven countries in Europe: the ENERGY-project. Pediatr Obes 2012; 7(5): 399-411.
34. Holm S, Lemm R, Thees O, Hilty LM. Enhancing agent-based models with discrete choice experiments JASSS 2016; 19(3).
35. Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 2002; 99 Suppl 3: 7280-7.
36. Abdou M, Hamill L, Gilbert N. Designing and Building an Agent-Based Model. In: Heppenstall AJ, Crooks AT, See LM, Batty M, eds. Agent-Based Models of Geographical Systems: Springer; 2012: 141-65.
37. Evans D. Uncertainty and error. In: Heppenstall AJ, Crooks AT, See LM, Batty GD, eds. Agent-based model of geographical systems. London: Springer; 2012.
38. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness

of dynamic-agent models in understanding place effects on health. Am J Epidemiol 2008; 168(1): 1-8.
39. Nianogo RA, Arah OA. Agent-based modeling of noncommunicable diseases: a systematic review. Am J Public Health 2015; 105(3): e20-31.
40. Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. The ODD protocol: A review and first update. Ecol Modell 2010; 221(23): 2760-8.
41. Grimm V, Berger U, Bastiansen F, et al. A standard protocol for describing individual-based and agent-based models. Ecol Modell 2006; 198(1-2): 115-26.
42. Rutter H, Savona N, Glonti K, et al. The need for a complex systems model of evidence for public health. Lancet 2017.
43. Schaefer DR, Adams J, Haas SA. Social networks and smoking: exploring the effects of peer influence and smoker popularity through simulations. Health Educ Behav 2013; 40(1 Suppl): 24S-32S.
44. Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential segregation. Am J Prev Med 2011; 40(3): 303-11.
45. Orr MG, Galea S, Riddle M, Kaplan GA. Reducing racial disparities in obesity: simulating the effects of improved education and social network influence on diet behavior. Ann Epidemiol 2014; 24(8): 563-9.
46. Zhang D, Giabbanelli PJ, Arah OA, Zimmerman FJ. Impact of different policies on unhealthy dietary behaviors in an urban adult population: an agent-based simulation model. Am J Public Health 2014; 104(7): 1217-22.
47. Widener MJ, Metcalf SS, Bar-Yam Y. Agent-based modeling of policies to improve urban food access for low-income populations. Appl Geogr 2013; 40: 1-10.
48. Chen HJ, Xue H, Kumanyika S, Wang Y. School beverage environment and children's energy expenditure associated with physical education class: an agent-based model simulation. Pediatric Obesity 2016.
49. Trogdon JG, Allaire BT. The effect of friend selection on social influences in obesity. Economics & Human Biology 2014; 15: 153-64.
50. Jin X, White RA. An agent-based model of the influence of neighbourhood design on daily trip patterns. Comput Environ Urban Syst 2012; 36(5): 398-411.

51. Yang Y, Diez Roux AV. Using an agent-based model to simulate children's active travel to school. Int J Behav Nutr Phys Act 2013; 10(1): 67.
52. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. A spatial agent-based model for the simulation of adults' daily walking within a city. Am J Prev Med 2011; 40(3): 353-61.
53. Yang Y, Diez Roux AV, Auchincloss AH, Rodriguez DA, Brown DG. Exploring walking differences by socioeconomic status using a spatial agent-based model. Health Place 2012; 18(1): 96-9.
54. Yang Y, Diez-Roux A, Evenson KR, Colabianchi N. Examining the impact of the walking school bus with an agent-based model. Am J Public Health 2014; 104(7): 1196-203.
55. Yin L. Assessing Walkability in the City of Buffalo: Application of Agent-Based Simulation. J Urban Plan D 2013; 139(3): 166-75.
56. Orr MG, Kaplan GA, Galea S. Neighbourhood food, physical activity, and educational environments and black/white disparities in obesity: a complex systems simulation analysis. J Epidemiol Community Health 2016; 70(9): 862-7.
57. Badland H, White M, Macaulay G, et al. Using simple agent-based modeling to inform and enhance neighborhood walkability. Int J Health Geogr 2013; 12: 58.
58. Hammond RA, Ornstein JT. A model of social influence on body mass index. Ann N Y Acad Sci 2014; 1331: 34-42.
59. El-Sayed AM, Seemann L, Scarborough P, Galea S. Are network-based interventions a useful antiobesity strategy? An application of simulation models for causal inference in epidemiology. Am J Epidemiol 2013; 178(2): 287-95.
60. Zhang J, Tong L, Lamberson PJ, Durazo-Arvizu RA, Luke A, Shoham DA. Leveraging social influence to address overweight and obesity using agent-based models: the role of adolescent social networks. Soc Sci Med 2015; 125: 203-13.
61. Trogdon JG, Allaire BT. The effect of friend selection on social influences in obesity. Econ Hum Biol 2014; 15: 153-64.
62. Zhang J, Shoham DA, Tesdahl E, Gesell SB. Network interventions on physical activity in an afterschool program: an agent-based social network

study. Am J Public Health 2015; 105 Suppl 2: S236-43.
63. Bollobás B. Random graphs. London: Academic Press; 1985.
64. Harris TE. Contact interactions on a lattice. Ann Probab 1974; 2: 969-88.
65. Keeling MJ, Eames KT. Networks and epidemic models. J R Soc Interface 2005; 2(4): 295-307.
66. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998; 393(6684): 440-2.
67. Barabasi AL, Albert R. Emergence of scaling in random networks. Science 1999; 286(5439): 509-12.
68. Bahr DB, Browning RC, Wyatt HR, Hill JO. Exploiting social networks to mitigate the obesity epidemic. Obesity (Silver Spring) 2009; 17(4): 723-8.
69. Wang Y, Xue H, Chen HJ, Igusa T. Examining social norm impacts on obesity and eating behaviors among US school children based on agent-based model. BMC Public Health 2014; 14: 923.
70. Grimm V, Railsback SF. Individal-based modeling and ecology. Princeton, NJ: Princeton University Press; 2005.
71. Valente TW. Network interventions. Science 2012; 337(6090): 49-53.

8

SUMMARY / SAMENVATTING

Summary

Health behaviors, such as smoking and unhealthy eating, are among the top leading preventable risk factors of non-communicable diseases; still more than 20% of the global population smokes, around 13% is obese, approximately a quarter of the adult population does not fulfill the guidelines for physical activity, and less than a quarter of the population meets the recommendations of fruit and vegetable consumption. The prevalence of unhealthy behaviors is consistently larger among individuals in lower socioeconomic groups as compared to higher socioeconomic groups.

Determinants of health behaviors cannot be fully understood in isolation of the context in which they are shaped and sustained. Features of the physical environment, such as the availability of facilities, may constrain, reward or induce the behavior of individuals. The social environment provides opportunities for sharing norms around behaviors, social support for behavioral decisions, and peer influence. Environmental factors may also contribute to socioeconomic inequalities in unhealthy behaviors, for example because lower socioeconomic groups reside more often in neighborhoods less supportive for health behaviors, including, for example, poorer access to facilities and less social support.

Recent studies on the relevance of social networks for health behaviors showed that smoking and obesity may spread from person-to-person, suggesting that unhealthy behaviors are contagious. These findings have changed the public health relevance of the social network landscape. Nevertheless, scientific evidence for contagiousness of health behavior within social networks remains scarce, mainly due to the inherent complexity to collect longitudinal information on health behaviors in dynamic social networks.

It is increasingly recognized that individual and environmental determinants interact with and feedback on each other, creating a complex system. Making the pathways and features of such a system explicit is important, because this may help to better identify the etiology of health behaviors and to optimize ways to promote healthy behavior and prevent unhealthy behaviors. Systems approaches, such as agent-based modeling, therefore receive increasing attention. Agent-

based models (ABMs) can simulate a dynamic system in which individuals interact with each other and their environment. These models can be used to test the (long term) effects of intervention scenarios and compare them to a counterfactual. Although ABM has been successfully adopted in many fields of research, it has hardly entered the fields of social epidemiology and behavioral research thus far.

In this thesis we aim to explore and quantify the importance of social networks as a determinant of health behaviors, and to investigate the usefulness of agent-based models as a tool for assessing the impact of interventions to reduce socioeconomic inequalities in health behaviors. The specific objectives of this thesis are:

1. To investigate to which extent the spread of unhealthy behaviors and infectious diseases share similarities and how infectious disease modeling could be applied for health behavioral research.
2. To quantify the associations between social networks and smoking, sports participation and overweight, and whether these associations vary by type of social network tie.
3. To develop two agent-based models to explore the potential impact of interventions aimed at reducing income inequalities in food consumption and sports participation.

In *Chapter 2*, we describe important similarities between the spread of unhealthy behaviors and infectious diseases, and how infectious disease modeling can be applied for health behavioral research. Apart from unhealthy behaviors being contagious, we have identified six additional similarities with infectious diseases: (1) the presence of super-spreaders resembles opinion leader or role models; (2) heterogeneity in susceptibility resembles variation in adopting unhealthy behaviors; (3) heterogeneity in infectiousness is similar to variation in social influence; (4) both infections and unhealthy behaviors tend to cluster within populations, (5) the influence of the physical environment promotes or discourages the spread of infections and unhealthy behaviors, and (6) vaccination resembles social inoculation. An important implication of these similarities is that the dynamics of individual interactions and the environment at large as modeled in existing infectious disease models, and in particular individual-based models or

ABMs, can be exploited for health behavioral research.

In the next two chapters, we quantify associations between social networks and smoking, sports participation and overweight, and whether these associations vary by type of social network tie.

In *Chapter 3*, we investigate whether the prevalence of health-related behaviors and overweight in neighborhoods is associated with changes in smoking, sports participation and overweight over 13 years of follow-up in Dutch adults in the city of Eindhoven. In search for specific contextual determinants of these behaviors, much emphasis has been placed on environmental characteristics, but the prevalence of health behaviors as contextual variable has surprisingly little been studied. The reasoning behind examining neighborhood prevalence of health behaviors stems from the idea that behaviors spread from person-to-person. Individuals may change behavior because other people in their environment engage in this behavior (observational learning). We have found that living in neighborhoods with a high prevalence of non-smoking, no sports participation and overweight increase the odds of quitting smoking, quitting sports and becoming overweight. It suggests that norms or culture in favor of, for example, overweight may contribute to becoming overweight. This is an empirical novelty that adds to the existing explanations of why people engage in healthy or unhealthy behaviors and why this clusters in neighborhoods.

In *Chapter 4*, we investigate whether smoking behavior among social network members determines smoking cessation and relapse of adults. We have used longitudinal data of the Dutch Longitudinal Internet Studies for the Social sciences (LISS) panel collected in 2013 with a follow-up in 2014. Results show that respondents with the largest proportion of smokers in their social network are less likely to quit smoking and more likely to experience a relapse. We further investigate whether these associations would be different with the nature of social ties: close friends, household members and family outside the household. Smoking cessation and relapse are most strongly associated with the proportion of smokers among household members and friends. The proportion of smokers in family outside the household is not related to smoking cessation and smoking relapse. Our findings support that smoking cessation and smoking relapse are

influenced by smoking behaviors of members in the social network, and therefore also supports the idea that smoking might be contagious. Collectively reducing the proportion of smokers in social networks might not only encourage others to quit smoking too, but might also make smoking relapse less likely.

The two following chapters introduce two newly developed agent-based models. Both models are intended as proof of concepts to assess their usefulness in health behavioral research. They primarily focus on the interaction between agents and their physical environment and do not (yet) include the influence of social networks as discussed in *Chapter 2*.

Chapter 5 presents an agent-based model for income inequalities in food consumption, which includes households and food outlets as agents. The model captures the dynamics between households making food shopping decisions and food outlets (i.e. fruit/vegetable stores, supermarkets, or discount supermarkets) making decisions about startups or closures. Decisions about where to shop for food and whether to visit fast food outlets are based on individual and environmental factors: food preference, distance to a food shop, price of a food shop. These factors are valued differently by household with a high- and low-income level, resulting in income inequalities in healthy food consumption. Food shops could adapt to customer demand by closing or opening new food shops in the city. We use the model to explore the impact of three interventions to reduce healthy food consumption inequalities: (1) eliminating residential segregation based on income; (2) lowering the prices of healthy food; and (3) providing health education. Model predictions show that all interventions could reduce income inequalities in food consumption. Eliminating residential segregation shows the largest reduction, but this is partly the result of an unfavorable change in healthy food consumption among high income groups. Marked effects are only visible after five to ten years.

Chapter 6 presents an agent-based model for income inequalities in sports participation. The ABM simulates the interaction sports participation of individuals and closures or startups of sports facilities in the city of Eindhoven. Sports participation is determined by an individual's intention to do sports, which results from interactions between attributes of individuals (i.e. age, sex,

income), sports facilities (i.e. price, accessibility), and the social environment (i.e. social cohesion, social influence). This model has been calibrated to sports participation in Eindhoven and used to explore five intervention scenarios: (1) providing health education, (2) lowering prices of sports facilities, (3) increasing the availability of sports facilities, (4) improving neighborhood safety, and (5) combining all previous interventions simultaneously (i.e. multilevel intervention). Our model predicts that the interventions simulated can modestly reduce absolute income inequalities in sports participation over time, with a somewhat stronger effect for health education. Combining interventions in a multilevel approach will have the largest impact. Discernable effects are only visible after five to ten years. Our findings highlight that increasing sports participation and reducing income inequalities in sports participation requires sustained effort with population-level effects only visible in the long-term.

In *Chapter 7* the main findings of the studies are summarized and critically appraised. In this thesis, we argue that it may be beneficial to exploit models of infectious disease for health behavioral research, because the spread of unhealthy behaviors shows similarities with the spread of infectious diseases. Close collaboration with infectious disease modelers is therefore highly recommended, particularly because they have decades of experience in developing and using ABMs. We have also demonstrated in separate studies that healthy and unhealthy behaviors in personal social networks and neighborhoods are important for changes in health behavior. To which extent personal social networks and broader context of networks, such as neighbors or even neighborhoods, interact with each other to influence health behaviors is still poorly understood and therefore may be an important topic of future research. Our findings also support the idea that health behaviors might be contagious and indicate preference for tailored network-based interventions. Moreover, as health behaviors might spread to others, the effect of various interventions might be more effective than initially thought. Furthermore, we have developed two ABMs and have illustrated that these models have much potential to study complex public health problems. ABMs are in particular suitable to explore the long-term impact of interventions aimed to reduce socioeconomic inequalities in health behaviors. However, our models should be seen as proof-of-concepts; our findings cannot be interpreted literally, but are only indicative.

Agent-based modeling in health behavioral research is still in its infancy and requires further refinements to make it suitable to support decision makers.

Samenvatting

Gezondheidsgerelateerde gedragingen, zoals roken, zijn belangrijke risicofactoren van ziekte en sterfte. Ondanks de vele maatregelen om ongezond gedrag terug te dringen, rookt nog steeds meer dan 20% van de bevolking, is ongeveer 13% obees, voldoet ongeveer een kwart van de bevolking niet aan de beweegnorm, en eet minder dan een kwart van de bevolking voldoende groente en fruit. Daarbij komt ongezond gedrag vaker voor onder lage sociaaleconomische groepen dan hoge sociaaleconomische groepen.

Het is bekend dat risicofactoren van gedrag niet volledig kunnen worden verklaard door persoonlijke factoren alleen. De context of de blootstelling aan omgevingsfactoren spelen eveneens een belangrijke rol bij de vorming en instandhouding van ongezond gedrag. Kenmerken van de fysieke omgeving, zoals de aanwezigheid van faciliteiten (b.v. snackbars en sportparken), kunnen gezondheidsgerelateerde gedragingen beletten of stimuleren. De sociale omgeving biedt kansen voor het delen van normen, sociale steun en beïnvloeding van peers. Deze omgevingsfactoren dragen ook bij aan de sociaaleconomische ongelijkheid in gezondheidsgerelateerde gedragingen. Personen uit de lagere sociaaleconomische groepen wonen vaker in een minder gunstige omgeving voor gezond gedrag, en ze hebben vaak minder sociale steun voor zulk gedrag dan personen uit hogere sociaaleconomische groepen.

Recente studies tonen aan dat ongezond gedrag zich mogelijk verspreidt van persoon-tot-persoon in sociale netwerken. Dit zou kunnen betekenen dat ongezond gedrag als het ware besmettelijk is. Ondanks dat deze inzichten de kijk op de relevantie van sociale netwerken voor public health hebben veranderd, blijft verder empirisch onderzoek naar de besmettelijkheid van gedrag in sociale netwerken schaars. Dit hangt nauw samen met de complexiteit om longitudinale data te verzamelen over gezondheidsgerelateerde gedragingen in dynamische sociale netwerken.

Het is bekend dat de individuele factoren en omgevingskenmerken van gezondheidsgerelateerde gedragingen op een complexe manier onderling afhankelijk zijn van elkaar, waardoor er een complex causaal web ontstaat. Het

is belangrijk om de verbanden en kenmerken van dit samenhangend systeem expliciet te maken, omdat het kan helpen bij het identificeren van oorzaken van gedrag en optimale interventies om gezond gedrag te promoten en ongezond gedrag te voorkomen. Systeemmethoden, zoals agent-based modellen (ABMs), zijn daar uitermate geschikt voor. ABMs simuleren een dynamisch systeem waarin individuen met elkaar en met hun omgeving interacteren. Bovendien kunnen deze modellen worden gebruikt om langetermijneffecten van interventies te testen. Het bestuderen van systemen door middel van ABMs vindt in volksgezondheidsonderzoek amper plaats, met uitzondering van studies naar de verspreiding van infectieziekten.

Het onderzoek in dit proefschrift heeft als doel om het belang van sociale netwerken als riscofactor van gezondheidsgerelateerde gedragingen te kwantificeren en om het nut van ABMs om de impact van interventies op gezondheidsgerelateerde gedragingen exploratief te onderzoeken. De specifieke doelen van dit proefschrift zijn:

1. Het bestuderen van de overeenkomsten tussen de verspreiding van ongezond gedrag en de verspreiding van infectieziekten, en in hoeverre infectieziekten modellen zouden kunnen worden toegepast voor onderzoek naar ongezond gedrag.
2. Het kwantificeren van associaties tussen sociale netwerken en roken, sportparticipatie en overgewicht, en in hoeverre deze associaties variëren met het type sociale contact.
3. Het ontwikkelen van twee ABMs die de impact van interventies, gericht op het reduceren van inkomensongelijkheid in voeding en sportparticipatie, kunnen voorspellen.

Hoofstuk 2 beschrijft de overeenkomsten tussen de verspreiding van ongezonde gedragingen en de verspreiding van infectieziekten. Ook wordt er gekeken naar relevante concepten van infectieziektemodellen die zouden kunnen worden toegepast op onderzoek naar ongezond gedrag. Naast besmettelijkheid van ongezond gedrag hebben we zes andere overeenkomsten geïdentificeerd: (1) de aanwezigheid van zogenaamde “super-spreaders” (superverspreiders), (2) heterogeniteit in vatbaarheid, (3) heterogeniteit in besmettelijkheid, (4) clustering

van de infectie, (5) het belang van de omgeving, (6) vaccinatie. De belangrijkste implicatie van deze overeenkomsten is dat de dynamische interacties tussen individuen en de omgeving in infectieziektemodellen, in het bijzonder individual-based modellen of ABMs, kunnen worden geëxploiteerd voor onderzoek naar ongezond gedrag.

De volgende twee hoofstukken richten zich op het tweede doel: het kwantificeren van associaties tussen sociale netwerken en roken, sportparticipatie en overgewicht, en in hoeverre deze associaties variëren met het type sociale contact.

In *Hoofdstuk 3* bestuderen we in hoeverre de prevalentie van ongezond gedrag en overgewicht in buurten gerelateerd is aan veranderingen in rookgedrag, sportparticipatie en overgewicht. De studie richt zich op volwassenen in Eindhoven en heeft een follow-up van 13 jaren. De prevalentie van gezondheidsgerelateerde gedragingen als contextuele determinant van gedragingen is verrassend weinig bestudeerd. De reden om de prevalentie van gedrag in een buurt te onderzoeken kan worden herleid tot de eerdere bevinding dat gedrag mogelijk besmettelijk is. Het is bekend dat observatie van het gedrag van anderen in de omgeving kan leiden tot gedragsveranderingen ("observational learning"). Onze studie laat zien dat respondenten uit buurten met een hoge prevalentie van niet-rokers, niet-sporters en overgewicht een verhoogde kans hebben op het stoppen met roken, stoppen met sporten en het krijgen van overgewicht. Dit suggereert dat bijvoorbeeld een buurtcultuur of norm ten gunste van bijvoorbeeld overgewicht bijdraagt aan het krijgen van overgewicht.

In *Hoofdstuk 4* bestuderen we de relatie tussen roken onder sociale contacten en het stoppen en weer beginnen met roken. Hiervoor hebben we longitudinale data verzameld via het LISS panel in 2013 met een follow-up in 2014. Respondenten met veel rokers in hun sociale netwerk hebben een kleinere kans te stoppen met roken en een verhoogde kans weer te beginnen met roken. Deze associaties zijn het sterkst wanneer het gezinsleden of naaste vrienden die roken betreft. Roken onder familieleden die niet deel uit maken van het gezin is niet gerelateerd aan het stoppen of weer beginnen met roken. Deze bevindingen ondersteunen de gedachte dat het rookgedrag zich verspreidt in sociale netwerken en dat roken

mogelijk besmettelijk is. Collectief terugdringen van roken in sociale netwerken zal niet alleen leiden tot meer stoppen, maar kan ook de kans op een terugval verlagen.

In de volgende twee hoofdstukken worden twee ABMs geïntroduceerd. Beide modellen zijn ontwikkeld als proof-of-concept met als doel de relevantie voor onderzoek naar gezondheidsgerelateerde gedragingen aan te tonen. De invloed van sociale netwerken op gezondheidsgerelateerde gedragingen, zoals uitgelegd in *Hoofdstuk 2*, wordt in deze modellen (nog) niet ondersteund.

Hoofdstuk 5 introduceert een ABM voor inkomensongelijkheid in voeding. De agents in dit model zijn huishoudens en winkels. Het model omvat de dynamiek tussen het winkelgedrag van huishoudens en het opengaan van winkels (groente- en fruitwinkel, supermarkt, budget supermarkt, fastfoodketen) of het sluiten van bestaande, onrendabele winkels. Het winkelgedrag van huishoudens wordt bepaald door de afstand tot de aanwezige winkels, de prijzen en voorkeur voor type voeding. Het belang van deze factoren is afhankelijk van het inkomensniveau. Met het model is het winkelgedrag in Eindhoven nagebootst. Vervolgens hebben we het model gebruikt om drie interventies door te rekenen: (1) het wegnemen van residentiële segregatie, (2) het verlagen van prijzen van gezonde voeding en (3) het aanbieden van gezondheidseducatie. Onze modelpredicties laten zien dat alle interventies de absolute inkomensongelijkheid in voeding kunnen reduceren. Het wegnemen van residentiële segregatie heeft de grootste impact op ongelijkheid in voeding, maar is mede het gevolg van een ongunstig effect voor de hoge inkomensgroep. Significante effecten zijn pas na vijf tot tien jaren te zien.

Hoofdstuk 6 beschrijft een ABM voor inkomensongelijkheid in sportparticipatie. Dit model simuleert de interacties tussen sportparticipatie onder volwassenen en het sluiten en opengaan van sportfaciliteiten in Eindhoven. Sportparticipatie wordt bepaald door de intentie van een individu om te sporten, welke verder wordt beïnvloed door leeftijd, geslacht, inkomen, bereikbaarheid en prijzen van sportfaciliteiten, buurtveiligheid, sociale cohesie en de sociale invloed van buren. In totaal zijn vijf interventiescenario's getest: (1) het aanbieden van gezondheidseducatie, (2) het verlagen van prijzen van sportfaciliteiten, (3) het verhogen van het aanbod van sportfaciliteiten, (4) het verbeteren van de

buurtveiligheid en (5) het combineren van alle voorgaande interventies. Alle interventies laten een bescheiden afname zien in de absolute inkomensongelijkheid in sportparticipatie, met gezondheidseducatie als gunstigste optie. De grootste impact kan worden bereikt wanneer alle interventies worden gecombineerd. Grote effecten zijn zichtbaar na vijf tot tien jaren. Onze bevindingen benadrukken dat het verhogen van sportparticipatie en het verlagen van de inkomensongelijkheid in sporten in Eindhoven veel inzet zal vragen en pas op langere termijn impact zal hebben.

In *Hoofdstuk 7* worden de bevindingen van dit proefschrift beschreven en kritisch bediscussieerd. We concluderen dat het goed zou zijn om infectieziektemodellen te gebruiken voor onderzoek naar gezondheidsgerelateerde gedragingen gezien de overeenkomsten tussen de verspreiding van ongezond gedrag en infectieziekten. Hierbij is een samenwerking met infectieziektemodelleurs noodzakelijk, omdat zij jarenlange ervaring hebben met het ontwikkelen en toepassen van ABMs. Ook hebben we in twee studies gedemonstreerd dat gezonde en ongezonde gedragingen in sociale netwerken en buurten belangrijke factoren zijn voor gedragsveranderingen. In hoeverre naaste sociale contacten en andere sociale contacten (zoals buren of buurten) met elkaar interacteren in de beïnvloeding van gezondheidsgerelateerde gedragingen is niet bekend en vormt een belangrijk onderwerp voor toekomstig onderzoek. Onze bevindingen ondersteunen verder het idee dat gedrag mogelijk besmettelijk is. Omdat gedrag (en dus de positieve veranderingen als gevolg van een interventie) zich mogelijk verspreidt zouden interventies effectiever kunnen zijn dan in eerste instantie gedacht. Onze twee ABMs studies laten zien dat het mogelijk is een systeem na te bootsen en onderschrijven de mogelijke bruikbaarheid van ABMs om complexe volksgezondheidsproblemen mee te bestuderen. ABMs zijn vooral geschikt om de langetermijneffecten van maatregelen gericht op het verkleinen van gezondheidsverschillen in gedrag te bestuderen. Omdat onze modellen beschouwd moeten worden als een proof-of-concept zijn onze resultaten slechts indicatief. Ondanks de toenemende aandacht voor ABMs in volksgezondheidsonderzoek, staat het gebruik ervan nog steeds in zijn kinderschoenen. Verdere ontwikkelingen en verfijningen zijn noodzakelijk om ABMs op termijn bruikbaar te maken als werktuig om daadwerkelijk besluitvorming te ondersteunen.

9

DANKWOORD

Dit proefschrift heeft alleen tot stand kunnen komen dankzij de begeleiding, steun en hulp van de geweldige mensen in mijn sociale omgeving. Je zou misschien zelfs kunnen stellen dat de sociale omgeving net zo relevant is voor de totstandkoming van een proefschrift als de verspreiding van ongezond gedrag of infectieziekte. Graag wil ik iedereen bedanken voor hun bijdrage hieraan.

Jan Hendrik Richardus, promotor nummer één. Dank voor je heldere sturing en overzicht. Ik ben je zeer dankbaar voor je vertrouwen in mij dat mede heeft geleid tot onze samenwerking op het gebied van lepra.

Frank van Lenthe, promotor nummer twee. Mede dankzij je ontzettend goede begeleiding zijn mijn artikelen aanzienlijk beter geworden. Ik heb veel van je mogen leren en ben je zeer dankbaar voor het vertrouwen en de prettige samenwerking.

Sake de Vlas, promotor nummer drie(!). Drie promotoren is toch wel iets bijzonders, alhoewel...hoe langer je over je proefschrift doet, hoe waarschijnlijker dit wordt. Je was het meest kritisch van de drie. Er was niets dat je niet zag, zelfs de dubbele spaties vielen je op. Je kitiek was meestal terecht en maakte de stukken aanzienlijk beter. Ik heb heel veel van je kunnen leren en heb onze samenwerking als zeer prettig ervaren.

Alle commissieleden, dank dat jullie de tijd hebben genomen om mijn proefschrift door te nemen.

Pepijn van Empelen, je was mijn beoogde co-promotor, maar helaas ging je voor mijn eerste werkdag begon ergens anders aan de slag. Dank voor je begeleiding en steun met name aan het begin van mijn promotietraject.

Roel Bakker, toen ik aan mijn promotietraject begon had ik nog nooit een regel Javacode geschreven. Maar nu, dankzij jouw lessen en begeleiding ben ik in staat geweest onder andere de twee modellen in dit proefschrift zelf te programmeren. Ik ben je hier heel dankbaar voor.

Alle (oud) leden van de sectie infectieziekte en sociale epidemiologie. Heel veel dank voor de prettige samenwerking, de overleggen, de borrels, etentjes en uitjes. Heel graag wil een aantal van jullie in het bijzonder bedanken: Suzette,

Epke, Jan, Luc, Wilma S, Cherry, Nana, Yannan, Joost, Karen, Astrid, Peri, Anuj, Famke, Rinke, Natalie, Caroline, Wilma N en Qing.

Alle (oud) MGZ collega's bedankt voor alle koffies, lunches, lunch wandelingen, borrels, etentjes, gesprekken, overleggen en interesse. Mijn oud kamergenoten, dank voor jullie gezelligheid: Maaike, Jet, Anne. In het bijzonder wil ik Lea en sushi club bedanken voor het organiseren van de sushi avonden en ook de vele koffies en lunches.

Ineke, mijn oud kamergenoot en een beetje mijn lotgenoot. Beiden doen we iets langer over ons proefschrift dan gepland, maar uiteindelijk komen we er. Dank voor alle jaren als kamergenoot en de vele koffies, lunches daarna.

Sophie, dank dat je op deze dag mijn paranimf bent. Ik ben zeer vereerd. Samen met Ineke was jij mijn eerste kamergenoot en we waren een hechte kamer. Jullie hadden toen samen het voornemen mij alles te leren over vrouwen 'dingen'. Na al die jaren is de kennis over onderjurken, nagellak, verschillende haktypes, helaas, nog steeds present. Dank voor alle fijne gesprekken, koffies en de wekelijkse lunchwandelingen. Laten we dat vooral blijven doen!

Carmen, mijn tweede paranimf. Wat heb ik veel lol beleeft met jou als collega en kamergenoot. Dank voor alle (diepgaande) gesprekken en de vele koffies, lunches en wandelingen. Ik ben zeer vereerd dat jij mijn paranimf bent.

Vrienden en familie, ook jullie wil ik bedanken voor de steun en interesse.

Als laatst wil ik mijn ouders bedanken. Pa en ma, jullie hebben mij altijd gestimuleerd om mijn grenzen te verleggen en voor het hoogst haalbare te gaan. De dag van promotie is dan eindelijk aangebroken. Het is jammer dat pa dit niet meer kan meemaken. Dank voor alle steun. Jullie zijn mij zeer dierbaar!

10

ABOUT THE AUTHOR

Curriculum vitae

David Blok was born on February 1, 1985, in Rotterdam, the Netherlands. In 2003, he completed his secondary education at Comenius College in Capelle aan den Ijssel. Afterwards he started studying economics at the Erasmus University in Rotterdam. In 2009, he obtained his master's degree in Economics, specialized in macroeconomics, followed by a second Master's degree in Health Economics in 2010. In September 2010, he started his PhD project at the department of Public Health, Erasmus MC in Rotterdam, the results of which are presented in this thesis. In 2012, as part of his PhD program, he obtained a third Master degree in Epidemiology at the National Institute for Health Sciences (NIHES) in Rotterdam. During his PhD period, he has worked for one year as a modeler and programmer of the individual-based model for Neglected Tropical Diseases (NTDsim). In 2015, he started working as a mathematical modeler in Leprosy within the NTD modeling consortium. In January 2016, David received the Encouragement Award for best PhD publication of 2015 from the department of Public Health.

11

LIST OF PUBLICATIONS

This thesis

Blok DJ, de Vlas SJ, van Empelen P, van Lenthe FJ. The role of smoking in social networks on smoking cessation and relapse among adults: A longitudinal study. Prev Med 2017; 99: 105-10.

Blok DJ, de Vlas SJ, Bakker R, van Lenthe FJ. Reducing Income Inequalities in Food Consumption: Explorations With an Agent-Based Model. Am J Prev Med 2015; 49(4): 605-13.

Blok DJ, de Vlas SJ, van Empelen P, Richardus JH, van Lenthe FJ. Changes in smoking, sports participation and overweight: does neighborhood prevalence matter? Health Place 2013; 23: 33-8.

Blok DJ, van Empelen P, van Lenthe FJ, Richardus JH, de Vlas SJ. Unhealthy behaviour is contagious: an invitation to exploit models for infectious diseases. Epidemiol Infect 2013; 141(3): 667-9.

Other publications

Blok DJ, Crump RE, Sundaresh R, et al. Forecasting the new case detection rate of leprosy in four states of Brazil: A comparison of modelling approaches. Epidemics 2017; 18: 92-100.

Blok DJ, de Vlas SJ, Richardus JH. Finding undiagnosed leprosy cases. Lancet Infect Dis 2016; 16(10): 1113.

de Matos HJ, **Blok DJ**, de Vlas SJ, Richardus JH. Leprosy New Case Detection Trends and the Future Effect of Preventive Interventions in Para State, Brazil: A Modelling Study. PLoS Negl Trop Dis 2016; 10(3): e0004507.

de Vlas SJ, Stolk WA, le Rutte EA, Hontelez JA, Bakker R, **Blok DJ**, et al. Concerted Efforts to Control or Eliminate Neglected Tropical Diseases: How Much Health Will Be Gained? PLoS Negl Trop Dis 2016; 10(2): e0004386.

Hollingsworth TD, Adams ER, Anderson RM, Atkins K, Bartsch S, Basanez MG, Behrend M, **Blok DJ**, et al. Quantitative analyses and modelling to support

achievement of the 2020 goals for nine neglected tropical diseases. Parasit Vectors 2015; 8: 630.

Blok DJ, De Vlas SJ, Richardus JH. Global elimination of leprosy by 2020: are we on track? Parasit Vectors 2015; 8: 548.

Blok DJ, de Vlas SJ, Fischer EA, Richardus JH. Mathematical modelling of leprosy and its control. Adv Parasitol 2015; 87: 33-51.

12

PHD PORTFOLIO

PhD traning	Period	Workload
Skills training		
Master of Health Sciences (specialization in Epidemiology), Netherlands Institute for Health Sciences	2010-2012	70 ECTS
Deel-BKO certificate	2015-2017	20 hours
Teach the teacher I		
BKO workshop: Omgaan met groepen		
Workshop Scientific Integrity	2016	8 hours
Awards		
MGZ best PhD publication award	2015	
Conference presentations		
Poster presentation: *"Veranderingen in rookgedrag, sport participatie en overgewicht: speelt buurtprevalentie daarbij een rol?"* Nederlands Congres Volksgezondheid, Ede, The Netherlands	2013	16 hours
Oral presentation: *"Health-related behaviors are contagious: Changes in smoking, sports participation and overweight: does neighborhood prevalence matter?"* WEON, Utrecht, The Netherlands	2013	16 hours
Oral presentation: *"Changes in smoking, sports participation and overweight: Does neighborhood prevalence matter?"* International Society of Behavioral Nutrition and Physical Activity (ISBNPA), Ghent, Belgium	2013	32 hours
Oral presentation: *"An agent-based model for simulating scenarios of various interventions aimed at the reduction of socioeconomic inequalities in diet"* International Society of Behavioral Nutrition and Physical Activity (ISBNPA), San Diego, USA	2014	32 hours

Oral presentation: "*The Impact of Individual and Environmental Interventions on Income Inequalities in Sports Participation: Exploration with an Agent-Based Model*" International Society of Behavioral Nutrition and Physical Activity (ISBNPA), Victoria, Canada	2017	32 hours
Teaching activities		
Supervisor medical students theme 3.C.4. Community projects, Erasmus MC	2013-2018	60 hours
Lecturer medical students theme 3.C.1 VO: Regionale sterfteverschillen, Erasmus MC	2015-2018	40 hours
Teaching assistant NIHES course Biostatistics I	2014-2015	40 hours
Lecture "Agent-based modelling in public health", as part of NIHES course Public Health Research	2016	8 hours
Other activities		
Reviewer international scientific journals	2013-2018	
Organizing section outing Social Epidemiology	2014	8 hours
PhD day	2013 & 2014	16 hours