Natural experiments to evaluate local public health interventions

Famke J.M. Mölenberg

Natural experiments to evaluate local public health interventions

Natuurlijke experimenten voor evaluatie van lokale volksgezondheidsinterventies

Famke J.M. Mölenberg

MIX
Paper from responsible sources
FSC® C128610

Natural experiments to evaluate local public health interventions
Doctoral thesis, Erasmus University Rotterdam

Famke J.M. Mölenberg

ISBN: 978-94-6421-336-2

Cover design by Fleur van der Put | www.fleurisenl.com
Layout and design by Rowen Aker | www.persoonlijkproefschrift.nl
Printed by Ipskamp Printing | proefschriften.net

This thesis was printed with financial support of the Department of Public Health, Erasmus Medical Center and of the Erasmus University Rotterdam.

Natural Experiments to Evaluate Local Public Health Interventions

Natuurlijke experimenten voor evaluatie van lokale volksgezondheidsinterventies

Proefschrift

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

Prof.dr. F.A. van der Duijn Schouten

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

vrijdag 25 juni 2021 om 13.00 uur

door

Famke Johanna Maria Mölenberg
geboren te Zutphen.

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	prof.dr. F.J. van Lenthe prof.dr.ir. A. Burdorf
Overige leden:	prof.dr. J.L.W. van Kippersluis prof.dr. S Denktas dr. E.M.F. van Sluijs

CONTENTS

John Snow moved the focus upstream, tackling the root of the problem, preventing, rather than curing disease. He is now considered to be the father of epidemiology.

Anna Coote, The wisdom of prevention. London: New economics foundation, 2012

1

General introduction

WORLDWIDE CHALLENGES FOR PUBLIC HEALTH

Despite governmental commitments to reduce the global burden of non-communicable diseases (NCDs), cardiovascular diseases, cancers, respiratory diseases, and diabetes have remained a major public health challenge. In 2017, 74% of all global deaths are due to NCDs (1). It has been estimated that more than 34 million deaths or 84% of all NCD deaths could be prevented through health policies, population-based interventions and access to high-quality health care (1). Important modifiable risk factors related to the incidence of the above-mentioned chronic diseases and a reduced life-expectancy are tobacco use, unhealthy diet, alcohol consumption and physical inactivity (2). Disability Adjusted Life Years (DALYs) is used to quantify the years lost in good health due to premature death and due to living with diseases and injuries. The burden of specific risk factors differs across countries. Figure 1 shows the distribution of DALYs attributable to the modifiable risk factors in the Netherlands for the year 2015 (3). Tobacco use is the leading risk factor, and the total disease burden is expected to decrease by 9.4% if smoking is completely eliminated in the population. Some risk factors will overlap, thus the combined disease burden will be lower than the sum of the two individual risk factors.

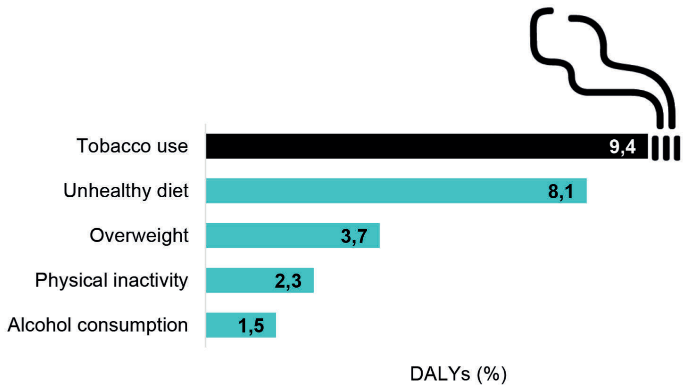


Figure 1: The disease burden of modifiable risk factors in the Netherlands in 2015, based on Hilderink et al. (3)

Abbreviation: Disability Adjusted Life Years (DALYs).

With the exception for tobacco use, no progress has been made in reducing the prevalence of potentially modifiable risk factors over the past decades. Diet quality did not substantially deteriorate nor improve (2), while physical inactivity increased in most high-income countries (4, 5). The worldwide prevalence of overweight (body-mass index [BMI] ≥ 25 kg/m²) among adults almost doubled, from 21.5% in 1975 to 38.9% in 2016 (6). The prevalence of obesity (BMI ≥ 30 kg/m²) nearly tripled, from 4.7% in 1975 to 13.1% in 2016. No country yet managed to significantly decline the proportion of the population with overweight or obesity (2, 6). At best, trends in obesity for children and adolescents seemed to have levelled off, at high levels, in some high-income countries (7).

In the Netherlands, like in most high-income countries, we hold a poor track record of reducing overweight and obesity. The Dutch adult population with overweight increased from 35.1% in 1990 to 50.1% in 2018, and obesity increased from 6.1% in 1990 to 14.8% in 2018 (8). Without effective interventions, a continued increase in obesity can be expected in the next two decades. Projected trends suggest that the prevalence of overweight in adults will further rise to 62.2% in 2040 (3).

Tobacco use is an exception. The worldwide prevalence of tobacco smoking declined steadily (2). Also in the Netherlands, there was a steep decline in the prevalence of smoking among adults from 38.2% in 1990 to 21.7% in 2018 (9). In the next decades, adult smoking rates are expected to decrease to 13.6% in 2040, resulting from smoking cessation among adult smokers and a decrease in the prevalence of youth who initiate smoking (3).

Why has tobacco use successfully reduced over the past decades, while in the same period reducing – or at the very least stabilise – overweight and obesity remained such a public health challenge? These questions are typically addressed by studying individuals' risk for certain diseases, and how health-related behaviours can be changed to mitigate their disease risk. However, the large increase in overweight and obesity across societies calls for a different perspective. A shift in focus is needed from understanding why individuals are at risk of becoming overweight towards understanding why overweight is highly prevalent in most societies. What are the determinants of the increase of the prevalence of obesity, and subsequently, what type of interventions and policies are needed to target these drivers of the "obesity epidemic"?

The role of the environment in shaping individual behaviours

Changes in habits and practices of individuals that resulted in overweight and obesity cannot be seen separate from societal changes to which individuals are exposed. Our contemporary society has been described as a so-called "obesogenic environment" that increasingly promotes unhealthy behaviour, such as a high energy intake and sedentary behaviour (10). The past decades, the environment has unfavourable changed towards offering more energy-dense and ultra-processed foods at low costs (11, 12), increased portion size of energy-dense, nutrient-poor foods (13-15), and urban space devoted to cars (16, 17). In such environments, the possibilities to adopt health-related behaviours are hindered. Not only the physical environment has changed, but also changes in the economic, social, cultural, political and legal environment to which individuals are exposed may influence health-related behaviours. Changes in our way of life, and the inability of our bodies for an adequate and timely biological response to those changes, likely contributed to disease occurrence (18).

In light of the increasing prevalence of potentially modifiable risk factors, it is emphasised that creating an environment in which it is easier for individuals to adopt health-related behaviours is essential to combat the burden of NCDs (19-21). Health promotion activities taking place in environments that are supportive of health-related behaviours have the potential to achieve sustained behaviour change. Swinburn and colleagues have outlined a framework to categorise obesity determinants and solutions (11). Figure 2 shows this framework, and it recognises that the determinants of obesity operate at various levels. The more proximal determinants of obesity relate to the behaviour of individuals, whereas the more distal determinants relate to the environment to which individuals are exposed. The majority of interventions rely on educational and motivational strategies to achieve behavioural change in individuals. Interventions targeting populations are less often implemented and often lack rigorous evaluation, thus little is known about the effects of interventions that are most likely to influence the population as a whole (22). The framework specifically focusses on the determinants of obesity, and a similar categorisation can be helpful for addressing other health-related behaviours.

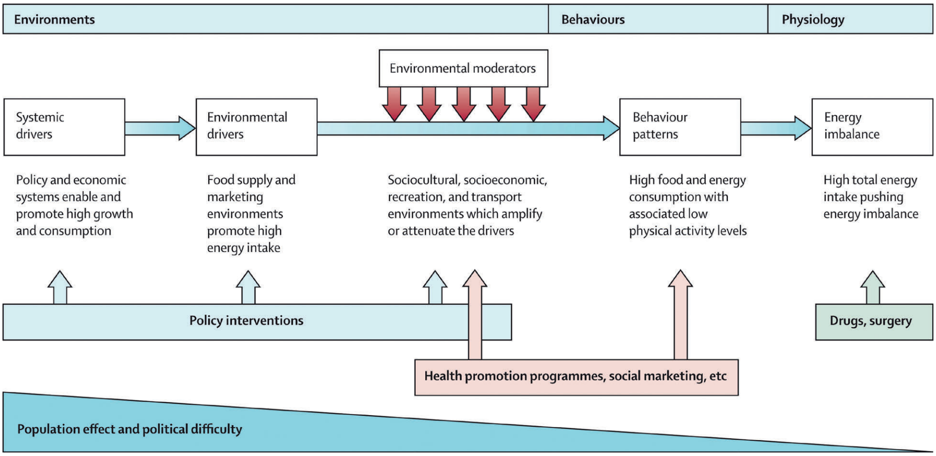


Figure 2: Framework categorising obesity determinants and solutions as presented by Swinburn et al. (11)

Strategies to reduce health inequalities

Environmental interventions may also be of importance to reduce health inequalities. In almost all societies, lower educated individuals have a lower life-expectancy, and are also experiencing more health problems as compared to their higher educated counterparts. For the period 2015-2018, the life-expectancy in the Netherlands of individuals with a primary education or less is four and a half year lower than that of individuals with a bachelor or master's degree (23). More striking is the 14 years difference lived in good health (24). Smoking, unhealthy diet, and physical inactivity contribute substantially to these inequalities (25, 26).

Although a variety of interventions such as counselling programs, fruit and vegetable distribution programs, and signs to promote stair use have been implemented in schools, companies and neighbourhoods, most were not effective in reducing inequalities in health-related behaviours and obesity (27-29). A likely explanation is that higher educated people may be in a better position to benefit from interventions that directly target individuals (30). Environmental interventions that do not heavily rely on the capacity of individuals to adopt health-related behaviours may provide valuable instruments to reduce socioeconomic inequalities in the population (31, 32). For example, governmental action through tobacco control policies have been essential in achieving large reductions in smoking over the past decades, showing that taxation and smoke-free policies can lead to behaviour change in the population (33, 34). There is also evidence indicating that the benefits of taxation were larger among lower socioeconomic groups, thereby reducing socioeconomic inequalities in smoking (35-37). Environmental interventions are likely needed to reduce the unequal distribution of health-related behaviours and obesity in the population.

Natural experiments to evaluate real world interventions

Evaluating interventions delivered at the population-level is challenging. Classically, randomised controlled trials (RCTs) are considered the “gold standard” to evaluate the effectiveness of interventions (38). The principle of an RCT is that persons are randomly assigned by the researcher to an intervention or control condition. The major benefit of this approach is that all known and unknown factors associated with the intervention or outcome are evenly distributed across the intervention and control group, and will not bias the outcome whether the intervention is beneficial or not. RCTs are useful for evaluating interventions with relatively short and simple causal pathways (38, 39). The first trials included tens of participants assigned to either intervention or control condition. The scale and complexity of RCTs have substantially increased towards studies randomly assigning thousands of participants to Mediterranean diets (40), school-based physical activity programmes (41, 42), and tens of thousands of participants to screening interventions (43-45). The history of clinical trials in medicine is presented in Panel 1.

It is unlikely that evidence of public health interventions that may yield substantial benefits for population health and health inequalities will be solely derived from RCTs. Some interventions are impractical, unfeasible, or unethical to randomise, while other interventions are simply not under control by researchers and therefore do not allow to be randomised (48). Public health interventions targeting populations represent large-scale “natural experiments”. There is a growing recognition that natural experiments may provide valuable insights on the effectiveness of public health interventions that are not amenable for randomisation (48-51).

Panel 1 – The history of randomized controlled trials (46, 47)

In 1747, James Lind was the first who conducted a controlled trial to identify a treatment of scurvy. On board of the Salisbury, he selected twelve patients and assigned each pair of patients to one of the six treatments. The patients that were assigned to an intervention of two oranges and one lemon a day for a period of six days experienced the best recovery. This was the first trial conducted in medical history.

The first RCT in medicine was carried out in 1946 by the UK Medical Research Council on determining the best treatment for pulmonary tuberculosis. Patients were randomly assigned to streptomycin and bed-rest or the standard treatment at that time of bed-rest alone for 15 months. The results of the RCT showed that mortality was lower among the patients assigned to streptomycin. The statistician of this project, Sir Bradford Hill, is widely known for his work on identifying causal relationships between exposures and disease outcomes in epidemiological studies.

Natural experiments lack a clear definition, but the key feature is that the change in exposure to the intervention is caused by external shocks or changes to a system that are outside researchers control (48). The shift in exposure in the population is used to assign individuals to intervention and control groups. While under ideal circumstances groups are exchangeable, it is not uncommon that potential confounding remains. Studies based on natural experiments can strengthen the causal interpretation of the findings when having a good understanding of the intervention and exposure processes, selecting appropriate study design and analytical methods, and transparent reporting of the study findings (48). Some disciplines use a more narrow definition of natural experiments, emphasising the importance that interventions should be randomly assigned or as good as random (52). It is unlikely that the majority of important public health lessons can be learned from these type of experiments. For example, a sudden increase in income due to winning the lottery may offer opportunities to study the impact of income on health outcomes, however, it remains questionable if similar health effects can be expected following specific policies that cause an increase in income. The latter is likely to be more informative to address public health issues. Therefore, a broader definition of natural experiments is warranted in our field of research. In this thesis we will refer to natural experiments as defined by the UK Medical Research Council, with its emphasis on the shift in exposure that has not been manipulated by the researcher (48).

The history of natural experiments goes back to John Snow in the mid-nineteenth century during the cholera outbreak in London, England. He reported on one of the first natural experiments, presented in Panel 2. More recently, natural experiments are used in public health research to evaluate policies and changes made to the environment that are not manipulated by researchers, and for which evidence cannot be derived from RCTs. Effective population-level approaches to support health-related behaviours include tobacco control policies (33, 34), sugar-sweetened beverages taxes (53, 54), and the installation of cycling infrastructure (55, 56). Evidence derived from natural experiments has contributed to the evidence base.

Practice-based evidence derived from natural experiments may provide valuable insights in the effectiveness of health promotion activities when implemented in real-life settings. These insights are essential to support policymaking; evidence from RCTs and laboratory experiments do not always replicate in the field. Natural experiments have long been an underutilised tool in public health (49). The guidance on natural experiments by the UK Medical Research Council (48) has contributed to new evidence derived from natural experiments that evaluated governmental actions at the national level. However, their usage for evaluating local public health interventions remained scarce.

The responsibility of local governments to safeguard population health has increased over the years. Local health policies and environmental interventions hold the potential to improve population health. For its evaluation we need to rely on opportunities that built on natural experiments. Despite increasing calls to do so, interventions targeting populations often lack rigorous evaluation. Uncertainty about methodological approaches and limited funding may drive this “inverse evidence law”, whereby little is known about the effects of interventions that are most likely to influence the population as a whole (22). In this thesis, we aim to identify and evaluate interventions introduced at the local level that address modifiable risk factors in the population through natural experiments.

Panel 2 – The history of natural experiments (57, 58)

One of the first natural experiments is the famous fieldwork of John Snow during the cholera outbreaks of 1849 and 1853-54 in London, England. At that time, two water companies were responsible for the water supply of the population. He collected data on cholera deaths per household and determined the company that provided the water during the two outbreaks. During the first cholera outbreak in 1849, both water companies supplied households with water from the region downstream of the River Thames, and death rates were comparable across households. Between the two outbreaks, one of the companies moved the pipe upstream, so that fresh water entered the households. Snow compared the cholera outbreak of 1849 with the outbreak of 1853-54 and showed that the mortality rates were lower for households that were supplied by the company that used the more upstream water. In absence of randomisation, he convincingly showed that contaminated water caused the cholera outbreaks based on the evidence derived from a natural experiment.

AIMS

It is important to identify and evaluate local public health interventions that are likely to improve the health of the population, but also to reflect on the methodological challenges that one may encounter for the evaluation of such natural experiments. Therefore, the current thesis addresses two objectives:

I. To evaluate local public health interventions aimed at improving health-related behaviours and aimed at reducing health inequalities in the population;
II. To appraise the methodology to evaluate local public health interventions using natural experiments.

OUTLINE OF THIS THESIS

This thesis is divided into two parts. Part one of this thesis focusses on the evaluation of local public health interventions through natural experiments. Specifically, chapter 2 focusses on the introduction of physical activity spaces in deprived neighbourhoods to increase outdoor play in children. Chapter 3 shows the inequalities of the food environment and its association with children's body composition measures. Chapter 4 presents the association between elite sport events and adult sport participation, and the shift in policy priorities towards promoting recreational sport participation. Chapter 5 and 6 focus on smoke-free policies in private or outside areas to reduce tobacco smoke exposure and respiratory disease in children.

Part two of the thesis reflects on approaches and methods used to evaluate interventions using natural experiments. Chapter 7 presents a systematic review on the methods used to evaluate infrastructural interventions to promote cycling. Chapter 8 presents a framework that was developed to evaluate if declines in response rates are biasing trends in health-related behaviours in health surveys. Chapter 9 describes the barriers we encountered to evaluate interventions using natural experiments in this public health PhD trajectory.

REFERENCES

1. Martinez R, Lloyd-Sherlock P, Soliz P, Ebrahim S, Vega E, Ordunez P, et al. Trends in premature avertable mortality from non-communicable diseases for 195 countries and territories, 1990-2017: a population-based study. The Lancet Global Health. 2020;8(4):e511-e23.
2. Murray CJL, Aravkin AY, Zheng P, Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1223-49.
3. Hilderink HBM, Verschuuren M. Volksgezondheid Toekomst Verkenning 2018: Een gezond vooruitzicht. Synthese. 2018.
4. Ng SW, Popkin BM. Time use and physical activity: A shift away from movement across the globe. Obesity Reviews. 2012;13(8):659-80.
5. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. The Lancet Global Health. 2018;6(10):e1077-e86.
6. World Health Organisation. Global Health Observatory data repository [updated 22-09-2017. Available from: https://apps.who.int/gho/data/node.main.A896.]
7. Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. The Lancet. 2017;390(10113):2627-42.
8. Volksgezondheidenzorg.info. Trend Volwassenen met overgewicht 1990-2018: Bilthoven: RIVM; [updated 25-05-2020. Available from: https://www.volksgezondheidenzorg.info/onderwerp/overgewicht/cijfers-context/trends#node-trend-overgewicht-volwassenen.]
9. Volksgezondheidenzorg.info. Trend roken volwassenen: Bilthoven: RIVM; [updated 25-05-2020. Available from: https://www.volksgezondheidenzorg.info/onderwerp/roken/cijfers-context/trends#node-trend-roken-volwassenen.]
10. Swinburn B, Egger G, Raza F. Dissecting Obesogenic Environments: The Development and Application of a Framework for Identifying and Prioritizing Environmental Interventions for Obesity. Preventive Medicine. 1999;29(6):563-70.
11. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. The Lancet. 2011;378(9793):804-14.
12. Monteiro CA, Moubarac JC, Cannon G, Ng SW, Popkin B. Ultra-processed products are becoming dominant in the global food system. Obes Rev. 2013;14(S2):21-8.

13. Kerr MA, Rennie KL, McCaffrey TA, Wallace JMW, Hannon-Fletcher MP, Livingstone MBE. Snacking patterns among adolescents: a comparison of type, frequency and portion size between Britain in 1997 and Northern Ireland in 2005. British Journal of Nutrition. 2008;101(1):122-31.

14. Steenhuis IHM, Leeuwis FH, Vermeer WM. Small, medium, large or supersize: trends in food portion sizes in the Netherlands. Public Health Nutrition. 2010;13(6):852-7.

15. Piernas C, Popkin BM. Food Portion Patterns and Trends among U.S. Children and the Relationship to Total Eating Occasion Size, 1977–2006. The Journal of Nutrition. 2011;141(6):1159-64.

16. Karsten L. It all used to be better? Different generations on continuity and change in urban children's daily use of space. Children's Geographies. 2005;3(3):275-90.

17. Norton PD. Street Rivals: Jaywalking and the Invention of the Motor Age Street. Technology and Culture. 2007;48:331 - 59.

18. Mackenbach JP. A History of Population Health: Rise and Fall of Disease in Europe: Brill Rodopi; 2020.

19. Frieden TR. A Framework for Public Health Action: The Health Impact Pyramid. American Journal of Public Health. 2010;100(4):590-5.

20. Capewell S, Capewell A. An effectiveness hierarchy of preventive interventions: neglected paradigm or self-evident truth? Journal of Public Health. 2017;40(2):350-8.

21. Marteau TM, White M, Rutter H, Petticrew M, Mytton OT, McGowan JG, et al. Increasing healthy life expectancy equitably in England by 5 years by 2035: could it be achieved? Lancet. 2019;393(10191):2571-3.

22. Ogilvie D, Egan M, Hamilton V, Petticrew M. Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go? J Epidemiol Community Health. 2005;59(10):886-92.

23. Volksgezondheidenzorg.info. Levensverwachting bij geboorte naar opleiding: Bilthoven: RIVM; [updated 11-11-2020. Available from: https://www.volksgezondheidenzorg.info/onderwerp/levensverwachting/cijfers-context/huidige-situatie#node-levensverwachting-bij-geboorte-naar-opleiding.]

24. Volksgezondheidenzorg.info. Gezonde levensverwachting bij geboorte naar opleiding: Bilthoven: RIVM; [updated 11-11-2020. Available from: https://www.volksgezondheidenzorg.info/onderwerp/gezonde-levensverwachting/cijfers-context/huidige-situatie#node-gezonde-levensverwachting-bij-geboorte-naar-opleiding.]

25. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimaki M, et al. Association of socioeconomic position with health behaviors and mortality. JAMA. 2010;303(12):1159-66.

26. Oude Groeniger J, Kamphuis CB, Mackenbach JP, van Lenthe FJ. Repeatedly measured material and behavioral factors changed the explanation of socioeconomic inequalities in all-cause mortality. J Clin Epidemiol. 2017;91:137-45.

27. Magnée T, Burdorf A, Brug J, Kremers SP, Oenema A, van Assema P, et al. Equity-specific effects of 26 Dutch obesity-related lifestyle interventions. Am J Prev Med. 2013;44(6):e57-66.

28. Robroek SJW, Oude Hengel KM, van der Beek AJ, Boot CRL, van Lenthe FJ, Burdorf A, et al. Socio-economic inequalities in the effectiveness of workplace health promotion programmes on body mass index: An individual participant data meta-analysis. Obes Rev. 2020;21(11):e13101.

29. Coenen P, Robroek SJW, van der Beek AJ, Boot CRL, van Lenthe FJ, Burdorf A, et al. Socioeconomic inequalities in effectiveness of and compliance to workplace health promotion programs: an individual participant data (IPD) meta-analysis. Int J Behav Nutr Phys Act. 2020;17(1):112.

30. Victora CG, Vaughan JP, Barros FC, Silva AC, Tomasi E. Explaining trends in inequities: evidence from Brazilian child health studies. The Lancet. 2000;356(9235):1093-8.

31. Adams J, Mytton O, White M, Monsivais P. Why Are Some Population Interventions for Diet and Obesity More Equitable and Effective Than Others? The Role of Individual Agency. PLOS Medicine. 2016;13(4):e1001990.

32. Capewell S, Graham H. Will Cardiovascular Disease Prevention Widen Health Inequalities? PLOS Medicine. 2010;7(8):e1000320.

33. Chaloupka FJ, Straif K, Leon ME. Effectiveness of tax and price policies in tobacco control. Tobacco control. 2011;20(3):235-8.

34. Frazer K, Callinan JE, McHugh J, van Baarsel S, Clarke A, Doherty K, et al. Legislative smoking bans for reducing harms from secondhand smoke exposure, smoking prevalence and tobacco consumption. Cochrane Database Syst Rev. 2016;2(2):CD005992.

35. Brown T, Platt S, Amos A. Equity impact of population-level interventions and policies to reduce smoking in adults: A systematic review. Drug and Alcohol Dependence. 2014;138:7-16.

36. Hill S, Amos A, Clifford D, Platt S. Impact of tobacco control interventions on socioeconomic inequalities in smoking: review of the evidence. Tobacco Control. 2014;23(e2):e89-e97.

37. Smith CE, Hill SE, Amos A. Impact of population tobacco control interventions on socioeconomic inequalities in smoking: a systematic review and appraisal of future research directions. Tobacco Control. 2020.

38. Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. Journal of Epidemiology and Community Health. 2003;57(7):527-9.

39. Cesar GV, Jean-Pierre H, Jennifer B. Evidence-Based Public Health: Moving Beyond Randomized Trials. American Journal of Public Health. 2004;94(3):400-5.

40. Estruch R, Ros E, Salas-Salvadó J, Covas M-I, Corella D, Arós F, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet. New England Journal of Medicine. 2013;368(14):1279-90.

41. Jansen W, Borsboom G, Meima A, Zwanenburg EJ-V, Mackenbach JP, Raat H, et al. Effectiveness of a primary school-based intervention to reduce overweight. International Journal of Pediatric Obesity. 2011;6(2Part2):e70-e7.

42. Corder K, Sharp SJ, Jong ST, Foubister C, Brown HE, Wells EK, et al. Effectiveness and cost-effectiveness of the GoActive intervention to increase physical activity among UK adolescents: A cluster randomised controlled trial. PLoS Med. 2020;17(7):e1003210.

43. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. The Cochrane database of systematic reviews. 2013;2013(6):CD001877-CD.

44. Lin JS, Piper MA, Perdue LA, Rutter CM, Webber EM, O'Connor E, et al. Screening for Colorectal Cancer: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA. 2016;315(23):2576-94.

45. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. New England Journal of Medicine. 2020;382(6):503-13.

46. Medical Research Council. Streptomycin treatment of tuberculous meningitis. The Lancet. 1948;251(6503):582-96.

47. Dunn PM. James Lind (1716-94) of Edinburgh and the treatment of scurvy. Archives of Disease in Childhood - Fetal and Neonatal Edition. 1997;76(1):F64-F5.

48. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health. 2012;66(12):1182-6.

49. Petticrew M, Cummins S, Ferrell C, Findlay A, Higgins C, Hoy C, et al. Natural experiments: an underused tool for public health? Public Health. 2005;119(9):751-7.

50. Bärnighausen T, Tugwell P, Røttingen JA, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series-paper 4: uses and value. J Clin Epidemiol. 2017;89:21-9.

51. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. Journal of Epidemiology and Community Health. 2020;74(2):203-8.

52. Dunning T. Natural Experiments in the Social Sciences: A Design-Based Approach. Cambridge: Cambridge University Press; 2012.

53. Backholer K, Sarink D, Beauchamp A, Keating C, Loh V, Ball K, et al. The impact of a tax on sugar-sweetened beverages according to socio-economic position: a systematic review of the evidence. Public Health Nutr. 2016;19(17):3070-84.

54. Teng AM, Jones AC, Mizdrak A, Signal L, Genç M, Wilson N. Impact of sugar-sweetened beverage taxes on purchases and dietary intake: Systematic review and meta-analysis. Obesity Reviews. 2019;20(9):1187-204.

55. Stappers NEH, Van Kann DHH, Ettema D, De Vries NK, Kremers SPJ. The effect of infrastructural changes in the built environment on physical activity, active transportation and sedentary behavior – A systematic review. Health & Place. 2018;53:135-49.

56. Panter J, Guell C, Humphreys D, Ogilvie D. Can changing the physical environment promote walking and cycling? A systematic review of what works and how. Health & Place. 2019;58:102161.

57. Snow J. On the mode of communication of cholera: John Churchill; 1855.

58. Smith GD. Commentary: Behind the Broad Street pump: aetiology, epidemiology and prevention of cholera in mid-19th century Britain. International Journal of Epidemiology. 2002;31(5):920-32.

Part I

Evaluation of local public health interventions

2

New physical activity spaces in deprived neighborhoods: does it change outdoor play and sedentary behavior? A natural experiment

Famke J.M. Mölenberg, J. Mark Noordzij, Alex Burdorf, Frank J. van Lenthe

Health & Place. 2019;58:102151

ABSTRACT

Introduction

We used the introduction of dedicated physical activity (PA) spaces in Rotterdam, the Netherlands, to study the impact of reducing distance to dedicated PA spaces on outdoor play and sedentary behaviour, and to evaluate if these effects were similar between population subgroups.

Methods

We included 1,841 Dutch children from the Generation R study who participated at two subsequent measurement waves when the children were, on average, 6.0 and 9.7 years old. None of these children lived within 600 meters of a dedicated PA space at baseline, and during follow-up 171 children became exposed to 13 new PA spaces within 600 meters from home. Individual-level fixed-effects models were used to evaluate changes in distances (determined by Geographical Information Systems) from home to the nearest new dedicated PA space, to parent-reported outdoor play and sedentary behaviour.

Results

The introduction of a dedicated PA space within 600 meters from home, and the reduction of the distance per 100 meter, did not affect outdoor play or sedentary behaviours. At $p<0.1$, significant interaction terms were found between the introduction of the PA spaces and indicators of family socioeconomic position. Although not statistically significant, stratified analyses showed a consistent pattern, suggesting that reducing the distance to the nearest PA space increased outdoor play for children from parents with lower levels of education. However, they also showed a non-significant increase in sedentary behaviours for children from families with net household income below average Dutch income, and for children from a non-Dutch ethnicity.

Conclusion

Introducing dedicated PA spaces may be a promising approach to increase outdoor play for children from more socioeconomically disadvantaged families, but larger studies are needed to contribute to the evidence.

INTRODUCTION

Promoting physical activity at young ages is a key strategy to combat childhood obesity (1-3). A supportive neighbourhood with access to physical activity (PA) spaces is considered to be important for this purpose. However, there is little robust evidence on the causal relationship between changes in the built environment and physical activity behaviour (4-6). Most studies have relied on cross-sectional data, leaving the question unanswered whether higher access to PA spaces make children living in neighbourhoods more physically active, or whether parents who would like their children to engage in physical activity reside in neighbourhoods with better opportunities to do so.

Randomized controlled trials (RCTs) are the gold standard to demonstrate causality, but it is difficult or perhaps even impossible to randomly assign play facilities. So-called “natural experiments” provide an alternative for situations in which the researcher lacks control over the intervention, but where the variation in access to play facilities can be used to allocate an intervention and control status to individuals (7). Due to the non-random introduction of PA spaces – presumably there where the need is largest – children who will and will not live closer to PA spaces after the follow up period may differ in many ways. However, as long as the change in physical activity behaviours within a child is independent from factors associated with the introduction of PA spaces, the introduction of PA spaces can be seen as an “exogenous” intervention. To the extent that time-invariant factors determine the introduction of PA spaces and the change in physical activity, fixed-effects analyses control for such (unobserved) confounding.

In the past years, two foundations established by Dutch sports legends (Richard Krajicek, former professional tennis player, Wimbledon champion; Johan Cruyff, former professional football player and coach) introduced new PA spaces in Dutch cities to encourage physical activity at young ages, with a special focus on children living in deprived neighbourhoods. The introduction of 18 dedicated PA spaces in the city of Rotterdam, the Netherlands, provides the unique opportunity to evaluate the impact of a changing built environment on health behaviour. Specifically, this study aimed to investigate the impact of the introduction of dedicated PA spaces on outdoor play and sedentary behaviour of children. The PA spaces specifically target deprived neighbourhoods; our secondary aim therefore was to evaluate whether the observed effects vary by family household income, parental education level, and ethnicity.

METHODS

Study design

We evaluated the introduction of 18 PA spaces between February 2008 and December 2015, using data from the Generation R prospective birth-cohort study (8). The timeline of the intervention and data collection is presented in Figure 1.

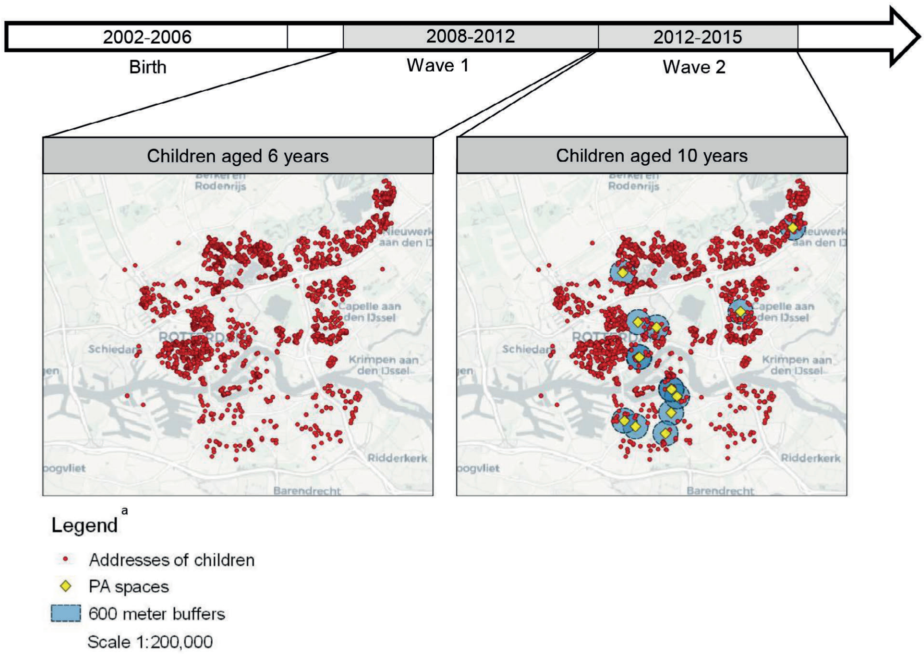


Figure 1: Timeline of the natural experiment

[a] The map was derived from OpenMapTiles (http://openmaptiles.org/)

The intervention: 18 new physical activity spaces in Rotterdam

In most cases the local government in Rotterdam applied for a dedicated PA space, but in exceptional cases (approximately 1 out of 10) residents initiated the application procedure. The applications are considered by the foundations on predefined criteria. Neighbourhoods that are eligible for a PA space are deprived, have low physical activity levels or sport participation rates among youth, or can otherwise show that the introduction of PA spaces is likely to be of benefit for children's development. Neighbourhood support is essential, and local residents are involved in the decision-making process about, for example, the design, location and the activities hosted on the new PA space. The local government is responsible for providing the additional funds needed for the introduction and maintenance of the facilities.

An impression of the PA spaces is given in Supplemental Figure 1. The PA spaces have many similarities, and target children aged 6-18 years. Although the design of the multifunctional PA spaces is tailored to the needs of the specific location, specific features like goals, colourful markings, and fences are present at most locations. PA spaces included a soccer field, basketball court, tennis field, playground equipment, or a combination here off. Some PA spaces additionally contained a mini-athletics track, panna-court, tennis table, skating rink, fitness items, volleyball field, or dance floor. All PA spaces are freely accessible, centrally located in the neighbourhood, often supervised during peak usage hours, and regularly host sports activities. The first Krajicek Playground in Rotterdam, the Netherlands, was opened in 2001; the first Cruyff Court in 2005.

Study population

Data from the Generation R study, a population-based, prospective birth cohort study were used to evaluate the introduction of the PA spaces. Invitations to participate in the Generation R study were sent out to all pregnant women who had an expected delivery date between April 2002 and January 2006 and who lived in the study area (Rotterdam, the Netherlands) at time of delivery. More information is presented in the design and cohort update paper (8). The Medical Ethics Committee of the Erasmus University Medical Center in Rotterdam approved the study. Written informed consent was obtained from all participants.

For this study, we included children who participated in two subsequent measurement waves when the children were on average 6.0 years old (February 2008 to January 2012) and 9.7 years old (February 2012 to December 2015) (n= 7,254). Two questionnaires were returned by 4,886 parents (67%). At each measurement wave we excluded children that no longer lived in Rotterdam (n=1,485), children with a missing or invalid residential address (n=137), children without repeated measures for outdoor play or sedentary behaviour (n=589), outliers with unrealistic levels of outdoor play >5 hour/day (n=51), and children without any information about time-varying net household income (n=85), sport participation (n=15), or active transport to school (n=16). Furthermore, we excluded children that moved houses during follow-up (n=333), to avoid selective migration. Clustering could occur within families, therefore we excluded younger siblings from the same mother (n=211).

Following recent recommendations to use observational data for an evaluation as if it was a trial (9), we excluded children that were exposed to dedicated PA spaces at baseline within 600 meters of their home (n=442). The population of analyses included 1,841 children without access to PA spaces. Due to different numbers of missing outcome variables, the population of analyses were different for outdoor play (n=1,607) and sedentary behaviour (n=1,545).

Distance to PA spaces using buffers

Information about the location of the PA spaces was obtained from the Krajicek and Cruyff foundations. The software QGIS was used to create Euclidian buffers of 600 meters around children's homes. Commonly used street network distances are not available in the Generation R study. Yet, the city planning of the Netherlands is characterized by an extremely high connectivity, including small alleys only accessible by feet or bike, perfectly suitable for children. In this context it is reasonable to assume that measurement error in the exposure induced by Euclidean distances is minimal. At both periods of measurement, the presence of existing and new dedicated PA spaces within buffers was determined, and this allowed us to distinguish the intervention group (no PA space in wave 1, a PA space in wave 2) from the control group (PA space absent at both waves). There is no information available about the actual distance Dutch children walk or cycle to visit a PA space. The buffer size of 600 meters was chosen based on the mean radius of a Rotterdam neighbourhood in 2008 (10). Importantly, the distance to PA spaces differ per individual, and are unrelated to neighbourhood borders. Children in the intervention group lived in 19 neighbourhoods. In 18 of these neighbourhoods some children were included in the intervention group, while other children from the same neighbourhood were included in the control group. It illustrates the often echoed critique on using neighbourhood boundaries, which have little meaning for children when it comes to access to facilities.

Distance to PA spaces using a continuous measure

The effects on playing outdoors are presumably small if children still live far away after the introduction of new play facilities. Indeed, 200 meters closer to a facility might matter more for those who originally lived 600 meters from a facility instead of for example 1600 meters from the nearest dedicated PA space. In a separate analysis, we truncated all distances above 600 meters to 600 meters (resulting in "no change in distance" for those who never lived within 600 meters) and investigated the impact of the absolute change in distance living closer to a PA space within the 600 meters buffer.

Outdoor play and sedentary behaviour

Outdoor play and sedentary behaviour were assessed by parent-reports. Outdoor play concerned exercise at school and outside school hours for an average week. At the age of 6 years, frequency (number of days) and duration (never, less than 30 minutes, 30-60 minutes, 1-2 hours, 2-3 hours, 3-4 hours) were asked for weekdays and weekend days separately. The mid-point of each category (e.g., 45 minutes for 30-60 minutes) was used to estimate the duration of a session. The frequency was multiplied by duration, and estimates for weekdays and weekend days were summed to obtain the average time spent playing outdoors in minutes per week at the age of 6 years. The same procedure was used at the age of 10 years, although answer options for frequency and duration slightly differed, and did not specify for weekdays and weekend days. Appendix 1 includes the questionnaires.

Sedentary behaviour was assessed through two separate questions on television viewing and computer game use for an average week. Questions included frequency and duration, for weekdays and weekend days separately. Again, the frequency was multiplied by the duration to calculate the average time spent television viewing and computer gaming. Time spent television viewing and computer gaming was summed to obtain the average time spent engaging in sedentary behaviours in minutes per week at the age of 6 and 10 years.

Physical activity behaviours

Other physical activity behaviours assessed by parent-reports at age 6 and 10 included the number of sport activities in which a child participated, and the number of days that children walked or cycled to school.

Socio-demographic variables

Parent self-reported highest obtained maternal and paternal education level at the child's age of 6 years were categorized according to the Dutch Standard Classification of Education into high (university degree), mid-high (higher vocational training, bachelor's degree), mid-low (>3 years general secondary school, intermediate vocational training), and low (no education, primary school, lower vocational training, intermediate general school, or ≤3 years general secondary school) (11). Information about maternal and paternal employment status (paid job, no paid job) was collected when the child was 6 years, whereas net household income (≤€2000/month, <€2000-€3200/month, >€3200/month) was collected at both time points. In accordance with Statistics Netherlands, a child's ethnic background was classified as native Dutch, other-Western background (countries in Europe, North-America and Oceania), and non-Western background (countries in Africa, Latin-America and Asia, including Turkey) based on the country of birth of the child's parents (12).

Statistical analyses

Baseline characteristics of the study population were presented as means with standard deviations (SDs) for normally distributed variables, medians with inter quartile range (IQR) for skewed variables, and percentages for categorical variables.

First, multi-variable linear regression models were constructed to evaluate the cross-sectional association between exposure to a PA space within 600 meter from home and outdoor play and sedentary behaviours at the age of 10 years. Subsequently, fixed-effects regression models were applied, which allow to control for measured time-variant and for unmeasured time-invariant confounders (13-15). As time-varying confounders, net household income, season of data collection, sport participation, and active transport to school were controlled for. If net household income, sport participation or active transport to school was missing at age 6, the value measured at age 10 was imputed (n=73 for income; n=32 for sport participation; n=69 for active transport to school), and *vice versa* (n=67 for income; n=25 for sport participation; n=13 for active transport to school). The following regression model was used:

$$y_{it} = \alpha_t + \beta_1 \, (intervention_{it}) + \beta_2 \, (x_{it}) + \mu_i + \varepsilon_{it}$$

Where y_{it} is the dependent variable of interest (e.g., outdoor play and sedentary behaviour) for individual i on time t, β_1 is the effect of the intervention for individual i on time t, β_2 is the effect of time-varying factors for individual i on time t, α_t accounts for time effects that are constant across individuals, μ_i accounts for time-invariant random errors on the individual-level, and ε_{it} accounts for normal sources of error that vary across individuals and time.

Five sensitivity analyses were conducted. First, we repeated the analyses using buffers of 400 and 800 meters to explore if larger effects were found for children living closer to the nearest PA space thereby receiving the highest exposure. Second, we included children who moved houses during follow-up, to evaluate if selective migration took place. Third, we excluded children for which the data were collected within 6 months after introduction of the new facility, to account for the novelty effect and assure that long term impact is obtained. Fourth, we used the median exposure time (the time between opening of the PA space and second measurement; 1.9 years) to stratify the sample into lower and higher exposure time, and assessed if this would influence the change in physical activity behaviours. Fifth, to be able to compare outcomes in the same group of children, we excluded children with missing data.

The dedicated PA spaces specifically target deprived neighbourhoods, which host relatively more persons from lower socioeconomic and ethnic minority groups. Interaction terms were introduced to assess a differential impact of new PA spaces for the intervention and control group by maternal and paternal education level, net household income and ethnicity. To retain statistical power, parental education level was dichotomized into higher (high, mid-high) and lower education level (mid-low, low). Likewise, net household income was dichotomized into higher (>€3200/month) and lower than average Dutch net household income (≤€3200/month). All analyses were conducted in R version 3.4.1, using the plm package for the fixed-effects analyses. Clustered sandwich estimators were used to allow for within-child correlation between error terms. Two-sided P-values<0.05 were considered statistically significant. Interactions were explored for P-for-interactions<0.10.

RESULTS

The intervention group consisted of 171 children who gained access to dedicated PA spaces within 600 meters of their home during follow-up. The children in the intervention group were more often of non-Western ethnicity, less often participated in sports, played ~1 hour/week more outdoors, and were more often from families with lower parental education level and lower net household income, as compared to children in the control group (Table 1). Children in the intervention group were exposed to 13 different PA spaces with, on average, 13 children (range: 1-55) being exposed per PA space.

Cross-sectional analyses at the age of 10 years showed that children in the intervention group played 40 min/week (95% CI: -6, 87) more outside as compared to children in the control group (Supplemental Table 1). For children from families with lower maternal education level, outdoor play was 96 min/week (95% CI: 18, 174) higher as compared to children in the control group. The difference in sedentary behaviour was 78 min/week (95% CI: -23, 179), and 101 min/week (95% CI: -86, 288), respectively.

Table 1: Characteristics of the study population (n=1,841) at age 6 (2008-2012) and 10 (2012-2015)

	Age 6			Age 10		
	Intervention group (n=171)	Control group (n=1,670)	P-value[a]	Intervention group (n=171)	Control group (n=1,670)	P-value
Child characteristics						
Age, years	6.0 ± 0.4[b]	6.0 ± 0.4	0.86	9.7 ± 0.2	9.7 ± 0.3	0.040
Sex, n (%)			0.10			
Girls	96 (56.1)	822 (49.2)				
Boys	75 (43.9)	848 (50.8)				
Ethnic background, n (%)			0.017			
Dutch	103 (60.2)	1,164 (69.7)				
Other-Western	22 (12.9)	200 (12.0)				
Non-Western	46 (26.9)	305 (18.3)				
Sport participation, n (%)[c]			0.088			0.68
No sport	103 (61.7)	857 (52.2)		19 (11.2)	169 (10.3)	
1 sport	54 (32.3)	643 (39.2)		94 (55.3)	848 (51.5)	
2 sports	10 (6.0)	127 (7.7)		47 (27.6)	528 (32.1)	
3 sports	0 (0.0)	15 (0.9)		10 (5.9)	101 (6.1)	
Outdoor play, min/week	693 ± 461	624 ± 418	0.051	404 ± 332	379 ± 320	0.34
Sedentary behaviour, min/week	699 ± 443	660 ± 487	0.35	1107 ± 763	1024 ± 689	0.16
Active transport to school, days/week[c]	3 (1-5)[d]	4 (1-5)	0.067	5 (4-5)	5 (4-5)	0.030
Family characteristics						
Maternal education level, n (%)			0.020			
High	39 (23.1)	576 (34.7)				
Mid-High	52 (30.8)	465 (28.0)				
Mid-Low	61 (36.1)	475 (28.6)				
Low	17 (10.1)	142 (8.6)				

Table 1: (Continued) Characteristics of the study population (n=1,841) at age 6 (2008-2012) and 10 (2012-2015)

	Age 6			Age 10		
	Intervention group (n=171)	Control group (n=1,670)	P-value[a]	Intervention group (n=171)	Control group (n=1,670)	P-value
Paternal education level, n (%)			0.006			
High	40 (24.8)	595 (38.5)				
Mid-High	47 (29.2)	370 (23.9)				
Mid-Low	48 (29.8)	406 (26.3)				
Low	26 (16.1)	175 (11.3)				
Maternal employment status, n (%)			0.73			
Paid job	130 (82.3)	1,297 (80.8)				
No paid job	28 (17.7)	308 (19.2)				
Paternal employment status, n (%)			0.44			
Paid job	152 (95.0)	1,454 (93.0)				
No paid job	8 (5.0)	109 (7.0)				
Net household income, n (%)[c]			0.048			0.39
≤€2000/month	29 (17.5)	247 (15.4)		32 (19.5)	251 (15.7)	
>€2000-€3200/month	57 (34.3)	428 (26.7)		38 (23.2)	359 (22.4)	
>€3200/month	80 (48.2)	927 (57.9)		94 (57.3)	991 (61.9)	

[a] P-values for differences between intervention and control group were obtained from ANOVA (normally distributed continuous variables), Kruskal-Wallis test (skewed continuous variables), or Chi-square test (categorical variables).
[b] Mean ± SD (all such values).
[c] Not-imputed values.
[d] Median; IQR in parentheses (all such values).

Buffers to assign exposure

The introduction of a dedicated PA space within 600 meters from home between the age of 6 and 10 years had no effect on outdoor play (-25 min/week (95% CI: -101, 51 min/week)) or sedentary behaviours (55 min/week (95% CI: -57, 167 min/week)), when controlling for the average decline in outdoor play, and increase in sedentary behaviours in the population (Table 2). The time-varying factors in the model showed that outdoor play was lower during autumn and winter, but did not differ by income, sport participation, or active transport to school. Sedentary behaviours was not determined by any of the time-varying factors.

Absolute distance to assign exposure

Findings were similar when using distance as continuous variable instead of dichotomous buffer size. Decreasing the distance to a dedicated PA space by 100 meters had no effect on outdoor play (-3 min/week (95% CI: -31, 25 min/week)), or sedentary behaviours (42 min/week (95% CI: -16, 99 min/week) (Table 2).

Sensitivity analyses

Using alternative buffer sizes, including children who moved houses, excluding children for which data was collected within 6 months after opening of a PA space, stratifying the analyses for children with lower and higher exposure time, and excluding children that had only data for outdoor play or sedentary behaviours, yielded essentially similar results (Supplemental Table 2).

Subgroup analyses

Stratified analyses were performed for all indicators of family socioeconomic position that were considered significant (P-for-interaction<0.10). Although estimates for each stratum did not reach statistical significance, a consistent pattern was found, suggesting that reducing the distance to the nearest PA space increased outdoor play for children from families with lower maternal or paternal education level (Figure 2; Supplemental Table 3). However, they also showed a non-significant increase in sedentary behaviours for children from families with net household income below average Dutch income, and for children from a non-Dutch ethnicity (Figure 2; Supplemental Table 4).

Table 2: The effect of the introduction of PA spaces on changes in outdoor play and sedentary behaviour

	Outdoor play (min/week, 95% CI)		Sedentary behaviour (min/week, 95% CI)	
Intervention / control (n)	(152 / 1,455)		(133 / 1,412)	
	Buffer <600 meter	Per 100 meter	Buffer <600 meter	Per 100 meter
Exposure to PA spaces	-25 (-101, 51)	-3 (-31, 25)	55 (-57, 167)	42 (-16, 99)
Time	-211 (-241, -181)	-213 (-242, -183)	394 (353, 434)	392 (352, 432)
Net household income				
≤€2000/month	-	-	-	-
>€2000-€3200/month	41 (-37, 120)	41 (-37, 120)	-75 (-187, 37)	-78 (-190, 35)
>€3200/month	63 (-27, 153)	63 (-27, 153)	-83 (-211, 44)	-87 (-215, 41)
Season of data collection				
Spring	-	-	-	-
Summer	29 (-13, 72)	29 (-13, 71)	-24 (-82, 35)	-24 (-82, 35)
Autumn	-154 (-197, -111)	-155 (-198, -112)	-54 (-114, 5)	-55 (-114, 5)
Winter	-252 (-293, -212)	-253 (-294, -212)	39 (-17, 96)	39 (-18, 96)
Active transport to school (days/week)	-5 (-16, 5)	-5 (-16, 5)	-4 (-19, 11)	-4 (-19, 11)
Sport participation (number of sports)	-22 (-47, 2)	-23 (-47, 2)	-26 (-59, 7)	-27 (-60, 6)

Values (95% confidence intervals) indicate changes in outdoor play time or sedentary behaviour time (in minutes per week) for children who were living closer to the nearest dedicated PA space, as compared to children without changes in distance.

Children from socioeconomically disadvantaged families

We explored if larger effects were found when using smaller buffer sizes for children from families with lower maternal education level. New PA spaces within 400 meters of home increased outdoor play non-significantly by 78 min/week (95% CI: -70, 226 min/week), within 600 meter by 45 min/week (95% CI: -72, 163 min/week), and within 800 meters by 25 min/week (95% CI: -86, 137 min/week) (Supplemental Table 5). To the contrary, no such exposure-response relationship was found for sedentary behaviours.

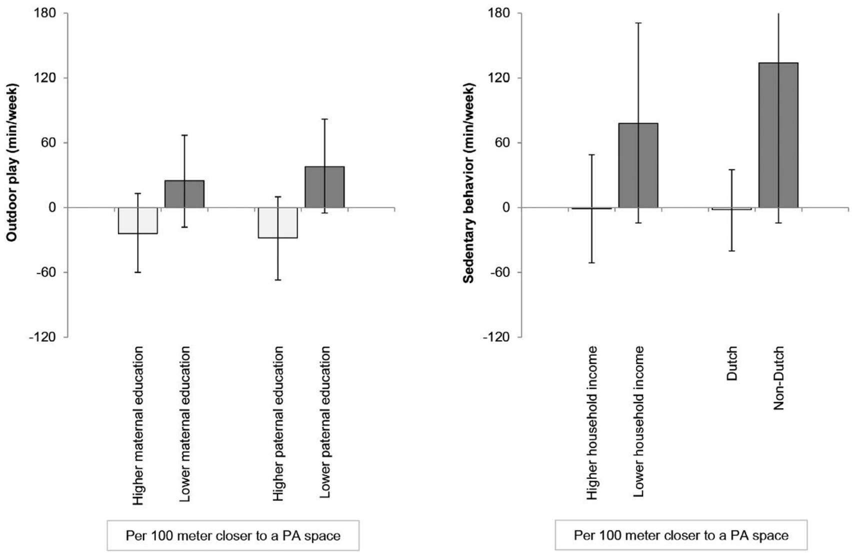


Figure 2: The effect of the introduction of PA spaces on changes in physical activity behaviours, stratified by indicators of family socioeconomic position

DISCUSSION

Children in the intervention group were more often from socioeconomically disadvantaged families, and had higher baseline levels of outdoor play, as compared to children in the control group. We found that the introduction of dedicated PA spaces within 600 meters of home in deprived areas in Rotterdam did not affect changes in outdoor play or sedentary behaviours between the age of 6 and 10 years. Although not statistically significant, stratified analyses suggested that the change in outdoor play was largest for children with parental education level up to 3 years general secondary school, or with intermediate vocational training. We also found that living closer to PA spaces non-significantly increased sedentary behaviours for children from families with net household income up to €3200/month (close to average net household income in the Netherlands), and for children from a non-Dutch ethnicity.

Strengths and limitations

Our study has several strengths. First, this study was strengthened by using a study design and selecting a population that closely mimics the situation that would have been appropriate if the study was conducted in an experimental setting, as recently proposed by Hernán et al. (9). Consequently, we only included children that did not live close to a PA space at baseline, and estimated the effect of the introduction of new facilities on physical activity behaviours within a group of children that did not move houses. The association between the intervention and outcomes reported in this study approximate the effect estimates that would have been obtained if randomization procedures would have allocated intervention and control status to children.

Second, the latter is strengthened by our fixed-effects analysis. Previously reported associations between outdoor play facilities and physical activity may be confounded. Perhaps most prominently, parental attitudes towards children's health may both determine living at a place where children can play outside and their physical activity behaviour. Clearly, even though we were able to control for such unobserved time-invariant confounding, time-variant factors may cause confounding (as it does using other methods).

Third, our measure of exposure was determined at the individual-level. Some children were allocated to the intervention group, while others living in the same neighbourhood were allocated to the control group. Despite the fact that the intervention was implemented in neighbourhoods, random-effect multilevel models seemed less appropriate; we do not expect a clustering of outcomes within neighbourhoods.

The main limitation of this study is the lack of power. In order to detect a 1 hour/week (SD: 7) difference in outdoor play between the intervention and control group with a power of 0.80, we would have needed around 400 children in each group. When conceptualizing the study, we were confident that we could include a sufficient number of children in the intervention group. During the study period, 18 dedicated PA spaces were built spread across the city. A 4-year difference between the first and second measure of physical activity behaviours seemed to be sufficiently large for children to experience a change in distance to the nearest dedicated PA space, but small enough to be able to attribute the change in outdoor play to the intervention. Unfortunately, we had to exclude nearly a fifth of the study population that were already living within 600 meters of a PA space at baseline, and the number of children being exposed to a new PA space during follow-up was much smaller than expected.

Another major limitation is that we had no GPS measurements available, and therefore could not identify whether children's physical activity occurred at the PA spaces. The finding of larger effect sizes when using smaller buffers in the group which seemed to be most responsive to the intervention provides some support that the impact of PA spaces was measured. Spatial and temporal certainty is needed to strengthen the evidence of built environment interventions (16, 17). We had to rely on parent reports, which may be prone to recall and social desirability biases, and did not allow to assess the level of physical activity intensity. For 93% of the children the same parent filled in the questionnaire twice, thereby offsetting some of the biases that may have occurred if different parents filled in the questionnaire. Similarly, the answer categories slightly changed over time. The questionnaire used at the age of 6 years better captured outdoor play for children that played frequently outside. We excluded children with unrealistic high levels of outdoor play, nevertheless, the major decline in outdoor play time may have been partly resulted from the change in questionnaire items. However, when evaluating the introduction of PA spaces, this is less of a problem since only small differences were seen in time spent playing outdoors at baseline between intervention and control group. The variety of activities that can be performed at the PA spaces, and the density of physical activity programs, may be of importance to stimulate behavioural change. Power issues, and incomplete information on the programming on the fields, did not allow to further explore this.

Children in the intervention group spent more time playing outdoors at baseline, and stratified analyses on baseline levels of outdoor play showed that children with low levels of outdoor play had largest increase in the time spent playing outdoors (results not shown). This may have resulted from the statistical phenomena regression to the mean, whereby children with extreme levels during the first measurement round fall back to the mean when measured a second time. Another explanation for the increase in outdoor play is that children with low baseline levels of physical activity are more responsive to the intervention. The effect found in the mid-tertile was not essentially different from the results in the main analyses. Careful examination of baseline characteristics is warranted when evaluating absolute changes over time using fixed-effects models.

When analysed in a cross-sectional way, we found that children at the age of 10 years with a PA space nearby played 0.5 hours/week more outside as compared to children without dedicated PA spaces around home. For children from families with lower maternal education level, outdoor play was 1.5 hours/week higher. These estimates are larger then found in the (natural) experimental setting, suggesting that both selection and causation mechanisms may explain the relationship between access to play facilities and physical activity.

An earlier study showed that Krajicek playgrounds attract more children and that their physical activity intensity is higher compared to children playing at regular PA spaces (18). Although not statistically significant, our study suggested some increase in time spent playing outdoors for children from more disadvantaged families, thus outdoor play did not shift only from other locations to dedicated PA spaces. Playing more outside at higher intensities could have a beneficial impact on children's health (19).

We found some evidence that sedentary behaviours increased following the introduction of PA spaces. A meta-analyses also showed that an increase physical activity does not necessarily reduce sedentary behaviours (20). A possible explanation is that children compensate their physical activity and sedentary behaviours between days (21-23). Thus, being active at one time will result in a decrease in physical activity and an increase in sedentary behaviours at another time. For example, compensation occurs when parents reward their children for playing outside by allowing screen time. Further research is needed to examine the potential compensation mechanisms following interventions to promote physical activity behaviours.

The introduction of dedicated PA spaces was confined to deprived neighbourhoods, however, the absolute level of deprivation varied across neighbourhoods. Previous work described Dutch deprived neighbourhoods as neighbourhoods "(...) with problems regarding employment, education, housing and the physical neighbourhood environment, social cohesion, and safety" (24). For the city of Rotterdam it was estimated that life expectancy differed by 7 years between neighbourhoods, and healthy life expectancy by 14 years (25). Segregation is not as present in Dutch cities as compared to some other countries, and higher income families do reside in deprived neighbourhoods. Therefore, the introduction of dedicated PA spaces was not restricted to lower income families only. Nearly half of the children in the intervention group were from families with net household income above Dutch average. The overrepresentation of children from higher socioeconomic groups in the Generation R study may have contributed to this finding.

The introduction of PA spaces in deprived neighbourhoods may result in, or come together with, more general neighbourhood changes that could affect outdoor play. We are not aware of any structural interventions or neighbourhood improvements implemented in neighbourhoods where new play facilities were built, but this could have biased our results. For example, road connectivity and neighbourhood safety could have been targeted by other programs in neighbourhoods where new PA spaces were introduced. This may have an effect on outdoor play directly, or indirectly by mediating factors such as increasing children's independent mobility (26-28). Further insights in mechanisms following built environment changes are needed to better understand subsequent behavioural changes (29). Strategies to increase independent mobility should be encouraged, since this may largely influence the usage of neighbourhood facilities.

To reduce health inequities in society, it is of great importance to identify strategies that improve health behaviours for those that are most at risk for developing disease later in life. In our main analyses, we did not find evidence that the introduction of PA spaces changed physical activity behaviours. Stratified analyses suggested that children from more socially disadvantaged families appear most responsive to the introduction of PA spaces, and there is possibly an increase in outdoor play and sedentary behaviours for these groups relative to more advantaged families. Municipalities should carefully verify if neighbourhoods applying for a PA space are likely to benefit from the introduction of these facilities. Negative aspects may also result from the introduction of PA spaces. Earlier research showed that dedicated PA spaces had a small but negative impact on social cohesion, violence, and perceived safety (30). The authors suggested that the PA spaces may attract various people, of which some may cause nuisance. This should be carefully monitored when implementing dedicated PA spaces.

It was encouraging to see a possible increase in outdoor play among these families, since these children more often do not participate in sports activities (31). Interventions often lack evidence concerning the equity effects of promoting physical activity in children (32). The presentation of subgroup estimates is not always appreciated, especially because they are prone to statistical malpractice (33). In our study, the intervention specifically target children living in more deprived neighbourhoods, which justifies the decision to present subgroup estimates for population characteristics that are associated with neighbourhood deprivation.

Researchers and policymakers often struggle how to evaluate the effectiveness of interventions within the built environment. There is a need to conduct more consequential research that informs policymakers how to improve population health (34, 35). Focusing on children in which the access to facilities changed during follow-up, as opposed to cross-sectional studies and studies in which children change neighbourhoods, is essential for this purpose. This study showed that natural experiments can be used for policy evaluation, however, finding a setting in which a substantial part of the population is experiencing differences in exposure is challenging. New technologies, such as GPS and wearable devices to objectively measure physical activity with smaller variance, and data linkage may further improve studies on the effect on environmental changes in physical activity and underlying mechanisms. We would encourage policymakers and researchers to look for relevant natural experiments within the built environment that may have contributed to population health. As such, investing in high quality measures and a good documentation of built environment changes would help in creating evidence-based cities in which health behaviours are promoted.

CONCLUSION

The introduction of dedicated PA spaces may increase the time spent playing outdoors for children from more socioeconomically disadvantaged families. Also increases in sedentary behaviours were observed, suggesting that compensation may have taken place. Larger studies unravelling the complexity of child behaviour are needed to design environments that support physical activity behaviours.

REFERENCES

1. Lobstein T, Baur L, Uauy R. Obesity in children and young people: a crisis in public health. Obesity Reviews. 2004;5:4-85.
2. Ebbeling CB, Pawlak DB, Ludwig DS. Childhood obesity: public-health crisis, common sense cure. The Lancet.360(9331):473-82.
3. World Health Organization. Population-based approaches to childhood obesity prevention. Geneva: World Health Organisation, 2012.
4. Committee on Environmental Health. The Built Environment: Designing Communities to Promote Physical Activity in Children. Pediatrics. 2009;123(6):1591-8.
5. Ding D, Sallis JF, Kerr J, Lee S, Rosenberg DE. Neighborhood Environment and Physical Activity Among Youth: A Review. American Journal of Preventive Medicine. 2011;41(4):442-55.
6. Timperio A, Reid J, Veitch J. Playability: Built and Social Environment Features That Promote Physical Activity Within Children. Current Obesity Reports. 2015;4(4):460-76.
7. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health. 2012;66(12):1182-6.
8. Kooijman MN, Kruithof CJ, van Duijn CM, Duijts L, Franco OH, van Ijzendoorn MH, et al. The Generation R study: design and cohort update 2017. European Journal of Epidemiology. 2016;31(12):1243-64.
9. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. American Journal of Epidemiology. 2016;183(8):758-64.
10. Statistics Netherlands. Kerncijfers wijken en buurten 2004-2008. Den Haag/Heerlen.
11. Statistics Netherlands. Standaard Onderwijsindeling 2003. Voorburg/Heerlen 2004.
12. Statistics Netherlands. Jaarrapport Integratie 2010. Den Haag/Heerlen 2010.
13. Allison PD. Using Panel Data to Estimate the Effects of Events. Sociological Methods & Research. 1994;23(2):174-99.
14. Allison PD. Fixed effects regression models: SAGE publications; 2009.
15. Cousens S, Hargreaves J, Bonell C, Armstrong B, Thomas J, Kirkwood BR, et al. Alternatives to randomisation in the evaluation of public-health interventions: statistical analysis and causal inference. Journal of Epidemiology and Community Health. 2011;65(7):576-81.

16. Dunton GF, Almanza E, Jerrett M, Wolch J, Pentz MA. Neighborhood Park Use by Children: Use of Accelerometry and Global Positioning Systems. American Journal of Preventive Medicine. 2014;46(2):136-42.

17. Humphreys DK, Panter J, Sahlqvist S, Goodman A, Ogilvie D. Changing the environment to improve population health: a framework for considering exposure in natural experimental studies. Journal of Epidemiology and Community Health. 2016;70(9):941-6.

18. Boonzajer Flaes SAM, Chinapaw MJM, Koolhaas CM, van Mechelen W, Verhagen EALM. More children more active: Tailored playgrounds positively affect physical activity levels amongst youth. Journal of Science and Medicine in Sport. 2016;19(3):250-4.

19. Ekelund U, Luan J, Sherar LB, et al. Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents. JAMA. 2012;307(7):704-12.

20. Pearson N, Braithwaite RE, Biddle SJH, van Sluijs EMF, Atkin AJ. Associations between sedentary behaviour and physical activity in children and adolescents: a meta-analysis. Obesity Reviews. 2014;15(8):666-75.

21. Rowland TW. The biological basis of physical activity. Medicine & Science in Sports & Exercise. 1998;30(3):392-9.

22. Frémeaux AE, Mallam KM, Metcalf BS, Hosking J, Voss LD, Wilkin TJ. The impact of school-time activity on total physical activity: the activitystat hypothesis (EarlyBird 46). International Journal Of Obesity. 2011;35:1277.

23. Ridgers ND, Timperio A, Cerin E, Salmon JO. Compensation of Physical Activity and Sedentary Time in Primary School Children. Medicine & Science in Sports & Exercise. 2014;46(8):1564-9.

24. Droomers M, Harting J, Jongeneel-Grimen B, Rutten L, van Kats J, Stronks K. Area-based interventions to ameliorate deprived Dutch neighborhoods in practice: Does the Dutch District Approach address the social determinants of health to such an extent that future health impacts may be expected? Preventive Medicine. 2014;61:122-7.

25. Jonker MF, van Lenthe FJ, Donkers B, Mackenbach JP, Burdorf A. The effect of urban green on small-area (healthy) life expectancy. Journal of Epidemiology and Community Health. 2014;68(10):999-1002.

26. Veitch J, Bagley S, Ball K, Salmon J. Where do children usually play? A qualitative study of parents' perceptions of influences on children's active free-play. Health & Place. 2006;12(4):383-93.

27. Veitch J, Salmon J, Ball K. Children's active free play in local neighborhoods: a behavioural mapping study. Health Education Research. 2008;23(5):870-9.

28. Moran MR, Plaut P, Merom D. Is the Grass Always Greener in Suburban Neighborhoods? Outdoors Play in Suburban and Inner-City Neighborhoods. International Journal of Environmental Research and Public Health. 2017;14(7):759.

29. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. Lancet. 2017; 9;390(10112):2602-2604.

30. Wittebrood K, Permentier M, Pinkster F. Housing, neighbourhoods and interventions. The 'empowered neighbourhoods' policy in perspective. Sociaal en Cultureel Planbureau, 2011. Report No.: 9789037700657.

31. Wijtzes AI, Jansen W, Bouthoorn SH, Pot N, Hofman A, Jaddoe VWV, et al. Social inequalities in young children's sports participation and outdoor play. The International Journal of Behavioral Nutrition and Physical Activity. 2014;11:155.

32. Love RE, Adams J, van Sluijs EMF. Equity effects of children's physical activity interventions: a systematic scoping review. International Journal of Behavioral Nutrition and Physical Activity. 2017;14(1):134.

33. Petticrew M, Tugwell P, Kristjansson E, Oliver S, Ueffing E, Welch V. Damned if you do, damned if you don't: subgroup analysis and equity. Journal of Epidemiology and Community Health. 2012;66(1):95-8.

34. Galea S. An Argument for a Consequentialist Epidemiology. American Journal of Epidemiology. 2013;178(8):1185-91.

35. Nandi A, Harper S. How Consequential Is Social Epidemiology? A Review of Recent Evidence. Current Epidemiology Reports. 2015;2(1):61-70.

SUPPLEMENTAL MATERIAL

Supplemental Figure 1	Impression of dedicated PA spaces in Rotterdam, the Netherlands
Supplemental Table 1	Cross-sectional analyses between exposure to PA spaces <600 meters and outdoor play and sedentary behaviours at the age of 10 years
Supplemental Table 2	Alternative comparison groups to evaluate the introduction of PA spaces on changes in outdoor play and sedentary behaviour
Supplemental Table 3	The effect of the introduction of PA spaces on changes in outdoor play, stratified by indicators of family socioeconomic position
Supplemental Table 4	The effect of the introduction of PA spaces on changes in sedentary behaviour, stratified by indicators of family socioeconomic position
Supplemental Table 5	Alternative comparison groups to evaluate the introduction of PA spaces on changes in outdoor play and sedentary behaviour for children from families with lower maternal education level
Appendix 1	Questionnaire items outdoor play and sedentary behaviour

Supplemental Figure 1: Impression of dedicated PA spaces in Rotterdam, the Netherlands[a]

[a] Received from the municipality of Rotterdam. Reproduced with permission.

Supplemental Table 1: Cross-sectional analyses between exposure to PA spaces <600 meters and outdoor play and sedentary behaviour at the age of 10 years

	Outdoor play		Sedentary behaviour	
	Intervention / control (n)	min/week (95% CI)	Intervention / control (n)	min/week (95% CI)
Whole group	194 / 1,606	40 (-6, 87)	177 / 1,550	78 (-23, 179)
Lower maternal education	81 / 581	96 (18, 174)	79 / 574	101 (-86, 288)
Higher maternal education	113 / 1,025	4 (-54, 61)	98 / 976	54 (-55, 164)

Values (95% confidence intervals) indicate the difference in outdoor play time or sedentary behaviour time (in minutes per week) for children who were living <600 meters from the nearest dedicated PA space, as compared to children living ≥600 meters. Estimates were obtained from multivariable linear regression models, adjusted for sex, ethnicity, net household income, season of data collection, sport participation, and active transport to school.

Supplemental Table 2: Alternative comparison groups to evaluate the introduction of PA spaces on changes in outdoor play and sedentary behaviour

	<400 meter	<600 meter	<800 meter	Per 100 meter
Outdoor play (min/week)				
Main analyses	11 (-85, 107)	-25 (-101, 51)	-19 (-82, 45)	-3 (-31, 25)
With movers[a]	21 (-59, 101)	21 (-45, 86)	20 (-38, 78)	9 (-16, 33)
Novelty effect[b]	21 (-80, 121)	-15 (-94, 65)	-15 (-82, 52)	2 (-27, 32)
Higher exposure time[c]	51 (-83, 186)	-35 (-129, 59)	-51 (-134, 33)	5 (-38, 47)
Lower exposure time[d]	-47 (-169, 76)	-11 (-130, 108)	20 (-70, 109)	-11 (-46, 25)
Complete cases[e]	35 (-66, 137)	-12 (-99, 75)	-1 (-72, 70)	-1 (-33, 32)
Sedentary behaviour (min/week)				
Main analyses	57 (-103, 216)	55 (-57, 167)	58 (-42, 158)	42 (-16, 100)
With movers[a]	79 (-79, 237)	46 (-46, 138)	48 (-38, 134)	32 (-18, 81)
Novelty effect[b]	49 (-117, 214)	61 (-52, 174)	90 (-15, 195)	45 (-17, 106)
Higher exposure time[c]	35 (-215, 285)	66 (-92, 224)	63 (-64, 190)	67 (-33, 166)
Lower exposure time[d]	83 (-91, 257)	44 (-107, 195)	55 (-94, 204)	16 (-36, 68)
Complete cases[e]	55 (-122, 232)	93 (-29, 215)	39 (-56, 135)	52 (-15, 120)

Values (95% confidence intervals) indicate changes in outdoor play time or sedentary behaviour time (in minutes per week) for children who were living closer to the nearest dedicated PA space, as compared to children without changes in distance. The model was adjusted for the time-varying factors net household income, season of data collection, sport participation, and active transport to school.

[a] Including children who moved houses during follow-up.

[b] Excluding children of which the data was collected within 6 months after being exposed to a PA space.

[c] Excluding children with time between opening of the PA space and the second measurement ≤1.9 year.

[d] Excluding children with time between opening of the PA space and the second measurement >1.9 year.

[e] Restricted to children with complete data on outdoor play and sedentary behaviour.

Supplemental Table 3: The effect of the introduction of PA spaces on changes in outdoor play, stratified by indicators of family socioeconomic position

Subgroup	n	Buffer <600 meter	P-for-interaction	Per 100 meter	P-for-interaction
Maternal education level			0.16		0.089
Higher	1,007	-68 (-169, 33)		-24 (-60, 13)	
Lower	587	45 (-72, 163)		25 (-18, 67)	
Paternal education level			0.082		0.011
Higher	934	-76 (-186, 35)		-28 (-67, 10)	
Lower	561	44 (-69, 157)		38 (-5, 82)	
Net household income			0.90		0.82
Higher	924	-22 (-122, 78)		-3 (-45, 38)	
Lower	683	-24 (-138, 90)		-1 (-39, 37)	
Ethnicity			0.50		0.66
Dutch	1,132	-26 (-118, 67)		-1 (-37, 35)	
Non-Dutch	475	-43 (-179, 93)		-11 (-58, 35)	

Values (95% confidence intervals) indicate changes in outdoor play time (in minutes per week) for children who were living closer to the nearest dedicated PA space, as compared to children without changes in distance, stratified by indicators of family socio-economic position. The model was adjusted for the time-varying factors net household income, season of data collection, sport participation, and active transport to school.

Supplemental Table 4: The effect of the introduction of PA spaces on changes in sedentary behaviour, stratified by indicators of family socioeconomic position

Subgroup	n	Buffer <600 meter	P-for-interaction	Per 100 meter	P-for-interaction
Maternal education level			0.075		0.15
Higher	954	5 (-106, 116)		17 (-24, 58)	
Lower	581	104 (-98, 305)		68 (-45, 181)	
Paternal education level			0.50		0.92
Higher	891	39 (-81, 158)		27 (-21, 75)	
Lower	545	26 (-145, 197)		0 (-61, 60)	
Net household income			0.061		0.066
Higher	897	-11 (-132, 110)		-1 (-51, 49)	
Lower	647	137 (-54, 328)		78 (-14, 171)	
Ethnicity			0.054		0.046
Dutch	1,099	-22 (-118, 73)		-2 (-40, 35)	
Non-Dutch	444	202 (-66, 470)		134 (-14, 282)	

Values (95% confidence intervals) indicate changes in sedentary behaviour time (in minutes per week) for children who were living closer to the nearest dedicated PA space, as compared to children without changes in distance, stratified by indicators of family socio-economic position. The model was adjusted for the time-varying factors net household income, season of data collection, sport participation, and active transport to school.

Supplemental Table 5: Alternative comparison groups to evaluate the introduction of PA spaces on changes in outdoor play and sedentary behaviour for children from families with lower maternal education level

	<400 meter	<600 meter	<800 meter	Per 100 meter
Outdoor play (min/week)				
Main analyses	78 (-70, 226)	45 (-72, 163)	25 (-86, 137)	25 (-18, 67)
With movers[a]	59 (-64, 181)	89 (-17, 195)	69 (-35, 173)	32 (-5, 69)
Novelty effect[b]	71 (-96, 238)	60 (-65, 184)	27 (-93, 146)	28 (-19, 75)
Higher exposure time[c]	71 (-115, 257)	22 (-108, 152)	-10 (-149, 129)	23 (-34, 81)
Lower exposure time[d]	101 (-99, 301)	82 (-122, 285)	71 (-97, 239)	28 (-28, 84)
Complete cases[e]	99 (-68, 265)	27 (-108, 163)	12 (-114, 138)	22 (-31, 76)
Sedentary behaviour (min/week)				
Main analyses	102 (-208, 412)	104 (-98, 305)	151 (-64, 365)	68 (-45, 181)
With movers[a]	81 (-171, 332)	85 (-89, 259)	134 (-52, 319)	46 (-54, 145)
Novelty effect[b]	103 (-234, 439)	130 (-77, 336)	221 (-2, 445)	75 (-50, 199)
Higher exposure time[c]	134 (-293, 562)	103 (-177, 383)	89 (-148, 327)	102 (-60, 263)
Lower exposure time[d]	46 (-336, 427)	109 (-168, 387)	231 (-135, 598)	10 (-110, 129)
Complete cases[e]	181 (-197, 559)	207 (-27, 440)	181 (-29, 391)	116 (-26, 259)

Values (95% confidence intervals) indicate changes in outdoor play time or sedentary behaviour time (in minutes per week) for children who were living closer to the nearest dedicated PA space, as compared to children without changes in distance. The model was adjusted for the time-varying factors net household income, season of data collection, sport participation, and active transport to school.

[a] Including children who moved houses during follow-up.

[b] Excluding children of which the data was collected within 6 months after being exposed to a PA space.

[c] Excluding children with time between opening of the PA space and the second measurement ≤1.9 year.

[d] Excluding children with time between opening of the PA space and the second measurement >1.9 year.

[e] Restricted to children with complete data on outdoor play and sedentary behaviour.

Appendix 1: Questionnaire items outdoor play and sedentary behaviour

Questionnaire outdoor play age 6

The questions below are concerned with the **exercise** you child has at school and outside of school hours. Keep an **average week** in mind when answering these questions and differentiate between weekdays (Mondays through Fridays) and week-ends (Saturdays and Sundays).
If you end up with an average of less than one day per week, mark the square 'never'.

On average, how many **weekdays** per week does your child **play outside**?

Never on weekdays	*1 day per week*	*2 days per week*	*3 days per week*	*4 days per week*	*Every weekday*
☐	☐	☐	☐	☐	☐

On average, how many **week-end days** per week does your child **play outside**?

Never in the week-end	*1 day in the week-end*	*2 days in the week-end*
☐	☐	☐

On the days that your child **plays outside, how long, on average**, does your child then play outside? Differentiate between weekdays and week-ends and answer according to the present season.

	*How long on **weekdays**?*	*How long on **week-ends**?*
Mornings	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Afternoons	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Evenings after dinner	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours

Questionnaire sedentary behaviours age 6

The questions below are concerned with watching television, videos and DVD's. Keep an **average week** in mind when answering these questions and differentiate between weekdays (Mondays through Fridays) and week-ends (Saturdays and Sundays). If you end up with an average of less than one day per week, mark the square 'never'.

—

On average, how many **weekdays** per week does your child watch **television/video/DVD**?

Never on weekdays	*1 day per week*	*2 days per week*	*3 days per week*	*4 days per week*	*Every weekday*
☐	☐	☐	☐	☐	☐

On average, how many **week-end days** per week does your child watch **television/video/DVD**?

Never in the week-end	*1 day in the week-end*	*2 days in the week-end*
☐	☐	☐

On the days that your **child** watches **television/video/DVD**, **how long** does he watch, on average? Differentiate here between weekdays and week-ends.

	*How long on **weekdays**?*	*How long on **week-ends**?*
Mornings	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Afternoons	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Evenings after dinner	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours

The questions below are concerned with **how much your child uses the computer**. By computer use we also mean playstation, gameboy, nintendo, etc. Keep an **average week** in mind when answering these questions and differentiate between weekdays (Mondays through Fridays) and week-ends (Saturdays and Sundays). If you end up with an average of less than one day per week, mark the square 'never'.

On average, how many **weekdays** in the week does your child use the **(game)computer**?

Never on weekdays	*1 day per week*	*2 days per week*	*3 days per week*	*4 days per week*	*Every weekday*
☐	☐	☐	☐	☐	☐

. On average, how many **week-end days** does your child use the **(game)computer**?

Never in the week-end	*1 day in the week-end*	*2 days in the weekend*
☐	☐	☐

On the days that your child uses the **(game)computer** at home, **how long** does he/she use it, on average? Please differentiate here between weekdays and week-ends

	*How long on **weekdays**?*	*How long on **week-ends**?*
Mornings	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Afternoons	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Evenings after dinner	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours

Questionnaire outdoor play age 10

The following questions concern your child's **exercise pattern**. During both school time and leisure time.

On average **how many days per week** does your child play **outside**?

- ☐ Never, continue to question 9
- ☐ 1 or 2 days per week
- ☐ 3 or 4 days per week
- ☐ 5 or more days per week

Approximately **how long** does your child approximately play **outside** per day? Only consider the days that your child plays outside.

- ☐ Less than 30 minutes per day
- ☐ 30 minutes to 1 hour per day
- ☐ 1 to 2 hours per day
- ☐ 2 to 3 hours per day
- ☐ 3 to 4 hours per day
- ☐ More than 4 hours per day

Questionnaire sedentary behaviours age 10

The following questions regard watching television. The answers to these questions should relate to an **average week** and differentiate between workdays (Monday thru Friday) and weekends (Saturday and Sunday). If your calculations result in less than 1 day per week, check the box 'never'.

On average, how many **weekdays** per week does your child watch **television/video/DVD**?

Never on weekdays	*1 day per week*	*2 days per week*	*3 days per week*	*4 days per week*	*Every weekday*
☐	☐	☐	☐	☐	☐

On average, how many **weekend days** per week does your child watch **television/video/DVD**?

Never in the weekend	*1 day in the weekend*	*2 days in the weekend*
☐	☐	☐

On the days that your child watches **television/video/DVD**, **how long** does he/she watch, on average? Differentiate here between weekdays and weekends.

	*How long on **weekdays**?*	*How long on **weekends**?*
Mornings	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Afternoons	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Evenings after dinner	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours

The questions below are concerned with **how much your child uses the (game)computer**. By computer use we also mean Playstation, Game boy, Nintendo, etc. Keep an average week in mind when answering these questions and differentiate between weekdays (Mondays through Fridays) and weekends (Saturdays and Sundays). If you end up with an average of less than one day per week, check the 'never' box.

On average, how many **weekdays** does your child use the **(game)computer**?

Never on weekdays	*1 day per week*	*2 days per week*	*3 days per week*	*4 days per week*	*Every weekday*
☐	☐	☐	☐	☐	☐

On average, how many **weekend days** does your child use the **(game)computer**?

Never in the weekend	*1 day in the weekend*	*2 days in the weekend*
☐	☐	☐

On the days that your child uses the **(game)computer** at home, **how long** does he/she sit there? Differentiate here between weekdays and weekends.

	*How long on **weekdays**?*	*How long on **weekends?***
Mornings	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Afternoons	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours
Evenings after dinner	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours	☐ Never ☐ Less than 30 minutes ☐ 30-60 minutes ☐ 1-2 hours ☐ 2-3 hours ☐ 3-4 hours

3

Socioeconomic inequalities in the food environment and body composition among school-aged children: a fixed-effects analysis

Famke J.M. Mölenberg, Joreintje D. Mackenbach, Maartje P. Poelman, Susana Santos, Alex Burdorf, Frank J. van Lenthe

Submitted

ABSTRACT

Introduction

There is limited evidence regarding socioeconomic inequalities of exposure to the food environment and its contribution to childhood obesity.

Methods

We used data from 4,235 children from the Generation R study, a large birth-cohort conducted in the city of Rotterdam, the Netherlands. We included 11,277 person-observations of body-mass index (BMI) and 6,240 person-observations of DXA-derived fat mass index (FMI) and fat-free mass index (FFMI) when children were between 4 and 14 years. We applied log binomial regression models to evaluate changes in the relative and absolute exposure of fast food outlets, and the healthiness of the food environment within 400 meters from home by maternal education. Furthermore, we used individual-level fixed-effects models to study changes in the food environment to changes in BMI, FMI and FFMI.

Results

At the age of 6 years, 32.1% and 23.1% of the children from lower and higher educated mothers were exposed to at least 4 fast food outlets, respectively. Over a median period of 7 years, the absolute (RR = 1.20 (95% CI: 1.09 to 1.30)) and relative (RR = 1.27 (95% CI: 1.19 to 1.37)) amount of fast food outlets increased more for children from lower as compared to higher educated mothers. Likewise, their food environments became unhealthier (RR = 1.10 (95% CI: 1.03 to 1.18)). Changes in the food environment were not associated with subsequent changes in BMI, FMI and FFMI. For children from lower educated mothers not exposed to fast food at first, we found some evidence that the introduction of fast food was associated with small increases in BMI.

Conclusion

Our findings provide evidence of widening inequalities in an already poor food environment. Access to more fast food outlets does not seem to have an additional impact on BMI in contemporary contexts with ubiquitous fast food outlets.

INTRODUCTION

Childhood obesity is a major public health concern due to its widespread prevalence and rapid increase in the past decades (1, 2). The food environment has also changed considerably, mainly towards a higher exposure of food outlets in residential areas (3), and a higher offer of high-energy and ultra-processed foods (4, 5). It is likely that food environments have contributed to rising childhood obesity. However, evidence for a causal influence of changing food environments on childhood obesity is limited (6-9). Studies have been mainly cross-sectional, and cannot rule out the effects of residential self-selection based on preferences and resources related to both the food environment and obesity. Some studies have offset some of the selection processes by applying fixed-effects models taking into account measured time-varying variables (e.g., income) and unmeasured time-invariant variables (e.g., neighbourhood preference) (10-14). We identified one study in children, and this US study counterintuitively showed that the increase in fast food outlets around home was associated with small reductions in body-mass index (BMI) (10). This finding warrants further exploration in other settings.

Changes in the food environment and their impact on obesity have been hypothesised to differ between individuals from higher and lower socioeconomic groups. A systematic review of 21 studies showed that fast food access was higher in more deprived compared to less deprived areas (15). Individuals from lower socioeconomic position also spent more time in their neighbourhood, thereby being more exposed to the food environment around home (16). Although this may possibly result in a differential impact of the food environment on overweight and obesity across socioeconomic groups, a recent systematic review of mostly cross-sectional studies could not confirm this (17). Longitudinal studies focussing on socioeconomic inequalities in the food environment and related changes in obesity are needed.

The Generation R study is a birth-cohort study in which 4 objective measures of BMI and 2 indices of fat mass obtained from dual-energy x-ray absorptiometry (DXA) scanners were available for children between the age of 4 and 14 years. Linkage with yearly updated food environment measures created the unique possibility to evaluate whether changes in the food environment were associated with changes in body composition. We hypothesised that children with parents of lower socioeconomic position (SEP) lived in neighbourhoods characterised by a higher number of unhealthy food outlets, and that over time, this resulted in unfavourable changes in body composition. First, we evaluated if the exposure to food environment around the home address differentially evolved over time for children from lower and higher educated mothers. Second, we studied if the impact of changes in the food environment on changes in measures of body composition differed between children from lower and higher educated mothers.

METHODS

Study design

This study used data collected from the Generation R study, a prospective birth-cohort study in the city of Rotterdam, in the Netherlands (18). We included objectively determined measures of BMI at the age of 4, 6, 10, and 14 years, and DXA-derived measures of fat mass at the age of 6 and 10 years. At each time-point, home addresses were linked with food environment data of the preceding year. The latter were only available from 2004 onwards, therefore we did not include outcomes collected in preceding years.

Study population

Invitations to participate in the Generation R study were sent out to all pregnant women who had an expected delivery date between April 2002 and January 2006 and who lived in the study area (Rotterdam, the Netherlands) at time of delivery (18). The Medical Ethics Committee of the Erasmus University Medical Centre in Rotterdam approved the study (MEC 217.595/2002/20). Written informed consent was obtained from parents at child ages of 4, 6, and 10 years, and from parents and children at the age of 14 years.

In total, 9,901 children and their parents participated in the Generation R study at baseline. Children with at least two outcome measures that could be linked to the food environment between the ages of 4 years and 14 years (n=5,418) were included. Younger siblings from the same mother (n=397), and children from which information on maternal education level was missing (n=441) were excluded. In total, 4,594 children were eligible for the present study. We included observations for the address where the child lived for the longest period between the age of 4 and 14 years, and children without two consecutive observations on one address (n=359) were excluded. The sample for the main analysis included 4,235 children with 11,277 person-observations for BMI, and 6,240 person-observations for fat mass (Figure 1).

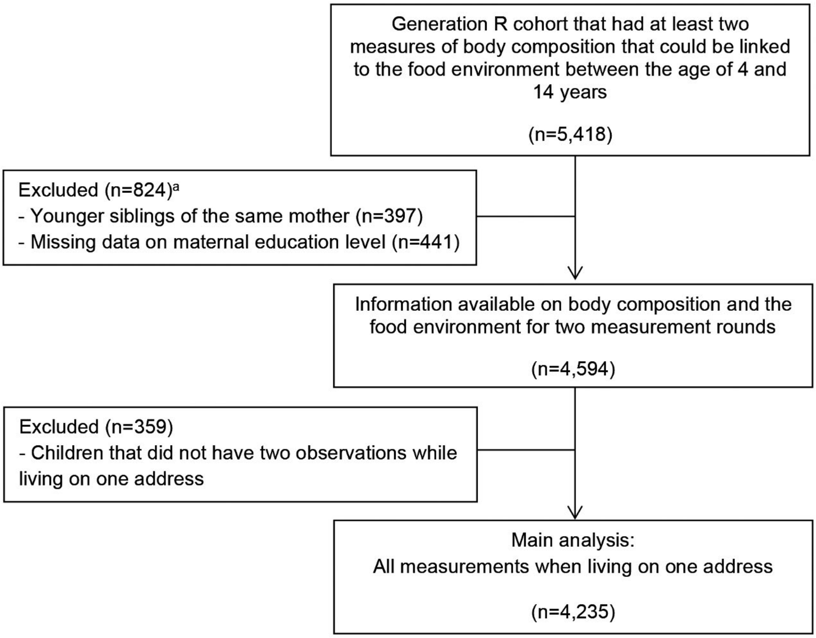


Figure 1: Flow diagram

[a] More than one reason could apply

Food environment

Data on location and type of food outlet were obtained from Locatus, a commercial company that collects information on food retailers in the Netherlands by means of yearly field audits. A validation study using field audit data found an overall "good" to "excellent" agreement for both the location and classification (19). Children's home addresses and the location of food retailers in the year preceding the outcome measure were mapped using a geographic information system (GIS) (GIS-ArcGIS 10, ESRI, Redlands, CA, USA). At each time-point, the food environment was characterized in a Euclidean buffer of 400 meter around the home, representing a walking distance of approximately 5 minutes. Dutch cities are characterised by high connectivity that is supportive of active forms of transportation. Therefore, a distance of 400 meters seems an appropriate distance to assess the food environment in this context.

We considered all fast food outlets, grillrooms/kebab shops, take away outlets, and ice cream shops that were classified by Locatus and available in the database as 'fast food outlets'. The 'absolute fast food exposure' reflects the total number of fast food outlets within the 400 meter buffer around home. Because the dominance of fast food outlets over other types of food outlets may drive purchasing behaviour and diet quality (13, 20), we also calculated the 'relative fast food exposure' by dividing the total number of fast food outlets by the total number of food outlets.

The food environment consists of a variety of outlets that may all influence food choices and ultimately obesity. To capture the healthiness of the food environment, we also calculated a healthiness score of all food outlets. Scores were used derived from a Delphi study that obtained input from 20 Dutch academics and nutrition experts (21). All experts were requested to score the healthiness of a list of food outlets ranging from -5 points (very unhealthy) to +5 points (very healthy). The range of scores and rationales were shared among all experts and discussed in several rounds until consensus was reached. Fast food was considered the most unhealthy food outlet (-4.9 points), whereas the green-grocer was considered the most healthy food outlet (4.8 points). We calculated the average healthiness score of the food environment within 400 meters from home. Addresses without any food outlet were excluded, since it is uncertain what value to assign to these addresses.

Body composition measures

At the age 4 years, body height and weight measurements were performed during routine visits at the Child Health Centres. At the age of 6, 10 and 14 years, children visited the research centre in the Erasmus Medical Center for detailed physical examinations. During all examinations, height and weight were measured without shoes and heavy clothing. Height was measured to the nearest millimetre by a stadiometer (Holtain Limited, Dyfeld, UK). Weight was measured to the nearest gram using an electronic scale (SECA, Almere, the Netherlands). BMI was calculated [weight (kg)/height (m^2)] and age- and sex-specific standard deviation scores (SDS) for BMI were obtained from Dutch reference growth charts (Growth Analyzer 4.0, Dutch Growth Research Foundation) (22).

Body composition was measured at the research centre at the age of 6 and 10 years using a DXA scanner (iDXA, GE-Lunar, 2008, Madison, WI, USA) and enCORE software version 12.6. Children were placed without shoes, heavy clothing, and metal objects in supine position on the DXA table. We calculated fat mass index (FMI) [fat mass (kg)/height (m^3)] and fat-free mass index (FFMI) [fat-free mass (kg)/height (m^3)] according to standard procedures (23). We calculated SDS for FMI and FFMI [(observed value−mean)/SD] on the basis of the cross-sectional sample distribution within the Generation R study population to enable comparisons of effect estimates for FMI and FFMI.

Maternal education level

We used maternal education level to stratify the analyses by SEP. The highest education level attained was established by questionnaire at the child's age of 6 years, and categorised according to the Dutch Standard Classification of Education into lower (up to intermediate vocational training), and higher (higher vocational training, or above) (24). We could not explore patterns for smaller steps of education level, due to the small number of children per subgroup.

Other sociodemographic variables

Sociodemographic characteristics obtained at baseline and during follow-up visits by means of questionnaires included age, sex, ethnicity, and net household income. In accordance with Statistics Netherlands (25), a child's ethnic background was classified as native Dutch, other-Western background, and non-Western background based on the country of birth of the child's parents. Net household income was asked for at the child's age of 6, 10, and 14 years, and categorized into low (≤€2000/month), intermediate (<€2000-€3200/month and <€2000-€3300/month at the age of 14 years), and high (>€3200/month and >€3300/month at the age of 14 years).

Statistical analyses

All analyses were stratified by maternal education level. Characteristics of the children at the age of 4, 6, 10, and 14 years were presented as means with standard deviations (SDs) for normally distributed variables, medians with interquartile range (IQR) for skewed variables, and percentages for categorical variables. Descriptive statistics were used to characterize the within-person changes in the food environment over time. Log binomial regression models were used to evaluate the association between maternal education level and the risk that the food environment became more unhealthy over time. Associations were expressed in rate ratios (RR) for the relative fast food exposure, absolute fast food exposure and healthiness of the food environment.

Fixed-effects linear regression models were used to study within-person changes in the food environment and the association with within-person changes in body composition measures. This method allowed to control for measured time-variant variables and for unmeasured time-invariant variables (26). All models were specified using the first-difference model, and evaluated the change between two waves available, which were not necessarily two consecutive waves. Models were adjusted for the time between measurements. The specification of the model is presented in the supplement (Appendix 1).

Three separate models were constructed to study the association between changes in the absolute and relative fast food exposure, and the healthiness of the food environment, with changes in BMI SDS between the age of 4, 6, 10, and 14 years. Likewise, separate models were conducted for changes in FMI SDS and FFMI SDS between the age of 6 and 10 years. The results were presented per 10 %-point increase on the relative scale, per 1 fast food outlet increase on the absolute scale, and per 0.5 point increase in healthiness score of the food environment (which indicates a healthier environment). Units were based on the average change in the population between time-points (Supplemental Table 1).

The aforementioned analyses assume a linear exposure-response relationship, whereby every unit decrease or increase in the food environment has a similar effect. However, changes observed for children without any fast food outlet available (e.g., from 0 to 2 outlets) may have stronger effects than similar changes observed at the higher end of the distribution (e.g., from 5 to 7 outlets). Hence, we also performed analyses restricted to children without any fast food outlets during the first measurement round, and explored if the introduction of fast food outlets mattered for the subsequent changes in outcomes.

Two sensitivity analyses were conducted. First, we controlled for the time-varying factor net household income, since having a higher income may contribute to healthier dietary behaviours (27). Income was not collected at age 4 years, thus this analysis was restricted for the age of 6, 10, and 14 years. We excluded children for which net household income was missing at all time-points (n=222). Income did not change between time-points for 77.5% of the children. We imputed income measured at the nearest time-point assuming that income did not change (12.7% imputations). Second, we additionally included changes in the food environment that occurred when children moved houses. In this analyses, we included 4,594 children with 13,528 person-observations of BMI, and 7,856 person-observations on fat mass.

All analyses were conducted in R version 3.4.1, using the plm package for the fixed-effects analyses. Clustered sandwich estimators were used to allow for within-child correlation between error terms. Two-sided P-values<0.05 were considered statistically significant.

RESULTS

The median period between the first and last measurement was 7.1 years (IQR: 3.8 to 7.6 years). At all time-points, children from lower educated mothers were exposed to more fast food, both on the relative and absolute scale. These children were also exposed to unhealthier food outlets, and had a higher BMI and FMI than children from higher educated mothers (Table 1). Some children were exposed to a healthier food environment over time, but a larger proportion of children were exposed to an unhealthier food environment (Figure 2). Over time, the food environment of children from lower educated mothers as compared to higher educated mothers was more likely to have changed towards an environment offering more fast food, both relatively (RR = 1.27 (95% CI: 1.19 to 1.37)) and absolutely (RR = 1.20 (95% CI: 1.09 to 1.30)). A shift towards unhealthier food outlets (RR = 1.10 (95% CI: 1.03 to 1.18)) was also more likely to occur for children from lower educated mothers.

At the age of 6 years, 32.1% of the children from lower educated mothers and 23.1% of the children from higher educated mothers were exposed to 4 or more fast food outlets in the direct vicinity of their home. Subsequent changes in fast food outlets, both on the relative and absolute scale, and the healthiness of the food environment were not associated with changes in body composition (Table 2). Small increases in

FMI SDS were seen for children from lower educated mothers with a higher relative fast food exposure over time, and this was not observed for children from higher educated mothers.

When focussing on children without fast food around home during the first measurement round, small increases in BMI were observed for children from lower educated mothers in which relative and absolute fast food exposure increased (Table 3). Primarily increases in FMI were observed, however, confidence intervals were wide. For children from higher educated mothers without fast food outlets at first, the introduction of fast food was not associated with body composition measures.

Additionally adjusting for the time-varying factor net household income did not change the findings (Supplemental Table 2a and 2b). None of the associations found in the main analyses were found when studying changes over time when additionally including changes resulting from residential moves (Supplemental Table 3a and 3b).

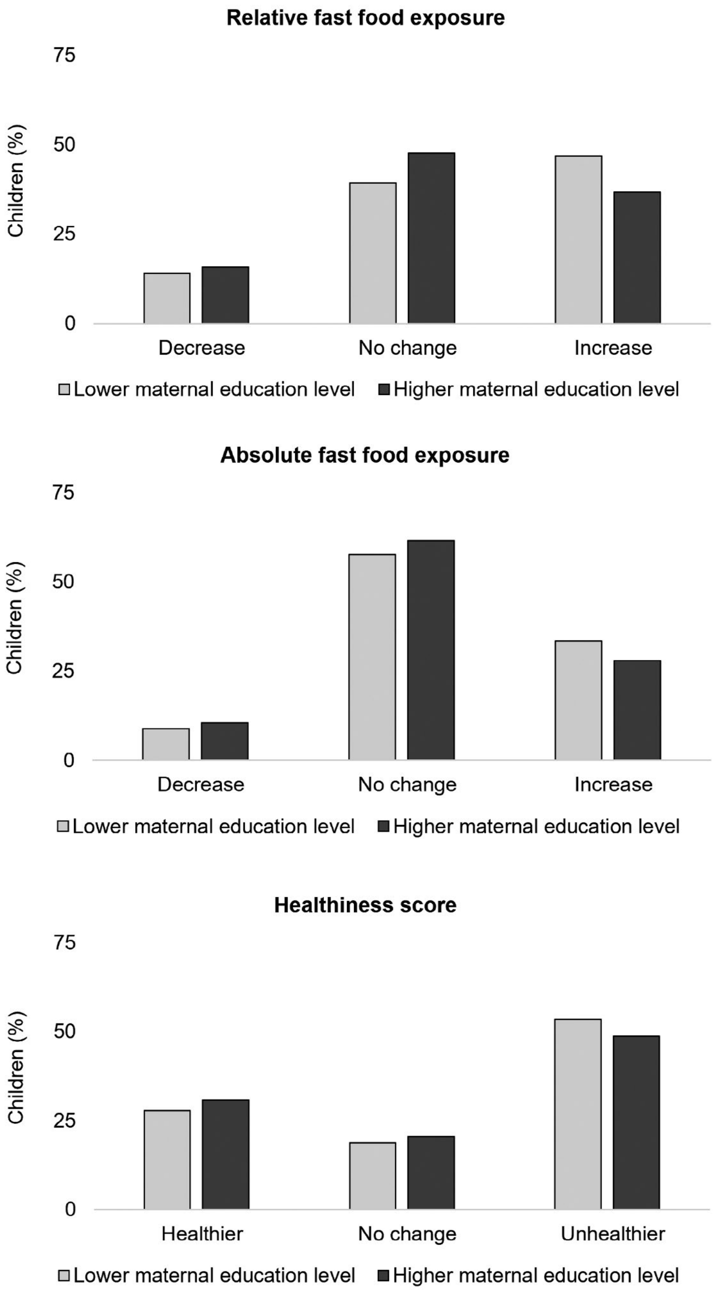


Figure 2: Changes in the food environment within children between the age of 4 and 14 years

Table 1: Characteristics of children from the Generation R study

	Lower maternal education level				Higher maternal education level			
	4 years (n=392)[a]	6 years (n=1,799)	10 years (n=1,961)	14 years (n=1,453)	4 years (n=393)	6 years (n=1,723)	10 years (n=1,978)	14 years (n=1,578)
Year of data collection[b]	2005-2008	2007-2011	2011-2014	2014-2019	2005-2008	2007-2011	2011-2014	2014-2019
Child characteristics								
Age, year	3.9 ± 0.1[c]	6.2 ± 0.6	9.8 ± 0.4	13.6 ± 0.4	3.9 ± 0.1	6.1 ± 0.4	9.8 ± 0.3	13.6 ± 0.3
Girls, n (%)	183 (46.7)	894 (49.7)	998 (50.9)	743 (51.1)	181 (46.1)	845 (49.0)	997 (50.4)	793 (50.3)
Ethnicity, n (%)								
Dutch	192 (49.0)	866 (48.1)	958 (48.9)	741 (51.0)	278 (70.7)	1,268 (73.6)	1,476 (74.7)	1,191 (75.5)
Non-Western	174 (44.4)	817 (45.4)	873 (44.5)	611 (42.1)	56 (14.2)	237 (13.8)	269 (13.6)	204 (12.9)
Other-Western	26 (6.6)	116 (6.4)	130 (6.6)	101 (7.0)	59 (15.0)	217 (12.6)	232 (11.7)	183 (11.6)
Net household income, n (%)								
Low	NA	500 (35.4)	522 (35.0)	311 (28.3)	NA	103 (6.5)	129 (7.2)	48 (3.4)
Inter-mediate	NA	515 (36.4)	432 (29.0)	374 (34.0)	NA	290 (18.4)	273 (15.3)	288 (20.7)
High	NA	399 (28.2)	538 (36.1)	414 (37.7)	NA	1,183 (75.1)	1,380 (77.4)	1,058 (75.9)
Measures of body composition								
BMI SDS	0.14 ± 1.11	0.36 ± 0.99	0.47 ± 1.13	0.47 ± 1.20	0.10 ± 0.85	0.11 ± 0.81	0.05 ± 0.92	0.00 ± 1.02
FMI SDS	NA	0.16 ± 1.21	0.34 ± 1.60	NA	NA	-0.28 ± 0.82	-0.41 ± 1.07	NA
FFMI SDS	NA	-0.01 ± 0.84	0.07 ± 0.83	NA	NA	-0.02 ± 0.78	-0.09 ± 0.73	NA

Table 1: (Continued) Characteristics of children from the Generation R study

	Lower maternal education level				Higher maternal education level			
	4 years (n=392)[a]	6 years (n=1,799)	10 years (n=1,961)	14 years (n=1,453)	4 years (n=393)	6 years (n=1,723)	10 years (n=1,978)	14 years (n=1,578)
Food environment within 400 meter around home								
Total food outlets, n	16 [1, 41][d]	9 [1, 31]	9 [1, 28]	8 [1, 27]	8 [1, 24]	6 [1, 21]	5 [1, 20]	5 [1, 20]
Fast food outlets, n	2 [0, 6]	2 [0, 5]	2 [0, 5]	2 [0, 5]	1 [0, 3]	1 [0, 3]	1 [0, 3]	1 [0, 4]
Relative fast food (%)	16 (14)	17 (16)	19 (17)	22 (20)	14 (18)	15 (17)	17 (18)	19 (19)
Healthiness score	-0.80 (1.10)	-0.75 (1.17)	-0.82 (1.21)	-0.94 (1.26)	-0.69 (1.24)	-0.65 (1.20)	-0.74 (1.18)	-0.82 (1.22)

[a] Data at age 4 years were retrieved from routine visits at the Child Health Centres that were less frequently visited by parents and their children than the research centre at age 6, 10 and 14 years. Furthermore, data collected prior to 2005 were excluded due to missing information on the food environment.

[b] Refers to the year in which body composition outcomes were measured. This was linked to characteristics of the food environment of the preceding year.

[c] Values are means ± SD (all such values)

[d] Values are medians [IQR] (all such values)

Number of missing values are presented in the Appendix 2

Table 2: Association between changes in the food environment and changes in body composition measures

	Lower maternal education level			Higher maternal education level		
	Persons	Person-observations	Estimate (95% CI)	Persons	Person-observations	Estimate (95% CI)
Relative fast food exposure (+10%-point)						
BMI (SDS)	2,153	5,605	0.01 (-0.01; 0.04)	2,082	5,672	0.01 (-0.01; 0.02)
FMI (SDS)	1,561	3,122	**0.04 (0.00; 0.08)**	1,559	3,118	0.01 (-0.01; 0.04)
FFMI (SDS)	1,561	3,122	0.00 (-0.02; 0.03)	1,559	3,118	0.01 (-0.01; 0.03)
Absolute fast food exposure (+1 outlet)						
BMI (SDS)	2,153	5,605	0.01 (-0.01; 0.02)	2,082	5,672	0.01 (-0.01; 0.02)
FMI (SDS)	1,561	3,122	0.02 (-0.00; 0.04)	1,559	3,118	0.01 (-0.01; 0.04)
FFMI (SDS)	1,561	3,122	0.01 (-0.01; 0.02)	1,559	3,118	-0.00 (-0.02; 0.01)
Healthiness score (+0.5 point)						
BMI (SDS)	1,812	4,558	0.00 (-0.02; 0.02)	1,657	4,363	0.01 (-0.01; 0.02)
FMI (SDS)	1,223	2,446	-0.01 (-0.03; 0.02)	1,170	2,340	0.00 (-0.02; 0.02)
FFMI (SDS)	1,223	2,446	-0.01 (-0.02; 0.01)	1,170	2,340	-0.00 (-0.02; 0.01)

Estimates were obtained from fixed-effects linear regression using a first-difference model specification. All analyses were adjusted for the time between measurements. BMI was collected at age 4, 6, 10 and 14 years, FMI and FFMI at age 6 and 10 years.

DISCUSSION

Children from lower educated mothers had more fast food outlets as well as an unhealthier food environment around home. These children were also at higher risk that the food environment unfavourably changed offering more fast food and other unhealthy food outlets. We observed that 32.1% of the children from lower educated mothers were exposed to 4 or more fast food outlets, whereas this was 23.1% of the children from higher educated mothers. Subsequent changes in fast food outlets or in the healthiness of the food environment were not associated with body composition measures. For children from lower educated mothers without fast food exposure during the first measurement round, small increases in BMI were seen following the introduction of fast food outlets.

Table 3: Associations between changes in the food environment and changes in body composition measures for children without fast food outlets around home during the first measurement round

	Lower maternal education level			Higher maternal education level		
	Persons	Person-observations	Estimate (95% CI)	Persons	Person-observations	Estimate (95% CI)
Relative fast food exposure (+10%-point)						
BMI (SDS)	705	1,879	**0.04 (0.00; 0.07)**	914	2,468	-0.02 (-0.04; 0.00)
FMI (SDS)	546	1,092	0.07 (-0.02; 0.16)	671	1,342	0.01 (-0.03; 0.04)
FFMI (SDS)	546	1,092	0.03 (-0.02; 0.07)	671	1,342	-0.00 (-0.02; 0.02)
Absolute fast food exposure (+1 outlet)						
BMI (SDS)	705	1,879	0.06 (-0.00; 0.12)	914	2,468	-0.04 (-0.12; 0.03)
FMI (SDS)	546	1,092	0.09 (-0.03; 0.20)	671	1,342	-0.08 (-0.20; 0.02)
FFMI (SDS)	546	1,092	-0.00 (-0.05; 0.05)	671	1,342	-0.02 (-0.07; 0.03)

Estimates were obtained from fixed-effects linear regression using a first-difference model specification. All analyses were adjusted for the time between measurements. BMI was collected at age 4, 6, 10 and 14 years, FMI and FFMI at age 6 and 10 years.

3

Strengths and limitations

A strength of the study was the use of repeated measures of the food environment and objective measures of body composition among a large cohort of school-aged children. A 2 to 4 year difference between time-points was sufficiently large so that a group of children experienced changes in the food environment around home, but small enough to be able to attribute the change in body composition to changes in the food environment. Some of the bias that may have arisen from residential self-selection was taken into account by using a fixed-effects approach. Focusing on children in which the exposure of food outlets around home changed during follow-up, as opposed to cross-sectional studies and studies in which children move across neighbourhoods, is essential for this purpose.

Our study also had some weaknesses. First, time-variant variables may have caused confounding. Instrumental Variable (IV) approaches may further eliminate bias from unobserved time-varying confounding, and two studies found that associations between the food environment and diet quality or BMI were stronger in magnitude using IV models compared to fixed-effects analyses (13, 14). Identifying a valid instrument to study built environment changes is challenging, and this was not further explored in the current study. Second, we lacked information of perceived availability of food outlets and purchasing behaviour. Two reviews indicated that perceived exposure of food outlets were yielding more consistent findings with dietary behaviours (6, 8). Furthermore, exposure to food outlets does not automatically imply that these outlets are accessible (28). Third, we were not able to evaluate changes occurring at other locations where children and their parents were exposed to food (e.g., school environment), or changes made to the physical activity environment (e.g., new playgrounds) that may also affect body composition. Assessments of the food environment based on addresses, rather than using administrative boundaries of neighbourhoods, may have prevented some of the biases resulting from changes occurring in the neighbourhood. Finally, the current evaluation lacked insights into the mechanism through which the food environment affect purchasing behaviours, dietary patterns and subsequently body composition.

The current evidence base defined and observed inequalities in the fast food environment in different ways. We categorized children by maternal education level, and found that children from lower educated mothers were more often exposed to an unhealthy food environment. A review reported that fast food locations were more prevalent in low-income neighbourhoods and neighbourhoods with high concentrations of ethnic minority groups (15). Although conceptualised differently, they point to similar conclusions that exposure to the food environment is not equally distributed across the population. Furthermore, we found that the food environment shifted towards a higher exposure of unhealthy foods, and this unhealthy shift was more likely to occur for children from lower educated mothers. Similar patterns were observed in a study conducted in New York City, US, suggesting that the number of unhealthy food outlets increased more rapidly in neighbourhoods with a higher initial number of unhealthy food outlets, and for disadvantaged neighbourhoods (29).

The counterintuitive finding of a US study, suggesting that BMI decreased following the increase in fast food outlets, was not found in our study (10). We did not find evidence that the food environment around home was associated with children's body composition measures in the total population. A likely explanation is that the introduction of 1 more fast food outlet or changes in healthiness of food outlets were too small to have a noticeable effect on obesity measures for children living in a so-called "fast food paradise", given the numerous food outlets around home. Our results gave some indications that fast food exposure was associated with small increases in BMI for children of lower educated mothers that were not exposed to fast food during the first measurement round, but not for children from higher educated mothers. DXA scans suggested that especially fat mass increased following the introduction of fast food outlets. In sensitivity analyses, no associations between the food environment and body composition were found when additionally including changes resulting from residential moves. This could suggest that other factors that change during residential relocation, such as the impact of changing schools, may contribute to body composition to a larger degree.

The finding that the food environment may not equally affect all population subgroups was also found in longitudinal studies using a fixed-effects approach conducted in adult populations in the US. Young adults were followed for 15 years and results suggested that fast food outlets were associated with higher fast food intake among low income families, but not among higher income families (11). Another study found that favourable changes in the food environment were related to BMI reductions in obese adults over a 10-year period, but not for normal weight or overweight adults (12). Furthermore, a review based on primarily cross-sectional studies reported that a higher fast food availability was associated with obesity for children from lower income households or neighbourhoods (9). These findings show the importance to thoroughly examine differential effects of the food environment on dietary behaviours and body composition outcomes.

The high and increasing number of fast food outlets, and the larger exposure to them of children from lower socioeconomic groups provide support to policies aimed at limiting the number of fast food outlets, especially in residential areas. To date, strategies to improve dietary habits mainly rely on nutritional education rather than on structural interventions to modify the food environment. That changes in BMI were small following the introduction of fast food outlets around home between a 2 to 4 year period should not limit efforts to do so. The cumulative effect of the food environment over time and across various daily activity settings was not explored, but may steadily impact on children's body composition. Comparing different generations may be useful for this purpose, however, disentangling the impact of the food environment from other causes that are likely to differ between generations is challenging. Study designs that allow to unravel causal mechanisms how fast food outlets influences body composition, for example by having information on fast food consumption or dietary patterns, is essential. Furthermore, the wider context in which food environments change and food choices are made is important to consider. It has been suggested that more food outlets could reflect general investments made to that area, which could positively or negatively influence children's body composition (29). More insights are needed to understand these type of complexities around childhood obesity (30).

CONCLUSION

This study provided evidence of widening socioeconomic inequalities in the food environment around children's homes growing up in a large urban area in the Netherlands over a period of 7 years. Children from lower educated mothers were at higher risk of unfavourably changing food environments. Exposure to more fast food outlets does not seem to have an additional impact on BMI in contemporary contexts with ubiquitous fast food outlets. These findings further support the importance of an equity-focus on the link between the food environment and body composition.

REFERENCES

1. Lobstein T, Baur L, Uauy R. Obesity in children and young people: a crisis in public health. Obes Rev. 2004;5(s1):4-85.
2. Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. The Lancet. 2017;390(10113):2627-42.
3. Pinho MGM, Mackenbach JD, den Braver NR, Beulens JJW, Brug J, Lakerveld J. Recent changes in the Dutch foodscape: socioeconomic and urban-rural differences. Int J Behav Nutr Phys Act. 2020;17(1):43.
4. Monteiro CA, Moubarac JC, Cannon G, Ng SW, Popkin B. Ultra-processed products are becoming dominant in the global food system. Obes Rev. 2013;14(S2):21-8.
5. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. The Lancet. 2011;378(9793):804-14.
6. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: A systematic review. Health & place. 2012;18(5):1172-87.
7. Osei-Assibey G, Dick S, Macdiarmid J, Semple S, Reilly JJ, Ellaway A, et al. The influence of the food environment on overweight and obesity in young children: a systematic review. BMJ Open. 2012;2(6):e001538.
8. Engler-Stringer R, Le H, Gerrard A, Muhajarine N. The community and consumer food environment and children's diet: a systematic review. BMC Public Health. 2014;14:522.
9. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CAM. The relationship of the local food environment with obesity: A systematic review of methods, study quality and results. Obesity (Silver Spring, Md). 2015;23(7):1331-44.
10. Sandy R, Liu G, Ottensmann J, Tchernis R, Wilson J, Ford OT. Studying the child obesity epidemic with natural experiments. National Bureau of Economic Research, 2009.
11. Boone-Heinonen J, Gordon-Larsen P, Kiefe CI, Shikany JM, Lewis CE, Popkin BM. Fast food restaurants and food stores: longitudinal associations with diet in young to middle-aged adults: the CARDIA study. Arch Intern Med. 2011;171(13):1162-70.
12. Barrientos-Gutierrez T, Moore KAB, Auchincloss AH, Mujahid MS, August C, Sanchez BN, et al. Neighborhood Physical Environment and Changes in Body Mass Index: Results From the Multi-Ethnic Study of Atherosclerosis. American Journal of Epidemiology. 2017;186(11):1237-45.

13. Rummo PE, Guilkey DK, Ng SW, Meyer KA, Popkin BM, Reis JP, et al. Understanding bias in relationships between the food environment and diet quality: the Coronary Artery Risk Development in Young Adults (CARDIA) study. Journal of Epidemiology and Community Health. 2017;71(12):1185-90.

14. Rummo PE, Guilkey DK, Ng SW, Meyer KA, Popkin BM, Reis JP, et al. Does unmeasured confounding influence associations between the retail food environment and body mass index over time? The Coronary Artery Risk Development in Young Adults (CARDIA) study. International Journal of Epidemiology. 2017;46(5):1456-64.

15. Fleischhacker SE, Evenson KR, Rodriguez DA, Ammerman AS. A systematic review of fast food access studies. Obes Rev. 2011;12(5):e460-e71.

16. Chum A, Farrell E, Vaivada T, Labetski A, Bohnert A, Selvaratnam I, et al. The effect of food environments on fruit and vegetable intake as modified by time spent at home: a cross-sectional study. BMJ Open. 2015;5(6):e006200.

17. Mackenbach JD, Nelissen KGM, Dijkstra SC, Poelman MP, Daams JG, Leijssen JB, et al. A Systematic Review on Socioeconomic Differences in the Association between the Food Environment and Dietary Behaviors. Nutrients. 2019;11(9):2215.

18. Kooijman MN, Kruithof CJ, van Duijn CM, Duijts L, Franco OH, van Ijzendoorn MH, et al. The Generation R study: design and cohort update 2017. European Journal of Epidemiology. 2016;31(12):1243-64.

19. Canalia C, Pinho MGM, Lakerveld J, Mackenbach JD. Field Validation of Commercially Available Food Retailer Data in the Netherlands. Int J Environ Res Public Health. 2020;17(6).

20. Clary CM, Ramos Y, Shareck M, Kestens Y. Should we use absolute or relative measures when assessing foodscape exposure in relation to fruit and vegetable intake? Evidence from a wide-scale Canadian study. Preventive Medicine. 2015;71:83-7.

21. Timmermans J, Dijkstra C, Kamphuis C, Huitink M, van der Zee E, Poelman M. 'Obesogenic' School Food Environments? An Urban Case Study in the Netherlands. International Journal of Environmental Research and Public Health. 2018;15(4):619.

22. Fredriks AM, van Buuren S, Wit JM, Verloove-Vanhorick SP. Body index measurements in 1996-7 compared with 1980. Arch Dis Child. 2000;82(2):107-12.

23. Wells JCK, Cole TJ, team As. Adjustment of fat-free mass and fat mass for height in children aged 8 y. International Journal of Obesity. 2002;26(7):947-52.

24. Statistics Netherlands. Standaard Onderwijsindeling 2003. Voorburg/Heerlen2004.

25. Statistics Netherlands. Jaarrapport Integratie 2010. Den Haag/Heerlen2010.

26. Allison PD. Fixed effects regression models: SAGE publications; 2009.

27. Drewnowski A, Specter SE. Poverty and obesity: the role of energy density and energy costs. The American Journal of Clinical Nutrition. 2004;79(1):6-16.

28. Clary C, Matthews SA, Kestens Y. Between exposure, access and use: Reconsidering foodscape influences on dietary behaviours. Health & Place. 2017;44:1-7.

29. Berger N, Kaufman TK, Bader MDM, Rundle AG, Mooney SJ, Neckerman KM, et al. Disparities in trajectories of changes in the unhealthy food environment in New York City: A latent class growth analysis, 1990-2010. Soc Sci Med. 2019;234:112362.

30. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. Lancet. 2017; 9;390(10112):2602-2604.

SUPPLEMENTAL MATERIAL

Appendix 1	Model specification
Appendix 2	Number of missings
Supplemental Table 1	Change in the food environment between time-points
Supplemental Table 2a	Associations between changes in the food environment and changes in body composition measures adjusted for income
Supplemental Table 2b	Associations between changes in the food environment and changes in body composition measures for children without fast food outlets around home, adjusted for income
Supplemental Table 3a	Associations between changes in the food environment and changes in body composition, including children that moved houses
Supplemental Table 3b	Associations between changes in the food environment and changes in body composition for children without fast food outlets around home, including children that moved houses

Appendix 1: Model specification

The following regression model was used:

$$y_{it-1} = \alpha_t + \beta_1 (food\ environment_{it-1}) + \beta_2 (x_{it-1}) + \mu_i + \varepsilon_{it-1}$$

Where y_{it} is the dependent variable of interest (e.g., BMI SDS, FMI SDS, FFMI SDS) for individual i on time t, β_1 is the effect of changes in the food environment for individual i on time t-1 (e.g., absolute fast food exposure, relative fast food exposure, healthiness score of the food environment), β_2 is the effect of time-varying factors for individual i on time t-1, α_t accounts for time effects that are constant across individuals, μ_i accounts for time-invariant random errors on the individual-level, and ε_{it} accounts for normal sources of error that vary across individuals and time.

3

Appendix 2: Number of missings

	4 year (n=785)	6 year (n=3,522)	10 year (n=3,939)	14 year (n=3,031)
Age	0	0	0	0
Sex	0	0	0	0
Ethnicity	0	1	1	0
Net household income	NA	532	665	538
BMI	0	0	0	0
FMI	NA	263	298	NA
FFMI	NA	263	298	NA

Supplemental Table 1: Change in the food environment between time-points

	Lower maternal education level (n=2,153)		Higher maternal education level (n=2,082)	
	Persons	Mean difference (SD)	Persons	Mean difference (SD)
Relative fast food exposure (%)				
4-6 year				
Decrease	78	-2.6 (3.5)	53	-3.8 (7.6)
Increase	90	3.4 (6.4)	80	4.8 (8.5)
No change	211	0 (0)	245	0 (0)
6-10 year				
Decrease	214	-9.1 (12.9)	256	-9.2 (12.3)
Increase	689	7.0 (9.9)	584	7.5 (10.3)
No change	707	0 (0)	779	0 (0)
10-14 year				
Decrease	249	-6.9 (12.3)	253	-10.4 (17.9)
Increase	540	7.7 (12.4)	496	7.3 (12.3)
No change	596	0 (0)	789	0 (0)
Absolute fast food exposure				
4-6 year				
Decrease	34	-1.2 (0.5)	20	-1.3 (0.6)
Increase	41	1.1 (0.3)	47	1.1 (0.5)
No change	304	0 (0)	311	0 (0)
6-10 year				
Decrease	164	-1.4 (0.7)	195	-1.2 (0.5)
Increase	456	2.1 (1.8)	399	1.6 (1.1)
No change	990	0 (0)	1,025	0 (0)
10-14 year				
Decrease	131	-1.3 (0.6)	134	-1.3 (0.6)
Increase	405	2.2 (2.0)	345	2.1 (1.9)
No change	849	0 (0)	1,059	0 (0)

Supplemental Table 1: (Continued) Change in the food environment between time-points

	Lower maternal education level (n=2,153)		Higher maternal education level (n=2,082)	
	Persons	Mean difference (SD)	Persons	Mean difference (SD)
Healthiness score				
4-6 year				
Decrease (unhealthier)	87	-0.4 (0.6)	91	-0.5 (0.7)
Increase (healthier)	98	0.2 (0.3)	60	0.3 (0.5)
No change	131	0 (0)	162	0 (0)
6-10 year				
Decrease (unhealthier)	599	-0.5 (0.7)	534	-0.6 (0.8)
Increase (healthier)	394	0.4 (0.6)	403	0.4 (0.5)
No change	270	0 (0)	280	0 (0)
10-14 year				
Decrease (unhealthier)	554	-0.5 (0.5)	511	-0.5 (0.6)
Increase (healthier)	280	0.5 (0.8)	307	0.6 (0.7)
No change	266	0 (0)	316	0 (0)

Supplemental Table 2a: Associations between changes in the food environment and changes in body composition measures adjusted for income

	Lower maternal education level			Higher maternal education level		
	Persons	Person-observations	Estimate (95% CI)	Persons	Person-observations	Estimate (95% CI)
Relative fast food exposure (+10%-point)						
BMI (SDS)	1,973	4,841	0.01 (-0.01; 0.04)	2,040	5,185	0.01 (-0.01; 0.02)
FMI (SDS)	1,456	2,912	**0.04 (0.00; 0.08)**	1,532	3,064	0.01 (-0.01; 0.03)
FFMI (SDS)	1,456	2,912	0.00 (-0.02; 0.03)	1,532	3,064	0.01 (-0.01; 0.03)
Absolute fast food exposure (+1 outlet)						
BMI (SDS)	1,973	4,841	0.01 (-0.00; 0.03)	2,040	5,185	0.00 (-0.02; 0.02)
FMI (SDS)	1,456	2,912	0.02 (-0.00; 0.05)	1,532	3,064	0.01 (-0.02; 0.03)
FFMI (SDS)	1,456	2,912	0.01 (-0.00; 0.02)	1,532	3,064	-0.00 (-0.02; 0.01)
Healthiness score (+0.5 point)						
BMI (SDS)	1,649	3,909	-0.00 (-0.02; 0.02)	1,621	3,958	0.01 (-0.01; 0.02)
FMI (SDS)	1,139	2,278	-0.01 (-0.04; 0.02)	1,149	2,298	0.00 (-0.02; 0.02)
FFMI (SDS)	1,139	2,278	-0.02 (-0.03; 0.01)	1,149	2,298	-0.00 (-0.02; 0.01)

Estimates were obtained from fixed-effects linear regression using a first-difference model specification. All analyses were adjusted for the time between measurements and net household income. Income was not collected at age 4 years, thus this analyses was restricted for the age of 6, 10, and 14 years. Missing income was imputed using income measured at the nearest time-point. BMI was collected at age 6, 10 and 14 years, FMI and FFMI at age 6 and 10 years.

Supplemental Table 2b: Associations between changes in the food environment and changes in body composition measures for children without fast food outlets around home, adjusted for income

	Lower maternal education level			Higher maternal education level		
	Persons	Person-observations	Estimate (95% CI)	Persons	Person-observations	Estimate (95% CI)
Relative fast food exposure (+10%-point)						
BMI (SDS)	666	1,676	**0.04 (0.00; 0.07)**	897	2,268	-0.02 (-0.04; 0.01)
FMI (SDS)	514	1,028	0.08 (-0.01; 0.17)	662	1,324	0.01 (-0.03; 0.04)
FFMI (SDS)	514	1,028	0.03 (-0.02; 0.08)	662	1,324	0.00 (-0.02; 0.02)
Absolute fast food exposure (+1 outlet)						
BMI (SDS)	666	1,676	0.07 (-0.01; 0.15)	897	2,268	-0.06 (-0.14; 0.02)
FMI (SDS)	514	1,028	0.12 (-0.00; 0.24)	662	1,324	-0.08 (-0.17; 0.03)
FFMI (SDS)	514	1,028	0.02 (-0.03; 0.06)	662	1,324	-0.01 (-0.06; 0.03)

Estimates were obtained from fixed-effects linear regression using a first-difference model specification. All analyses were adjusted for the time between measurements and net household income. Income was not collected at age 4 years, thus this analyses was restricted for the age of 6, 10, and 14 years. Missing income was imputed using income measured at the nearest time-point. BMI was collected at age 6, 10 and 14 years, FMI and FFMI at age 6 and 10 years.

Supplemental Table 3a: Associations between changes in the food environment and changes in body composition, including children that moved houses

	Lower maternal education level			Higher maternal education level		
	Persons	Person-observations	Estimate (95% CI)	Persons	Person-observations	Estimate (95% CI)
Relative fast food exposure (+10%-point)						
BMI (SDS)	2,353	6,778	0.01 (-0.01; 0.02)	2,241	6,750	0.00 (-0.01; 0.01)
FMI (SDS)	1,976	3,952	0.02 (-0.01; 0.04)	1,952	3,904	0.00 (-0.02; 0.02)
FFMI (SDS)	1,976	3,952	0.01 (-0.01; 0.02)	1,952	3,904	0.01 (-0.00; 0.02)
Absolute fast food exposure (+1 outlet)						
BMI (SDS)	2,353	6,778	0.00 (-0.04; 0.01)	2,241	6,750	0.00 (-0.00; 0.01)
FMI (SDS)	1,976	3,952	0.00 (-0.01; 0.01)	1,952	3,904	0.01 (-0.01; 0.02)
FFMI (SDS)	1,976	3,952	0.00 (-0.01; 0.01)	1,952	3,904	-0.00 (-0.01; 0.01)
Healthiness score (+0.5 point)						
BMI (SDS)	2,071	5,533	-0.00 (-0.01; 0.01)	1,883	5,161	0.01 (-0.00; 0.02)
FMI (SDS)	1,529	3,058	-0.01 (-0.03; 0.01)	1,407	2,818	0.00 (-0.02; 0.02)
FFMI (SDS)	1,529	3,058	-0.01 (-0.02; 0.00)	1,407	2,818	-0.01 (-0.02; 0.00)

Estimates were obtained from fixed-effects linear regression using a first-difference model specification. All analyses were adjusted for the time between measurements. BMI was collected at age 4, 6, 10 and 14 years, FMI and FFMI at age 6 and 10 years.

Supplemental Table 3b: Associations between changes in the food environment and changes in body composition for children without fast food outlets around home, including children that moved houses

	Lower maternal education level			Higher maternal education level		
	Persons	Person-observations	Estimate (95% CI)	Persons	Person-observations	Estimate (95% CI)
Relative fast food exposure (+10%-point)						
BMI (SDS)	714	2,066	0.02 (-0.01; 0.05)	945	2,796	-0.01 (-0.03; 0.01)
FMI (SDS)	628	1,256	0.03 (-0.03; 0.09)	804	1,608	-0.01 (-0.03; 0.01)
FFMI (SDS)	628	1,256	**0.03 (0.00; 0.07)**	804	1,608	0.00 (-0.02; 0.02)
Absolute fast food exposure (+1 outlet)						
BMI (SDS)	714	2,066	0.00 (-0.02; 0.03)	945	2,796	-0.01 (-0.02; 0.01)
FMI (SDS)	628	1,256	-0.01 (-0.05; 0.02)	804	1,608	-0.01 (-0.04; 0.02)
FFMI (SDS)	628	1,256	0.02 (-0.03; 0.06)	804	1,608	-0.01 (-0.02; 0.01)

Estimates were obtained from fixed-effects linear regression using a first-difference model specification. All analyses were adjusted for the time between measurements. BMI was collected at age 4, 6, 10 and 14 years, FMI and FFMI at age 6 and 10 years.

4

Hosting elite sport events to target recreational sport participation: an interrupted time series analysis

Famke J.M. Mölenberg, Frouwkje de Waart, Alex Burdorf, Frank J. van Lenthe

International Journal of Sport Policy and Politics 2020;12(4):531-543

ABSTRACT

Introduction

Increasing sport participation by hosting elite sport events is a much-debated policy. This study evaluated the effect of hosting elite sport events on sport participation against the background of a shift in focus towards targeting recreational sport participation.

Methods

We included 10 international elite sport events organised between 2000 and 2017 in the city of Rotterdam, the Netherlands. Sport-specific participation in the past year was obtained from a biannual cross-sectional survey. Per event, interrupted time series analyses were employed using 3 pre- and post-event measures. Data were summarised by means of random-effects meta-analyses. We tested for group differences to evaluate if events organised in the period with a more explicit aim to target recreational sport participation, and associated policies, had a larger impact.

Results

Three events concerning cycling, table tennis and gymnastics were followed by an increase in sport-specific participation 1-year after the event was organised, whereas the korfball event was followed by a decrease. The pooled effect of the 10 events did not show any change in sport-specific participation (0.2 %-point (95% CI: -0.3; 0.8)). Significant group differences by period were found. More recent events targeting sport participation were followed by an increase in sport-specific participation (1.1 %-point (95% CI: 0.0; 2.1)), but not for other events (-0.3 %-point (95% CI: -0.6; 0.1)). No group differences were found for the number of visitors and location.

Conclusion

Hosting elite sport events that explicitly target sport participation may increase sport participation among citizens. Longitudinal data following individuals over time are needed to support this finding.

INTRODUCTION

Physical inactivity is an important risk factor for many diseases (1) and sports participation can contribute considerably to leisure-time physical activity and health (2). Stimulating sports participation by hosting elite sport events, although a much-debated policy, can leave a legacy making the investment worthwhile (3). Increased participation in the population, in turn, may provide a wider 'pool' for talent from which professional athletes of the future can be selected. This process has been described as the 'virtuous cycle' of sport, and is often the reason why governments invest in elite sport (4).

Theoretical frameworks suggest that hosting elite sport events may increase participation in sports activities by means of a so-called "trickle-down" effect (5-7). The trickle-down effect describes the process by which the general population is inspired by elite sports, professional athletes and sport events, to engage in sport activities and physical activity behaviours (7), either directly or indirectly (8). Direct trickle-down effects may result from watching elites sport performances, whereby successes and the presence in the media of athletes inspire persons, and, as a consequence increase enrolment in sport activities and physical activity behaviours in the population. Indirect trickle-down effects may result from improvements in the sports infrastructure (e.g., the provision of new sport facilities or improvements made to existing sport facilities), temporal side-events organised along the main event, and an improved "sports culture" in the city. For example, the legacy of hosting the start of the 2007 Tour de France Grand Depart in London (UK) consisted of new cycling lanes across the city (9), and this may indirectly facilitate the uptake of cycling among the host population.

The direct trickle-down effects fit well within theories of behaviour change, and two theoretical frameworks for elite sport events have been supported with empirical evidence. The transtheoretical model of change (TTM) was used as a framework to study the impact of major sport events in the UK (10, 11). The TTM categorizes people into one of five stages based on intentions towards engagement in sport activities and actual behaviour (12). The findings suggested that people attending events were belonging to the stages closest to behaviour change (10, 11), and that progression towards more sport participation was observed following attending the event (11). The Theory of Planned Behaviour proposes that attitude, subjective norm and perceived behavioural control influence behaviour directly, and indirectly through intentions (13). Two studies showed that the intention to participate in sport activities increased within the Australian and Canadian population after hosting the Olympic Games (14, 15).

Although studies found positive changes in intentions following elite sport events (11, 14, 15), to date there is no convincing evidence that hosting elite sport events resulted in increases in sport participation or physical activity (16-18). Weeds and colleagues reviewed the evidence on legacies of Olympic Games and sport events including 11 studies, and concluded that more robust evaluation is needed to inform the development of legacies following sport events (16). McCartney and colleagues reviewed 2 studies on sport participation and 5 studies on health outcomes, and also concluded that the available evidence was not sufficient to draw conclusions (17). More recently, Annear summarised 3 reviews and additionally 6 newly published articles, and reported that most studies did not find evidence for sustained changes in adult physical activity following large sport events (18). In summary, to date there is little robust evidence on the legacy of sport events and, hence, it remains to be seen whether hosting elite sport events may increase sport participation or physical activity, and thereby result in health benefits (16-18).

There is a growing recognition that hosting sport events could have a wide range of impacts that go beyond promoting sport participation and physical activity. Impacts outsides the sports domain include new employment opportunities, economic revenues, infrastructural improvements, neighbourhood regeneration and housing projects, and increased feelings of social cohesion and pride (17, 19, 20). Evidence suggests that these changes positively impact upon population heath (21). A comprehensive review by McCartney on major sport events included 54 studies with a wide range of outcomes (17). Economic outcomes were reported by 18 studies, but the long-lasting impact of hosting events on economic growth and employment remained unclear. Four out of the five included studies showed that the restriction of car use and promotion of public transportation during major events were associated with short term reductions in traffic volume, congestion or air pollution. However, negative impacts such as economic losses and inflation were also reported in some studies. Again, the authors concluded that high quality evidence is needed to predict positive and negative impacts from hosting sport events.

The only study that used a longitudinal design was the ORiEL study, aimed at evaluating the impact of the urban regeneration associated with the London 2012 Olympic Games (UK) on adolescent physical activity, mental health and well-being (22). After 18 months of follow-up, no differences were found in physical activity between adolescents living in the intervention borough (where the majority of urban regeneration activities took place) as compared to those living in comparison boroughs (23). Urban regeneration and the Olympic Games had no impact on depressed persons, and on well-being. Also this study with high quality data could not show that physical activity increased or that health impacts were achieved following the hosting of elite sport events.

Recently, it has increasingly becoming the focus of attention that long term planning and supportive activities are needed to achieve any impact from hosting sport events (7, 24). The majority of events did not explicitly aim to promote sport participation (16, 25), but were mainly hosted to generate economic revenues. Likewise, most studies evaluating sport events focused on the economic impact, and less often included outcomes relevant for public health (17). It remains unknown if we can expect a sport legacy of hosting sport events that aimed to promote recreational sport participation and physical activity in the population.

The city of Rotterdam, the Netherlands, has a rich history of organizing major sport events. The aim of hosting elite sport events shifted from a city-marketing perspective (2000-2009), to a more health-oriented perspective (2010-2017) (26-28). From 2010 onwards, policies were implemented aiming towards more engagement of citizens (29, 30). First, events were more often organised within the city centre to attract more visitors. Second, an increasing number of side-events were organised along the main events to introduce the sport to visitors, and to city residents. To our knowledge, no study has yet evaluated whether sport events that aimed at increasing sport participation were more likely to do so, as compared to events that were mainly organised for other purposes.

This study aimed at evaluating the effect of hosting elite sport events on recreational sport participation among citizens. Our secondary aim was to evaluate if the shift in focus towards sports participation resulted in higher sport participation in more recent events.

METHODS

Elite sport events

The Rotterdam Topsport Foundation (a foundation funded by the city government for the support of local elite clubs and the acquisition and organisation of elite sport events) provided information about large international elite sport events hosted between 2000 and 2017. This included World Championships (WC), European Championships (EC), and other major events that were organised once in the city of Rotterdam, the Netherlands. Annual events were not included, because these events could not be related to changes in sport participation across a specific time-point. Soccer events were also not included, since soccer receives much media attention throughout the year. Events that attracted at least 5,000 visitors were considered for inclusion; 22 events were identified. We excluded events without sport-specific participation data for 3 measurement rounds before and after the event (n=8), and events with the smallest number of visitors that promoted the same sport (n=4). Accordingly, changes in sport-specific participation were related to 10 singular elite sport events.

Events targeting sport participation

Since 2010, all elite sport events hosted in the city of Rotterdam and funded through the event funds are accompanied by side-events along the main event for the purpose of creating societal impact. In the bidding phase, event organizers need to provide a detailed plan how sport participation in the population will be targeted, and they are requested to budget a minimum of 10% of the total requested amount from the municipality (or a minimum of €7,500) for the organisation of side-events. Funding is only provided if these criteria are met.

For events before 2010 the aim to increase sport participation, and forthcoming actions including the organisation of side-events, were not documented given that it was not a policy goal for the organisation of sport events prior to 2010. Therefore, we could not separate events that targeted increasing sport participation from those primarily interested in generating economic benefits and city marketing. We categorised events by year, and separated events organised in the period without a specific aim to target sport participation (before 2010) from events organised in the period with a specific aim to target sport participation (2010 onwards).

Study design

For each event, interrupted time series analyses were employed to evaluate if events significantly interrupted the underlying time trend in the percentage of people participating in that specific sport. Sport participation data over the past year were obtained from a biannual cross-sectional survey. Data were used from 6 measurement waves around the time the event was organised. We assumed that events would result in an immediate change in the level of sport participation measured in the next round of data collection, which was, on average, 1-year after the event. For events that significantly changed sport-specific participation levels after 1-year, we examined if the change sustained up to 5-years after the event.

Study population

The cross-sectional biannual leisure-time survey among citizens of Rotterdam [in Dutch: *Vrijetijdsonderzoek*] was used to obtain sport participation data from 1995 until 2017. Under the Dutch law for medical scientific research with human subjects, questionnaire surveys are not subject to approval by an institutional ethics committee. Participants were randomly selected from the municipal registration database. Respondents were informed that by filling out the questionnaire permission was given to use the data anonymously for research purposes. All surveys included questions on demographics, leisure-time related topics, and a dedicated section on sport participation.

During 12 measurement waves, a total of 136,696 subjects were invited to participate in the survey, and 35,587 subjects returned the questionnaire. Response percentages declined from 47% in 1995 to 15% in 2017 (Supplemental Table 1). In two surveys by design, half of the participants did not receive questions on sport participation (n=6,287 subjects), and those were excluded. The total study sample for analyses included 27,235 participants aged between 16 and 75 years old for which data on age, sex, ethnicity, income, city district and sport participation were available (Supplemental Table 1).

Sport participation

A standardized questionnaire [in Dutch: *Richtlijn Sportonderzoek]* was used to ask for any sport participation over the past year. Participants could tick multiple boxes from a list of sports. The list was updated over the years, and included indoor, outdoor, organised, and un-organised activities. Sport-specific participation included multiple sport activities for cycling (all recreational types of cycling), judo (all defensive sports), sailing (all water sport activities, but not swimming or rowing), and short track (all ice skating activities).

Socio-demographic variables

Age, sex, and ethnicity were derived from the municipal registration database. Ethnicity was based on the country of birth and classified as Western, and non-Western, following the definition used by Statistics Netherlands (31). For each survey, the social minimum income and modal income in that specific year was used as the reference to classify self-reported household income into low (below social minimum income), mid-low (social minimum to modal income), mid-high (modal to 2x modal income), and high (>2x modal income).

Statistical analyses

We used an interrupted time series design that allowed to evaluate the change in sport-specific participation for 1, 3, and 5 years after the event, as compared to the trend in sport-specific participation for the 5 years before the event. For example, we evaluated if the percentage of people who played tennis in the population changed after hosting the Davis Cup Semi-final, as compared to the trend in tennis participation for the 5 years before the event. We used linear multivariable regression models adjusted for the pre-event trend in sport-specific participation, and individual-level characteristics age, sex, income and ethnicity. From this model we obtained the absolute change in sport-specific participation following events expressed as percentage points. The models were run separately for each event. We visualised the pre-event trend and the absolute change in sport-specific participation following events. Details about the model specification can be found in Appendix 1. We quantitatively summarised the data by using random-effects meta-analyses (32), with the effect of each event weighted by the inverse of its variance. Heterogeneity between events was evaluated with the I^2 test statistic (33). Forest plots were made to visualise the results.

We tested for group differences to evaluate if more recent events with the explicit aim to target sport participation had a differential effect on sport participation. More specifically, we tested for differences for events targeting sport participation (no (before 2010) *vs* yes (2010 onwards)), the number of visitors (based on the median: <37500 *vs* ≥37500), location (city centre *vs* stadium in outskirts), and side-events (yes *vs* no). We conducted a falsification test in which we estimated what the change in sport participation would have been in the hypothetical scenario that events organised before 2010 would all have taken place in 2010. For each event organised before 2010, we run a model in which we set the year of the event at 2010, and evaluated what the effect on sport-specific participation would have been if the trend in sport-specific participation was interrupted in 2010. This ensured that an observed effect for events organised from 2010 onwards was not reflecting the increase in sport participation seen in the population (34). We expected to find no statistically significant intervention effect in this falsification test.

Two sensitivity analyses were conducted. First, we additionally included events with sport participation data available for at least 1 pre-event measure and 1 post-event measure around the time that the event was organised. Second, we excluded one event at a time from the analysis to verify if the results were attributable to a single event.

All analyses were conducted in SPSS version 24. The data were summarised in R version 3.4.1, using the forest plot package.

Table 1: Sport events organised in Rotterdam, the Netherlands between 2000 and 2017

Sport	Event	Period (duration)	Location	Number of visitors	Dutch result
Tennis	Davis Cup Semi-final	September 2001 (3 days)	Indoor hall Ahoy	30,000	Loss (2-3)
Korfball	World Championship	November 2003 (9 days)	Indoor hall Ahoy	20,000	Finals (gold)
Baseball	World Championship	September 2005 (16 days)	Baseball stadium Neptunus	134,000	Semi-finals (4th place)
Sailing	Ocean Race	June 2006 (1 day)	City centre	500,000	Gold for a Dutch boat
Hockey	Men's World Champions Trophy	June 2008 (9 days)	Hockey stadium HC Rotterdam	40,000	Semi-finals (bronze)
Judo	World Championship	August 2009 (5 days)	Indoor hall Ahoy	26,900	1 gold, 2 silver
Cycling	Tour de France Grand Depart	July 2010 (4 days)	City centre	1000000	No medals
Gymnastics	World Championship	October 2010 (9 days)	Indoor hall Ahoy	46,100	1 silver
Table tennis	World Championship	May 2011 (8 days)	Indoor hall Ahoy	35,000	No medals
Equestrian sport*	European Championship Dressage	August 2011 (5 days)	Dressage arena Kralingse bos	55,000	2 gold, 1 bronze
Squash	World Championship	October-November 2011 (10 days)	Indoor hall Vitoria; Finals in the city centre	7,500	No medals
Volleyball*	Women's European Championship	September 2015 (9 days)	Indoor hall Ahoy	26,500	Finals (silver)
Rowing*	World Championship	August 2016 (9 days)	Willem-Alexander Rowing Complex	19,800	No medals
Short track*	World Championship	March 2017 (3 days)	Indoor hall Ahoy	25,000	2 gold, 1 silver, 1 bronze

* Included in sensitivity analyses

RESULTS

Characteristics of sport events

Table 1 presents the description of 10 singular elite sport events hosted in the city of Rotterdam in the period 2000 to 2017, and the 4 events additionally included in sensitivity analyses. This included 9 WCs, 2 ECs, the Davis Cup semi-final tennis, the Ocean Race, and the Tour de France Grand Depart. The number of visitors ranged between 7,500 of the WC Squash in 2011 and 1 million spectators at the 2010 Tour de France Grand Depart. During these 14 events, 15 medals were won by the Dutch team of which 7 were golden medals.

Effect on sport participation

The pooled effect of the 10 events did not show any change in sport-specific participation 1-year after the event took place (0.2 %-point (95% CI: -0.3; 0.8)) (Figure 1). Heterogeneity across events was substantial (I^2 = 65%). Three events on cycling, gymnastics, and table tennis were followed by an increase in sport-specific participation 1-year after the event was organised, as compared to the pre-event period (Figure 2). The 2010 Tour de France Grand Depart had the largest impact. The number of cyclists increased by 3.0 %-point (95% CI: 1.1; 5.0) following the event, and remained higher the years after the event. The increase in the participation of gymnastics and table tennis did not sustain. To the contrary, participation in korfball changed by -0.9 %-point (95% CI: -1.7, -0.1) following the WC, and remained lower the years after the event. The results for events that were not followed by an increase in sport participation are presented in Supplemental Figure 1.

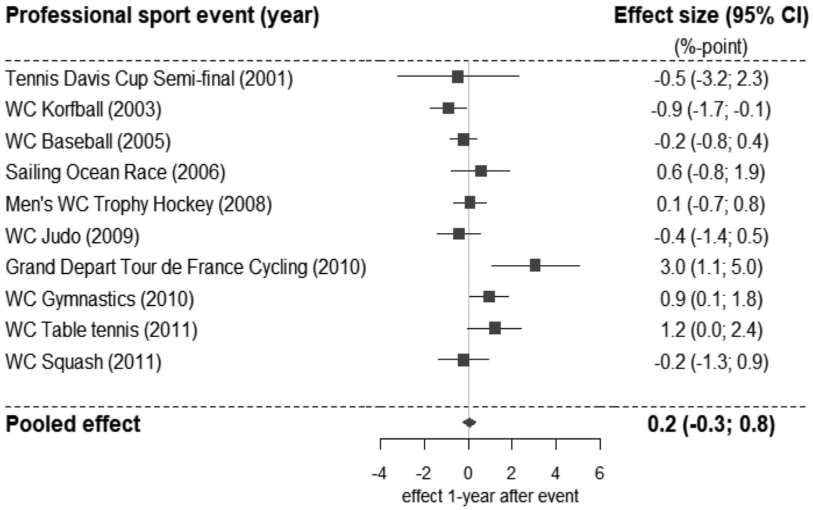


Figure 1: Forest plot of the effect of 10 elite sport events on sport-specific participation 1-year after the event took place

Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. Black squares indicate the point estimate for each event, with the horizontal lines representing the 95% CIs. The pooled effect of the 10 events was obtained using random-effects meta-analyses, with the effect of each event weighted by the inverse of its variance. The pooled effect is indicated with the diamond, and the width of the diamond corresponds with the 95% CI of the pooled effect.
Abbreviations: World Championship (WC); European Championship (EC).

Group differences by aim of events

There was a significant difference between events according to the period in which they were organised (Table 2). Events organised from 2010 onwards in which the more explicit aim was to target sport participation were followed by an increase in sport-specific participation of 1.1 %-point (95% CI: 0.0; 2.1), whereas no change was observed for events organised in the period without this specific aim (-0.3 %-point (95% CI: -0.6; 0.1)). Substantial heterogeneity remained for events organised after 2010 (I^2 = 69%). The change in sport participation did not differ by number of visitors or location of the event.

The result of the falsification test indicated that the increase in sport participation seen for events organised after 2010 was not reflecting the increase in sport participation seen in the population (Supplemental Figure 2).

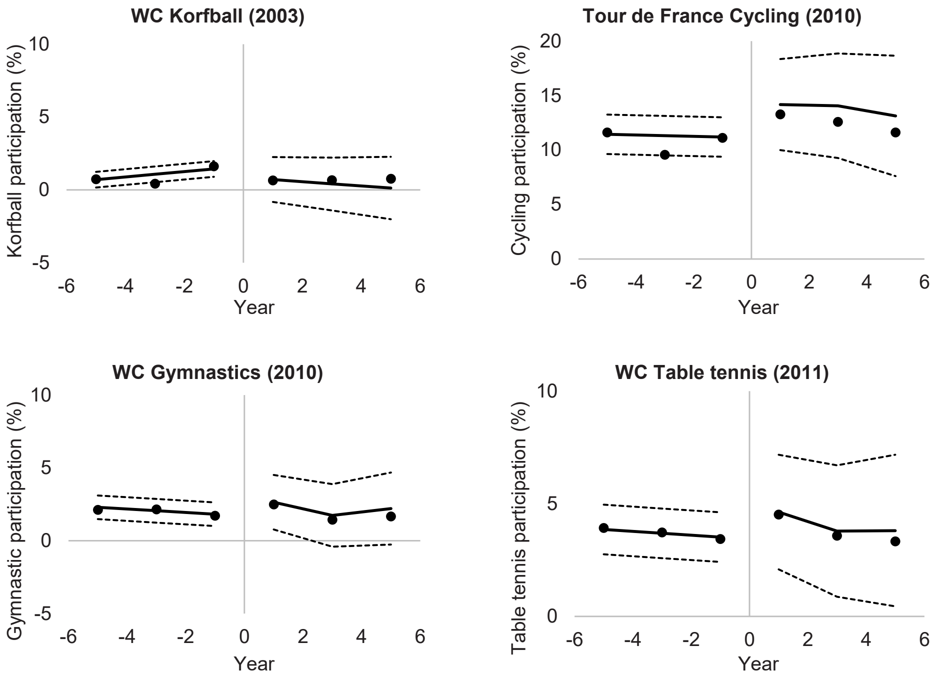


Figure 2: The sustained effect on sport-participation (% in the population) of events that significantly changed sport-participation 1-year after the event was organised

The population estimates (points) and results of the interrupted time series analyses (lines) are given. Dashed line represents the 95% CI around the estimate. Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. Year = 0 indicates the year that the event took place.
Abbreviations: World Championship (WC); European Championship (EC).

Sensitivity analyses

The findings were robust to sensitivity analyses with inclusion of 4 events with fewer data (Supplemental Figure 3 and 4, Supplemental Table 2). Excluding single events indicated that the Tour de France Grand Depart and the WC Korfball contributed most to the findings. The pooled effect of all events together slightly increased after exclusion of the korfball event (0.4 %-point (95% CI: -0.2; 0.9)). The main finding that events organised from 2010 onwards were followed by an increase in sport-specific participation slightly attenuated after exclusion of the cycling event (0.7 %-point (95% CI: -0.1; 1.5)) (Supplemental Table 3).

Table 2: The effect of elite sport events on sport-specific participation, stratified by indicators related to the policies that aimed to increase sport participation following events

	Events (n)	Sport-specific participation (95% CI) (%-point)	Heterogeneity (I^2 (%))	Group difference (P-value)
Targeting sport participation				0.039
Yes (after 2010)	4	1.1 (0.0; 2.1)	69	
No (before 2010)	6	-0.3 (-0.6; 0.1)	0	
Number of visitors				0.19
≥37500	5	0.6 (-0.2; 1.4)	73	
<37500	5	-0.2 (-1.0; 0.6)	54	
Location of the event				0.34
City centre	3	1.0 (-0.8; 2.8)	78	
Stadium in outskirts	7	0.0 (-0.5; 0.6)	62	

Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. All effect estimates were pooled effects obtained from random-effects meta-analysis, weighted by the inverse of its variance.

DISCUSSION

Three out of the 10 events were followed by an increase in sport-specific participation 1-year after the event was organised, and one was followed by a decrease. Only for the cycling event, the number of cyclists remained elevated up to 5-years following the event. The pooled effect of the 10 events, however, did not give evidence that sport-specific participation changed. Subgroup analyses suggested that only for events organised in the period in which the more explicit aim was to target recreational sport participation, an increase in sport-specific participation in the host population was observed. The result of the falsification test increased our confidence that we were capturing the effect of hosting elite sport events rather than time trends in sport participation.

The finding that sport events with the more explicit aim to target sport participation were followed by an increase in sport-specific participation is a novel finding. Existing evidence showed little impact of multi-sport events such as the Olympic and Paralympic Games on sport participation and physical activity (16-18, 23). Mega-events may be less (financially) accessible by the host population. It is promising that sport-specific participation rates increased after organizing sport events, but longitudinal data following individuals over time are needed to identify if people increased their physical activity levels.

The ORiEL study is the only study to date that evaluated the sport legacy of the London Games in 2012 by following people over time, but did not find evidence of an impact on adolescents physical activity, mental health or well-being (23). Qualitative research revealed that health was less of a priority in relatively deprived areas in which the study was conducted than more immediate concerns such as employment, safety and housing. This indicates that not all subgroups of the population may be able to benefit from sport events, and additional research is warranted to identify population subgroups that are impacted from hosting sport events.

In our study, events targeting cycling, gymnastics and table tennis were followed by an increase in sport participation among adults. Most notably, we found that the number of cyclists remained higher up to 5-years following the event. Possible reasons are that a large-scale awareness campaign started 100-days before the event took place, reaching a large part of the population, and the fact that it is relatively easy and safe to start recreational cycling in the Netherlands (35). Cycling also fits within the current era in which people prefer to sport individually (36). Interestingly, the routing of the time trail elites cycled in 2010 is still clearly signed in the city centre. Finally, other major cycling events were organised after 2010 in the Netherlands. The continued attention of the cycling sport in the media may have contributed to the sustained increase in cyclists.

For the other two events, an increase in sport participation was observed only directly following the gymnastic and table tennis event. Prolonged promotional activities may be needed to achieve sustained behavioural change. Furthermore, it may require more actions to increase sport participation for sports that are less accessible, and less embedded within the sports culture of a population. Further research is needed to identify what type of sport events have the highest potential to boost sport participation in the population.

The organisation of the WC korfball decreased sport participation. A possible explanation is that so-called "discouragement effects" may have appeared if the performance levels were perceived as unattainable (5). However, this may have occurred for all elite sport events under study, and therefore unlikely explains the contradictory finding. Another possibility is that the image of korfball decreased after the WC, because of the Dutch team winning all matches with large results. Alternatively, some measurement error could be present. We observed an increase in any type of sport in the year 2003, most likely due to random sampling error. Therefore, measurement error is a likely explanation of the finding that sport participation decreased following the WC korfball hosted in 2003.

A systematic review by Weeds and colleagues of 21 studies evaluating trickle-down effects resulting from the inspiration by elite sports, professional athletes and sport events concluded that a trickle-down effect was unlikely to engage new persons into sport activities, but that those already active may have increased their sport frequency or switched from activity (7). In our study, we lacked information on frequency of sport participation. The review also indicated that the trickle-down effect will mostly influence people with positive attitudes towards sport activities (7). We found the largest impact on sport participation levels following the Tour de France Grand Depart. It is likely that Dutch citizens have a positive attitude towards cycling given that cycling is one of the main modes of transportation (35, 37).

The inspirational function of elite performances, and the presence of professional athletes in the media when hosting events have also been suggested to be of importance. To date, there is only limited evidence of the trickle-down effect through elite success (5, 7, 38-40). In our study, we could not evaluate whether success by Dutch athletes in events hosted in the city of Rotterdam contributed to a larger increase in sport participation in the host population. In the current study, we could only compare performances across sport disciplines. The expected gold medal at the WC korfball have had little impact on korfball participation in the population, while some increases in gymnastics were seen following the unexpected silver medal during the WC gymnastics. Furthermore, we see that the Dutch team performed well in most events, which is not surprising given that decisions to host an event to some extent will rely on perceived importance and interest to the general population. The selection of events mirrors the events where typically medals are won by Dutch athletes during international tournaments. Without interest by the general population, there will be minimal economic spin-off, city marketing and changes in recreational sport participation in the population.

Strengths and limitations

This study was strengthened by the inclusion of multiple events, and we summarised the findings to generalise the results. The use of an interrupted time series design allowed to account for the pre-event trends. Sport participation estimates were taken from a single survey always conducted in September, which increases comparability of the estimates found. Sport participation was not restricted to memberships of sport clubs, but included all self-reported physical activity behaviours over the past year.

This study also had some limitations. There was no information on the frequency and duration of the sports performed. Ideally, a comparison group was included to control for time-varying factors that may have affected sport participation, but there were no data available for other cities that covered a similar period. A difference-in-difference design using sport-specific participation rates for sports that were not targeted by an event may be an alternative, especially when focusing on single events. However, as we studied multiple events combined the selection of a single control group for multiple events would not be possible. Furthermore, we could not use major sports as a control, since events were related to major sports. In absence of a control group, we conducted a falsification test. The result from the falsification test strengthened our believe that we are capturing the effect of hosting elite sport events. We selected 3 data points to correct for the pre-event trend, but more data points are needed to be more precise. Sport participation data were collected biannually, thus short term changes could not be detected. The decline in response may have influenced external validity, but the influence on sport participation in the population is expected to be small. In our analyses we focused on adults, however, earlier research indicated that children and adolescents might have been a better population for promoting physical activity following sport events (41, 42). We had no information on intermediate outcomes or on the mechanism how the events contributed to changes in sport participation.

Future recommendations

In this study we found some evidence that a city with a clear ambition to target sport participation via the hosting of elite sport events may influence sport participation among the general population. A careful selection of the type sport activity promoted through hosting elite sport event seems important, given that not all events were followed by an increase in sport participation. The increase is likely to occur among people with a positive attitude towards sport in general, or towards the specific sport activity promoted through the event. Furthermore, the sport infrastructure in a city must allow for a potential increase in sport participation. Funding for the development of programmes and side-events targeting sport participation before and after the event is critical to ensure a sustained increase in sport participation. It is important to monitor potential changes in the city that influence sport development and participation to inform future policy-making. Longitudinal data following individuals over time are needed to provide evidence whether more people become physically active, or that people shift between sports. It is essential to identify the type of activities needed to reach specific subgroups of the population, and to identify the elements of these activities that contribute to increasing sport participation.

CONCLUSION

Hosting elite sport events that specifically target sport participation in the general population may increase levels of sport participation among citizens. Not all events were followed by an increase in sport participation, therefore the type of event needs to be carefully selected. To capture the sport legacy of future elite sport events, it is important to study whether inactive people take up sport activities, if those already active increases the frequency of sport activities, or that people switch between sports.

REFERENCES

1. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. BMJ. 2016;354.
2. Eime RM, Harvey JT, Charity MJ, Casey MM, van Uffelen JGZ, Payne WR. The contribution of sport participation to overall health enhancing physical activity levels in Australia: a population-based study. BMC public health. 2015;15:806.
3. Hover P, Dijk B, Breedveld K, van Eekeren FJA, Slender H. Creating social impact with sport events. Utrecht, the Netherlands: Mulier Institute & Utrecht University; 2016.
4. Grix J, Carmichael F. Why do governments invest in elite sport? A polemic. International Journal of Sport Policy and Politics. 2012;4(1):73-90.
5. Hogan K, Norton K. The 'Price' of olympic gold. Journal of Science and Medicine in Sport. 2000;3(2):203-18.
6. Boardley ID. Can viewing London 2012 influence sport participation? – a viewpoint based on relevant theory. International Journal of Sport Policy and Politics. 2013;5(2):245-56.
7. Weed M, Coren E, Fiore J, Wellard I, Chatziefstathiou D, Mansfield L, et al. The Olympic Games and raising sport participation: a systematic review of evidence and an interrogation of policy for a demonstration effect. European Sport Management Quarterly. 2015;15(2):195-226.
8. Veal AJ, Toohey K, Frawley S. The sport participation legacy of the Sydney 2000 Olympic Games and other international sporting events hosted in Australia. Journal of Policy Research in Tourism, Leisure and Events. 2012;4(2):155-84.
9. Berridge G. The promotion of cycling in London: The impact of the 2007 Tour de France Grand Depart on the image and provision of cycling in the capital. Journal of Sport & Tourism. 2012;17(1):43-61.
10. Ramchandani G, Coleman Richard J, Bingham J. Sport participation behaviours of spectators attending major sports events and event induced attitudinal changes towards sport. International Journal of Event and Festival Management. 2017;8(2):121-35.
11. Ramchandani G, Coleman R, Christy E. The sport participation legacy of major events in the UK. Health Promotion International. 2017;34(1):82-94.
12. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. The American psychologist. 1992;47(9):1102.

13. Ajzen I, Madden TJ. Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. Journal of Experimental Social Psychology. 1986;22(5):453-74.

14. Potwarka LR. Exploring Physical Activity Intention as a Response to the Vancouver Olympics: an Application and Extension of the Theory of Planned Behavior. Event Management. 2015;19(1):73-92.

15. Bauman A, Bellew B, Craig CL. Did the 2000 Sydney Olympics increase physical activity among adult Australians? British Journal of Sports Medicine. 2015;49(4):243-7.

16. Weed M, Coren E, Fiore J. A systematic review of the evidence base for developing a physical activity and health legacy from the London 2012 Olympic and Paralympic Games. Canterbury: Centre for Sport, Physical Education and Activity Research (SPEAR), Canterbury Christ Church University, 2009.

17. McCartney G, Thomas S, Thomson H, Scott J, Hamilton V, Hanlon P, et al. The health and socioeconomic impacts of major multi-sport events: systematic review (1978-2008). BMJ. 2010;340:c2369.

18. Annear MJ, Shimizu Y, Kidokoro T. Sports mega-event legacies and adult physical activity: A systematic literature review and research agenda. Eur J Sport Sci. 2019;19(5):671-85.

19. Preuss H. The Conceptualisation and Measurement of Mega Sport Event Legacies. Journal of Sport & Tourism. 2007;12(3-4):207-28.

20. Chappelet J-L. Mega sporting event legacies: a multifaceted concept. Papeles de Europa. 2012(25):76-86.

21. Solar O, Irwin A. A conceptual framework for action on the social determinants of health. Geneva, Switzerland: WHO Document Production Services; 2010.

22. Smith NR, Clark C, Fahy AE, Tharmaratnam V, Lewis DJ, Thompson C, et al. The Olympic Regeneration in East London (ORiEL) study: protocol for a prospective controlled quasi-experiment to evaluate the impact of urban regeneration on young people and their families. BMJ Open. 2012;2(4):e001840.

23. Cummins S, Clark C, Lewis D, Smith N, Thompson C, Smuk M, et al. The effects of the London 2012 Olympics and related urban regeneration on physical and mental health: the ORiEL mixed-methods evaluation of a natural experiment. 2018;6:12.

24. Thomson A, Cuskelly G, Toohey K, Kennelly M, Burton P, Fredline L. Sport event legacy: A systematic quantitative review of literature. Sport Management Review. 2019;22(3):295-321.

25. Mahtani KR, Protheroe J, Slight SP, Demarzo MMP, Blakeman T, Barton CA, et al. Can the London 2012 Olympics ‘inspire a generation’ to do more physical or sporting activities? An overview of systematic reviews. BMJ Open. 2013;3(1):e002058.

26. City of Rotterdam. Ten-points-plan Rotterdam Sportscity. (in Dutch: Tienpuntenplan Rotterdam Sportstad). Rotterdam, the Netherlands 2001.
27. City of Rotterdam. Sportsnote Rotterdam 2016. (in Dutch: Sportnota Rotterdam 2016). Rotterdam, the Netherlands 2009.
28. City of Rotterdam. Sport moves Rotterdam: Sportsnote Rotterdam 2017-2020. (in Dutch: Sport beweegt Rotterdam: Sportnota Rotterdam 2017-2020). Rotterdam, the Netherlands 2016.
29. Rotterdam Festivals. Recalibration of event policies: Deeper into the city, furher into the world. (in Dutch: Herijking evenementenbeleid: Dieper in de stad, verder in de wereld). Rotterdam, the Netherlands: City of Rotterdam; 2010.
30. de Groot M, Blom S, van der Gugten M. Get the most from sport events. Evaluation policyframework pilots sport events. (in Dutch: Meer halen uit sportevenementen. Evaluatie VWS beleidskader pilots sportevenementen). Amsterdam, the Netherlands: DSP-groep BV; 2012.
31. Statistics Netherlands. Annual Report on Integration 2010 (in Dutch: Jaarrapport Integratie 2010). Den Haag/Heerlen, the Netherlands 2010.
32. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. Psychological Methods. 1998;3(4):486-504.
33. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine. 2002;21(11):1539-58.
34. Peirce RS, Frone MR, Russell M, Cooper ML. Relationship of financial strain and psychosocial resources to alcohol use and abuse: The mediating role of negative affect and drinking motives. Journal of Health and Social Behavior. 1994;35(4):291-308.
35. Pucher J, Buehler R. Making Cycling Irresistible: Lessons from the Netherlands, Denmark and Germany. Transport Reviews. 2008;28(4):495-528.
36. van den Dool R. Sports participation and physical activity without others increased (in Dutch: Zonder anderen sporten en bewegen neemt toe). Utrecht, the Netherlands: Mulier Institute; 2019.
37. Fishman E, Schepers P, Kamphuis CBM. Dutch Cycling: Quantifying the Health and Related Economic Benefits. American Journal of Public Health. 2015;105(8):e13-e5.
38. De Bosscher V, Sotiriadou P, van Bottenburg M. Scrutinizing the sport pyramid metaphor: an examination of the relationship between elite success and mass participation in Flanders. International Journal of Sport Policy and Politics. 2013;5(3):319-39.
39. Wicker P, Frick B. The inspirational effect of sporting achievements and potential role models in football: a gender-specific analysis. Managing Sport and Leisure. 2016;21(5):265-82.

40. Frick B, Wicker P. The trickle-down effect: how elite sporting success affects amateur participation in German football. Applied Economics Letters. 2016;23(4):259-63.

41. Carter RV, Lorenc T. A qualitative study into the development of a physical activity legacy from the London 2012 Olympic Games. Health Promotion International. 2013;30(3):793-802.

42. Wicker P, Sotiriadou P. The trickle-down effect: What population groups benefit from hosting major sport events. International journal of event management research. 2013;8(2):25-41.

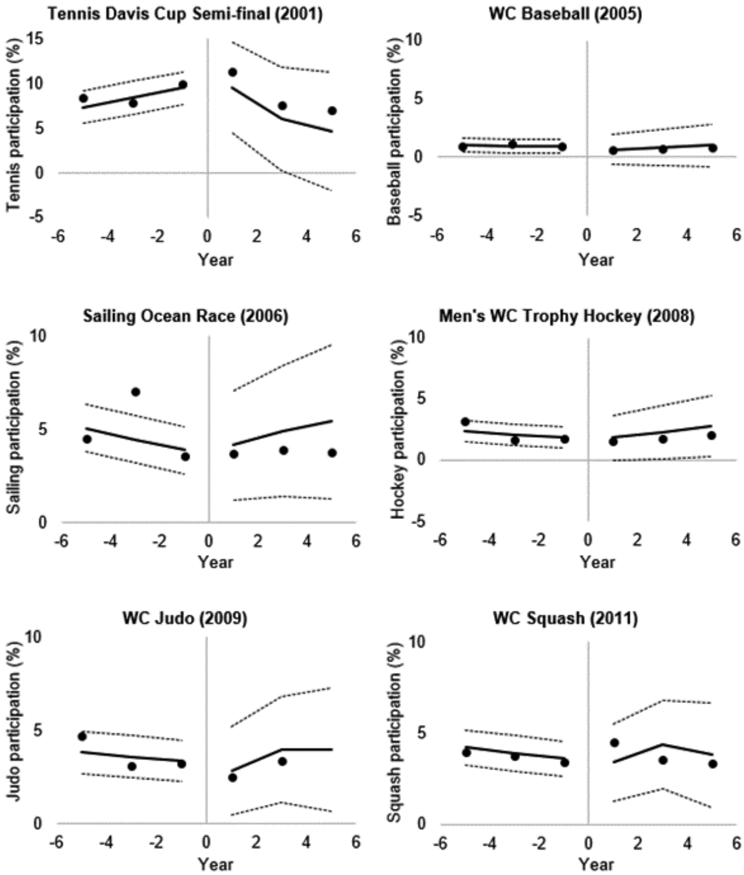
SUPPLEMENTAL MATERIAL

Supplemental Table 1	Population characteristics from 1995 to 2017 (n=27,235)
Supplemental Figure 1	The sustained effect on sport-participation (% in the population) of events that not significantly changed sport-participation 1-year after the event was organized
Supplemental Figure 2	Forest plot of the effect of 5 elite sport events on sport-specific participation 1-year after the event took place, in the hypothetical scenario that these events organized before 2010 would all have taken place in 2010
Supplemental Figure 3	The sustained effect on sport-participation (% in the population) of 4 events included for sensitivity analyses
Supplemental Figure 4	Forest plot of the effect of 14 elite sport events on sport-specific participation 1-year after the event took place
Supplemental Table 2	The effect of 14 elite sport events on sport-specific participation, stratified by indicators related to the policies that aimed to increase sport participation following events
Supplemental Table 3	The effect of 9 elite sport events (excluding the cycling event) on sport-specific participation, stratified by indicators related to the policies that aimed to increase sport participation following events
Appendix 1	Specification of the model

Supplemental Table 1: Population characteristics from 1995 to 2017 (n=27,235)

	1995	1997	1999	2001	2003	2005	2007	2009	2011	2013	2015	2017
Response (%)	46.8	43.5	54.0	48.6	53.8	42.8	36.7	20.1	33.1	23.0	20.5	14.8
Age, mean ± SD	41.0 ± 0.4	41.1 ± 0.5	41.4 ± 0.4	41.6 ± 0.4	39.6 ± 0.5	41.4 ± 0.3	42.4 ± 0.2	42.5 ± 0.3	42.3 ± 0.3	42.2 ± 0.3	42.8 ± 0.4	42.9 ± 0.4
Sex (%)												
Male	51.5	50.9	50.4	51.0	54.1	50.5	50.5	49.9	49.9	50.3	50.3	50.7
Female	48.5	49.1	49.6	49.0	45.9	49.5	49.5	50.1	50.1	49.7	49.7	49.3
Ethnicity (%)												
Western	77.2	75.8	73.4	70.5	73.7	66.7	65.0	65.0	62.9	63.4	62.2	62.6
Non-Western	22.8	24.2	26.6	29.5	26.3	33.3	35.0	35.0	37.1	36.6	37.8	37.4
Household income (%)												
Low	24.4	26.7	25.9	28.2	26.9	27.8	28.8	22.5	22.4	20.6	21.5	24.3
Mid-low	36.2	37.6	37.0	33.4	30.5	28.9	27.3	31.2	29.0	26.6	24.6	25.7
Mid-high	23.7	20.1	19.5	22.2	17.6	23.8	23.5	24.3	24.3	34.0	32.2	26.3
High	15.6	15.6	17.7	16.2	25.0	19.5	20.4	22.0	24.2	18.9	21.6	23.7

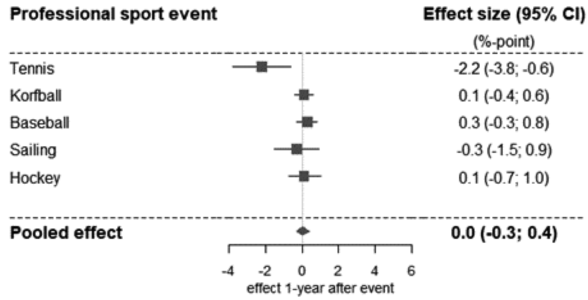
Estimates are weighted for age, sex, and ethnicity per city district to reflect the population of that year



4

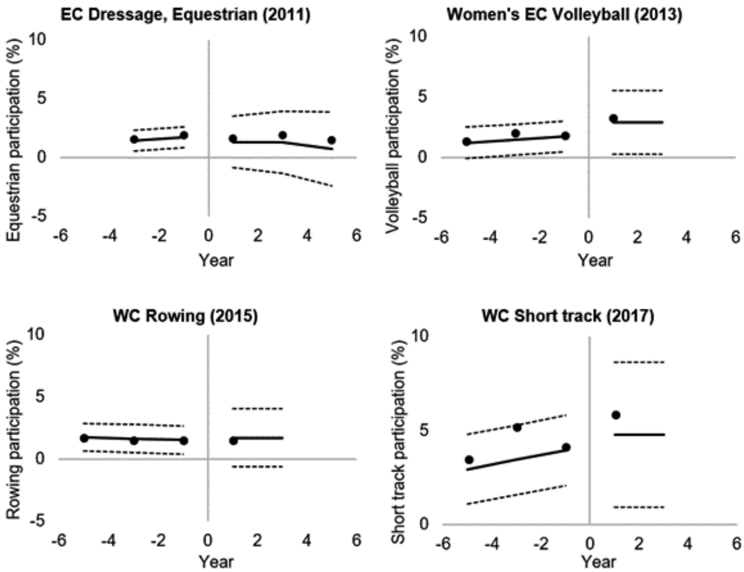
Supplemental Figure 1: The sustained effect on sport-participation (% in the population) of events that not significantly changed sport-participation 1-year after the event was organized

The population estimates (points) and results of the interrupted time series analyses (lines) are given. Dashed line represents the 95% CI around the estimate. Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. Year = 0 indicates the year that the event took place.



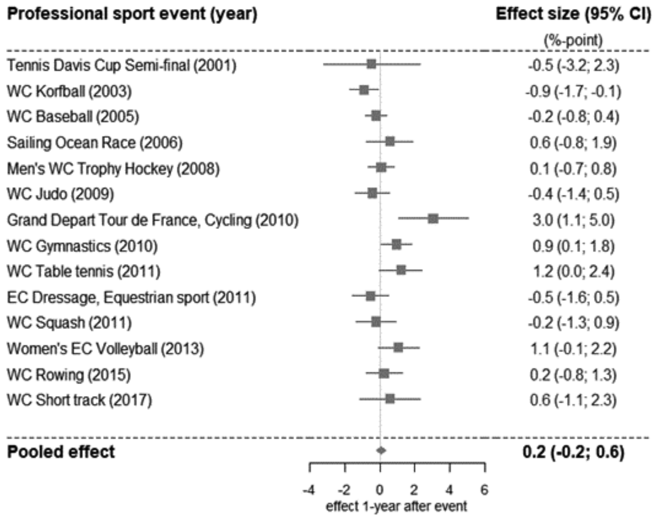
Supplemental Figure 2: Forest plot of the effect of 5 elite sport events on sport-specific participation 1-year after the event took place, in the hypothetical scenario that these events organized before 2010 would all have taken place in 2010

Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. Black squares indicate the point estimate of sport-specific participation 1-year after the hypothetical event took place for each event, with the horizontal lines representing the 95% CIs. The pooled effect of the 5 events was obtained using random-effects meta-analyses, with the effect of each event weighted by the inverse of its variance. The pooled effect is indicated with the diamond, and the width of the diamond corresponds with the 95% CI of the pooled effect.



Supplemental Figure 3: The sustained effect on sport-participation (% in the population) of 4 events included for sensitivity analyses

The population estimates (points) and results of the interrupted time series analyses (lines) are given. Dashed line represents the 95% CI around the estimate. Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. Year = 0 indicates the year that the event took place.



Supplemental Figure 4: Forest plot of the effect of 14 elite sport events on sport-specific participation 1-year after the event took place

Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. Black squares indicate the point estimate for each event, with the horizontal lines representing the 95% CIs. The pooled effect of the 14 events was obtained using random-effects meta-analyses, with the effect of each event weighted by the inverse of its variance. The pooled effect is indicated with the diamond, and the width of the diamond corresponds with the 95% CI of the pooled effect.
Abbreviations: World Championship (WC); European Championship (EC).

Supplemental Table 2: The effect of 14 elite sport events on sport-specific participation, stratified by indicators related to the policies that aimed to increase sport participation following events

	Events (n)	Sport-specific participation (95% CI) (%-point)	Heterogeneity (I^2 (%))	Group difference (P-value)
Targeting sport participation				0.021
Yes (after 2010)	8	0.6 (0.0; 1.2)	52	
No (before 2010)	6	-0.3 (-0.6; 0.1)	0	
Number of visitors				0.57
≥37500	6	0.4 (-0.3; 1.1)	71	
<37500	8	0.1 (-0.5; 0.7)	49	
Location of the event				0.36
City center	3	1.0 (-0.8; 2.8)	78	
Stadium in outskirts	11	0.1 (-0.3; 0.5)	51	

Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. All effect estimates were pooled effects obtained from random-effects meta-analysis, weighted by the inverse of its variance.

Supplemental Table 3: The effect of 9 elite sport events (excluding the cycling event) on sport-specific participation, stratified by indicators related to the policies that aimed to increase sport participation following events

	Events (n)	Sport-specific participation (95% CI) (%-point)	Heterogeneity (I^2 (%))	Group difference (P-value)
Targeting sport participation				0.044
Yes (after 2010)	3	0.7 (-0.1; 1.5)	41	
No (before 2010)	6	-0.3 (-0.6; 0.1)	0	
Number of visitors				0.35
≥37500	4	0.2 (-0.3; 0.8)	43	
<37500	5	-0.2 (-1.0; 0.6)	54	
Location of the event				0.85
City center	2	0.1 (-0.7; 1.0)	0	
Stadium in outskirts	7	0.0 (-0.5; 0.6)	62	

Analyses were adjusted for the pre-event trend and age, sex, income and ethnicity. All effect estimates were pooled effects obtained from random-effects meta-analysis, weighted by the inverse of its variance.

Appendix 1: Specification of the model

Pre-event data were graphically examined for linearity. Linear multivariable regression models were used to obtain the absolute change in sport-specific participation following events, expressed as percentage point change. The models were run separately for each event and specified as:

$$Y_{it} = \beta_0 + \beta_1 (T_{(pre-trend)}) + \beta_2 (T_1) + \beta_3 (T_3) + \beta_4 (T_5) + Z_i + \varepsilon_{it}$$

Where Y_{it} is a dichotomous outcome variable indicating whether individual *i* participated in the sport of interest in year t, β_0 is the intercept and represent the baseline sport participation level at $T = 0$, β_1 is the pre-event trend in sport participation for the 5 years before the event, β_2, β_3, and β_4 are dummy variables indicating the level change in sport participation for 1, 3, and 5 years after the event, Z_i is a vector of confounders (age, sex, income and ethnicity), and ε_{it} accounts for random sources of error. The pre-event trend and the absolute change in sport-specific participation following events were visualised using the weighted estimate of sport-specific participation at $T = 0$ as the intercept.

5

Impact of expanding smoke-free policies beyond enclosed public places and workplaces on children's tobacco smoke exposure and respiratory health: protocol for a systematic review and meta-analysis

Martá K. Radó, Famke J.M. Mölenberg, Aziz Sheikh, Christopher Millett, Wichor M. Bramer, Alex Burdorf, Frank J. van Lenthe, Jasper V. Been

BMJ Open. 2020;10(10):e038234

ABSTRACT

Introduction

Tobacco smoke exposure (TSE) has considerable adverse respiratory health impact among children. Smoke-free policies covering enclosed public places are known to reduce child TSE and benefit child health. An increasing number of jurisdictions are now expanding smoke-free policies to also cover outdoor areas and (semi)private spaces (indoor and outdoor). We aim to systematically review the evidence on the impact of these “novel smoke-free policies” on children’s TSE and respiratory health.

Methods

13 electronic databases will be searched by two independent reviewers for eligible studies. We will consult experts from the field and hand-search references and citations to identify additional published and unpublished studies. Study designs recommended by the Cochrane Effective Practice and Organisation of Care (EPOC) group are eligible, without restrictions on the observational period, publication date, or language. Our primary outcomes are: self-reported or parental-reported TSE in places covered by the policy; unplanned hospital attendance for wheezing/asthma; and unplanned hospital attendance for respiratory infections. We will assess risk of bias of individual studies following the EPOC or Risk Of Bias In Non-randomised Studies of Interventions tool, as appropriate. We will conduct separate random-effects meta-analyses for smoke-free policies covering (1) indoor private places, (2) indoor semiprivate places, (3) outdoor (semi)private places, and (4) outdoor public places. We will assess whether the policies were associated with changes in TSE in other locations (e.g., displacement). Subgroup analyses will be conducted based on country income classification (i.e., high, middle or low income) and by socioeconomic status. Sensitivity analyses will be undertaken via broadening our study design eligibility criteria (i.e., including non-EPOC designs) or via excluding studies with a high risk of bias. This review will inform policymakers regarding the implementation of extended smoke-free policies to safeguard children’s health.

Ethics and dissemination

Ethical approval is not required. Findings will be disseminated to the academics and the general public.

PROSPERO registration number: CRD42020190563

INTRODUCTION

Tobacco smoke exposure (TSE) is an important cause of adverse respiratory health outcomes in children worldwide. TSE includes both second-hand smoke (SHS) and third-hand smoke (THS) exposure. SHS refers to the inhalation of emitted smoke, while THS exposure refers to the uptake of tobacco smoke residuals from polluted surfaces after someone has finished smoking. TSE is estimated to cause 166,000 child deaths annually, and is associated with an increased risk of respiratory tract infections (RTI), wheezing, and asthma (1, 2).

Safeguarding of child health provides justification to advocate for strong tobacco control measures (3). Previous systematic reviews found that smoke-free policies in enclosed public places are successful in reducing adverse child health outcomes. Meta-analyses showed that these policies were followed by a 9.8% (95% confidence interval [CI]: 3.0 to 16.6) reduction in hospital attendance for asthma exacerbations, and an 18.5% (95% CI: 4.2 to 32.8) reduction in hospital attendance for lower RTIs (2). These health impacts of smoke-free policies are likely mediated through a reduction in SHS and potentially also in THS (1). Indeed, individual studies have demonstrated that smoke-free policies covering enclosed public areas are successful in decreasing child TSE in public places and even in private places, such as cars or homes, most likely via changes in social norms (4-8). Although in low and middle-income countries high background air pollution, poor economic conditions, and low awareness of tobacco-related harm might obscure the positive effects of smoke-free legislation, evidence suggests that smoke-free policies can have a similarly positive impact in these countries as in high-income countries (9, 10).

As a result, smoke-free legislation is increasingly recognised as an important policy tool to protect children from the adverse health effects of TSE (e.g., incorporated in Sustainable Development Goal 3.2, 3.4, 3.9, and 3.A that aim to improve health and well-being) (3). Recently, policies to provide additional protection of children to TSE beyond enclosed public places have been implemented in a number of places, both at a national and a subnational level. These include expansion to cover outdoor areas frequented by children such as school grounds, playgrounds and parks (11). Also, some countries and subnational regions have now implemented policies prohibiting smoking in semiprivate spaces such as shared housing or private spaces such as cars (12, 13).

The positive effects of smoke-free policies covering enclosed public places on child health are now well established (2, 14). A systematic synthesis of the evidence on the effectiveness of smoke-free policies covering outdoor and (semi)private spaces to reduce children's TSE and benefit their health is however currently lacking (2, 14). It is important to note that effectiveness of smoke-free policies covering indoor public places cannot be easily extrapolated to those covering outdoor or private spaces, due to various reasons (e.g., dilution of TSE in outdoor spaces, enforcement issues). Further, evidence from various countries indicates that smoke-free policies covering public enclosed places and workplaces are followed by reductions in TSE even in areas not covered by the policy through norm spreading. Whether novel smoke-free policies also have an impact on TSE in places not covered by the policy is unclear. Theoretically, there may be a reduction (via norm spreading) in such places, but there may also be displacement of smoking to areas not covered by the policy (13). Finally, effectiveness of smoke-free policies is likely to vary according to local rates of smoking and children's SHS exposure, and comprehensiveness of the policy as well as compliance and enforcement (15).

With an increasing number of institutions and governments now implementing smoke-free policies covering areas other than enclosed public spaces, optimal understanding of the effectiveness of such policies in protecting children is essential. To establish this evidence base, we will undertake a comprehensive systematic review of available studies assessing the impact of policies covering outdoor areas and (semi) private places (whether indoor or outdoor) on children's TSE and respiratory health. We believe this work will be instrumental to informing effective policymaking, both (inter)nationally and locally, to further protect children from the adverse effects of tobacco smoke.

METHODS

This protocol is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols (PRISMA-P) guideline for systematic review and meta-analysis protocols (16).

Patient and public involvement

Patients and the public were not involved in the design, development, conduct, reporting, or dissemination of this study.

Eligibility criteria

We will include studies based on pre-defined criteria as summarised in Table 1 and detailed in the text below (17).

I. Population

Studies are eligible when they include children between 0 and 16 years. Any study reporting (sub)populations in which at least 50% fits this age criterion will also be eligible. Some studies may have included specific subgroups, for example, children with asthma (18). We will include all studies that meet the inclusion criteria, irrespective of such additional restrictions in the populations.

II. Interventions

We will include studies that evaluated smoke-free policies that were introduced at any governmental level (e.g., cities, municipalities, regions, or countries) or any institutional level (e.g., school, or hospital). The included policies might concern the following areas: indoor private places (e.g., cars), indoor semiprivate places (e.g., multi-unit housing), outdoor (semi)private places (e.g., shared gardens), and outdoor public places (e.g., parks, school grounds, beaches, hospital grounds). We will exclude studies that solely evaluated policies covering enclosed public places (i.e., “traditional” smoke-free policies).

Smoke-free policies can be a part of a complex reform package (19). For example, the above described smoke-free policy might be introduced simultaneously with other tobacco control policies (e.g., increased taxes on tobacco products). We will include all studies that estimate the effect of smoke-free policies covering (semi)private or outdoor places irrespective of whether other policies were introduced at the same time. If the effect of the smoke-free policy covering (semi)private or outdoor places has been disentangled from other interventions, we will extract the effect estimates from the analyses that most closely reflects that of this smoke-free policy.

III. Comparison

We will include studies that estimate the counterfactual scenario (i.e., no change in smoke-free policy implemented) using any of the following comparators: (1) a comparison population where no change in the observed smoke-free policy occurred in the observational period, or (2) a comparison time period in which no change in the observed smoke-free policy occurred.

IV. Outcomes

We will include studies assessing the impact of the policy on two types of outcomes: (1) indicators of TSE and (2) respiratory health outcomes. We have specified a wide range of outcomes to obtain a comprehensive overview of the entirety of available evidence on the topic. We will in our interpretation focus primarily on a small set of primary outcomes next to a larger set of secondary outcomes. Any study that reported step changes (immediate change in incidence) and slope changes (gradual change in incidence over time) in one or more of these outcomes will be eligible.

Studies that report on the following TSE indicators will be included:

Primary outcome:

(1) TSE in places covered by the policy (as reported by the child, parent, or primary caregiver)

Secondary outcomes:

(1) TSE in places of which only some were covered by the policy, or in unspecified places (as reported by the child, parent, or primary caregiver)
(2) TSE in places not covered by the policy (as reported by the child, parent, or primary caregiver)
(3) Cotinine or other specific biomarkers of TSE quantified in body fluids, hair, nails, or on the skin
(4) TSE assessed by wearable devices

Note that the primary TSE outcome of interest assesses exposure specifically in places covered by the policy. That is, we aim to assess whether the policy was successful in reducing child TSE in those places that it was intended to cover. In addition, unintended effects of smoke-free policies on TSE have been described, both positive (spill-over; e.g., via increased adoption of voluntary home smoking bans following the implementation of smoke-free policy) and negative (displacement; e.g., via increased smoking at home following smoke-free vehicle regulation) (13). In order to identify such effects we will also identify studies that have assessed changes in TSE in places not covered by the policy.

Individual studies may report different estimates for TSE based on various definitions. We will prioritise child-reported TSE outcomes over parental or primary caregiver reported outcomes for studies reporting both. Furthermore, we will prioritise past-week TSE outcomes over longer or shorter recall periods. If past-week TSE is not available, longer recalls are prioritised over shorter recall periods.

The reduction in child TSE is anticipated to translate into child health benefits. Health outcomes are selected based on severity and responsiveness to changes in TSE within a reasonably short time frame (i.e., within one year), these primarily being respiratory outcomes (2, 14). Accordingly, we will include studies that report the following outcomes:

Primary outcomes:

(1) Unplanned hospital attendance for wheezing/asthma
(2) Unplanned hospital attendance for RTIs

Secondary outcomes:

(1) General incidence of wheezing/asthma
(2) General incidence of RTIs
(3) Otitis media with effusion (OME)
(4) Chronic cough
(5) Lung function (forced expiratory volume in one second (FEV1), forced vital capacity (FVC), FEV1/FVC ratio)

Table 1. Inclusion criteria based on PICOS strategy

PICOS	Inclusion criteria
Population	(1) 0-16 years old (2) (Sub)populations in which at least 50% is under 17 years of age
Intervention	Smoke-free policies instituted at any governmental or institutional level that restrict smoking in designated (semi)private places or any outdoor areas
Comparison	(1) A comparison population living in a location where no intervention was introduced or where the intervention did not change in the observational period (2) A comparison time period in which no intervention was introduced or the intervention did not change
Outcomes child TSE	Primary outcome: (1) TSE in places covered by the policy (as reported by the child, parent or primary caregiver) Secondary outcomes: (1) TSE in places of which only some were covered by the policy or in unspecified places (as reported by the child, parent, or primary caregiver) (2) TSE in places not covered by the policy (as reported by the child, parent, or primary caregiver) (3) Cotinine or other specific biomarkers of TSE quantified in body fluids, hair, nails, or on the skin (4) TSE assessed by wearable devices
Outcomes child health	Primary outcomes: (1) Unplanned hospital attendance for wheezing/asthma (2) Unplanned hospital attendance for RTIs Secondary outcomes: (1) General incidence of wheezing/asthma (2) General incidence of RTIs (3) OME (4) Chronic cough (5) Lung function (FEV1, FVC, FEV1/FVC ratio)
Study design	Included in the main analyses: (1) Randomised trials (2) Non-randomised trials (3) Interrupted time series (4) Controlled before-after studies Included in sensitivity analyses: (5) Prospective cohort studies (6) Retrospective cohort studies (7) Uncontrolled before-after studies

Abbreviations: carbon monoxide (CO); forced expiratory volume in one second (FEV1); forced vital capacity (FVC); otitis media with effusion (OME); Population, Intervention, Comparison, Outcomes, and Study design (PICOS); fine particulate matter (PM2.5); respiratory tract infection (RTI); tobacco smoke exposure (TSE).

Unplanned hospital attendance may include acute presentations to the accident and emergency department as well as hospital admissions. For the secondary outcomes, respiratory diseases may be based on different case definitions. For studies reporting estimates according to various definitions we will prioritise according to the following hierarchy: (1) based on physician diagnosis, (2) based on medication use (not applicable for OME), and (3) self-reported.

We will not include perinatal outcomes in this review as many of the policies for which we anticipate to identify evidence are specifically aimed at protecting children, not pregnant women (e.g., smoke-free policies in playgrounds).

V. Study designs

We will follow the guidelines of the Cochrane Effective Practice and Organisation of Care (EPOC) group to select eligible study designs (20). EPOC established criteria for high-quality methodologies in research. These include: (1) randomised trials, (2) non-randomised trials, (3) interrupted time series, and (4) controlled before-after studies. We will include studies that satisfy these criteria in our main analysis. Based on previous experiences, this restriction may sometimes lead to a low number of studies being eligible. In order to properly assess the available evidence in such instances, we will conduct sensitivity analyses by relaxing the inclusion criteria (2, 21). For the sensitivity analyses, we will additionally include (1) prospective cohort studies, (2) retrospective cohort studies, and (3) uncontrolled before-after studies.

VI. Further restrictions

There will be no restriction on publication date, the timeframe of the study, or the length of follow-up after the introduction of the policy. All databases will be searched from their date of inception, and we will update the search to supplement our review just before publication. We will include every study that meets the above-mentioned criteria regardless of the language. Studies published in languages other than English will be translated. Google Translate will be used for this purpose, and we will consult with a translator if necessary.

VII. Report type

Our systematic review will include published studies in scientific journals as well as "grey literature" (i.e., documents that are circulated in non-commercial academic channels or not indexed by major databases). Only full reports will be included in our review. Studies for which only an abstract was published will not be included since risk of bias for these studies cannot be adequately assessed. In these cases, however, we will contact the authors to ask if a full report is available.

Information sources

We will search for potential studies in the following electronic databases: (1) Embase.com, (2) Medline Ovid, (3) Web of Science, (4) PsycINFO Ovid, (5) CINAHL EBSCOhost, (6) Google Scholar, (7) IndMED, (8) KoreaMed, (9) EconLit, (10) WHO Global Health Library (including African Index Medicus [AIM], Latin America and the Caribbean Literature on Health Sciences [LILACS], Index Medicus for the Eastern Mediterranean Region [IMEMR], Index Medicus for South-East Asia Region [IMSEAR], Western Pacific Region Index Medicus [WPRIM]), (11) WHO Library Database [WHOLIS], (12) Scientific Electronic Library Online (SciELO), and (13) Paediatric Economic Database Evaluation (PEDE).

Search strategy

First, we will search for potentially relevant studies based on a search strategy that is the combination of MeSH terms and free text search. Our research team, including a librarian who is specialised in search strategy optimisation, has developed this search strategy. Search terms were tailored to each database. Search terms included four parts: (1) terms to identify smoke-free policies; (2) terms that identify children as the target population; (3) terms to identify asthma, wheezing, respiratory diseases, OME, chronic cough, lung function, or TSE as outcome; and (4) terms that exclude conference abstracts, letter to the editors, notes and editorials.

Second, we will search for additional studies by screening reference lists of included studies and their citations through Google Scholar. Moreover, we will contact with experts in the field to identify additional studies which may have been missed and any relevant ongoing or unpublished studies. Finally, we will update our search to add the most recent publications just before submitting the review for publication.

Study records

I. Data management

We will extract all records identified by the different sources into an EndNote Library, and use this software to automatically deduplicate the collected records according to the method previously described (22). Subsequently, we will manually identify and remove any duplicates that remain. At this stage, duplicates will be identified based on overlaps in the author names, the titles of the publication, the observed populations, the constructions of the treatment group and control groups, sample sizes, and the reported outcomes. In our final report, we will note the number of duplicates in the PRISMA flow diagram.

II. Selection process

After the deduplication process, we will select eligible papers from the remaining unique records. At the first step, two independent researchers will screen titles and the abstracts for eligibility. We will then obtain the full-text reports of studies that may fit eligibility criteria based on this assessment. At the next step, two independent researchers will assess eligibility based on the full-texts. Any discrepancies will be resolved after discussion with a third researcher. The involved researchers will not be blinded to information about the articles (e.g., author names and affiliations) at any stage.

III. Data collection process

Two researchers will independently extract data from the included studies to a customised data extraction form developed a *priori* that has been piloted using six eligible studies. After finishing the data extraction, they will confer their results with each other and create one final file. Again, any discrepancies will be resolved after discussion with a third researcher. We will contact authors for any relevant missing data.

In case overlapping populations are analysed in multiple studies, we will include according to the following hierarchy the study that (1) has the lowest risk of bias (see section *Risk of bias assessment*), (2) evaluates the most comprehensive policy, or (3) incorporates the largest sample size. In case one study reports multiple effect estimates for overlapping populations, we will select according to the following hierarchy the one derived from (1) the most adjusted model, (2) the longest observation period, (3) the most comprehensive policy, or (4) the largest treatment group.

Data items

From eligible studies we will extract the following items: (1) the first author's name, (2) his or her affiliation, (3) publication year, (4) type of publication (journal, book, dissertation etc.), (5) access information (URL or doi), (6) study design, (7) observational period (the beginning and end dates), (8) exact places covered by the policy, (9) whether the policy covers outdoor or indoor places, (10) whether the policy covers public or (semi)private places, (11) timing of policy, (12) institution or government that initiated the policy, (13) country or location where the policy had been implemented, (14) compliance with the policy, (15) enforcement of the policy, (16) eligibility criteria for inclusion, (17) description of control group(s), (18) population at risk, (19) number of participants or events, (20) control and treatment group size, (21) characteristics of the population and the treatment groups (e.g., age, gender, and socioeconomic status), (22) data source(s) used for the study, (23) definition(s) of outcome measure(s), (24) controlled confounders (if applicable), (25) applied statistical techniques to draw inference, (26) number of clusters (if applicable), (27) cluster size (if applicable), (28) whether the results were adjusted for clustering (if clustered study), (29) whether the intra-class correlation coefficient is reported and what it is (if clustered study), (30) the number of drop-outs and missing values, (31) techniques for handling missing values, (32) pre-intervention population at risk(n)/events(n)/rates(%) of outcome variable(s), (33) post-intervention population at risk(n)/events(n)/rates(%) of outcome variable(s), (34) association between smoke-free policy and outcome(s) (coefficients, confidence intervals, and p-values), (35) any unintended effects (e.g., TSE displacement), (36) proportion of children exposed to TSE in the location covered by the smoke-free policy, before and after the implementation of the policy, (37) risk ratios of respiratory disease for different levels of TSE according to specific locations covered by the smoke-free policy, (38) bias assessment (see section *Risk of bias assessment*), (39) elements supporting causal inference (see section *Elements supporting causal inference*), (40) any conflict of interest reported by the authors, and (41) the funding source(s).

Outcomes and prioritisation

As previously detailed, we will identify studies evaluating the effects of smoke-free policies on children's (1) TSE and (2) health outcomes. See the description of outcomes and our prioritisation in the *Eligibility criteria* section.

Risk of bias assessment

For each individual result, we will assess risk of bias using standardised assessment tools. For randomised trials, we will use the Quality Assessment Tool for Quantitative Studies, developed by EPOC (23). For non-randomised studies, we will use the Risk Of Bias In Non-randomised Studies of Interventions (ROBINS-I) tool (24). The ROBINS-I tool evaluates biases in the following domains: (1) bias due to confounding (see Appendix 1 for additional criteria), (2) bias in selection of participants into the study, (3) bias in classification of interventions, (4) bias due to deviations from intended interventions (see Appendix 2), (5) bias due to missing data, (6) bias in measurement of outcomes, and (7) bias in selection of the reported result. For each domain, we will rate the risk of bias on a 4-point scale ranging from critical risk of bias to low risk of bias (i.e., bias is comparable to a well-performed randomised trial). The ROBINS-I tool will be followed to give an overall estimate of the risk of bias. The risk-of-bias assessments will be conducted independently by two researchers and a third reviewer will be consulted to resolve disagreements.

Elements supporting causal inference

Based on information reported in the included studies, we will reflect on elements that may support causal inference and the robustness of the evidence using the UK Medical Research Council guidance on natural experiments (25). First, we will report whether the effect estimates were robust when using alternative comparison groups (e.g., the use of different populations living in distinct geographical areas as a comparison). Second, we will report whether a study estimated whether the smoke-free policy also had an "effect" on neutral outcomes that were not expected to change as a consequence of the intervention (e.g., hospital admission for appendicitis). Finally, we will report from the included studies any additional information from complementing research methodologies regarding possible underlying mechanisms supporting the quantitative study findings. These assessments will be based on the quantitative and qualitative information presented in the individual study reports. We anticipate that this information cannot be uniformised across studies, and as such will be reported narratively.

Data synthesis

I. Synthetisation of data

Obtaining comparable data is essential to facilitate meta-analysis. Studies should be sufficiently similar in terms of their designs and the definitions of their measurements. We will uniformise outcome assessments before conducting data analysis. First, if continuous outcomes appear on different scales across studies we will use standardised mean differences. If continuous outcomes appear on the same scale, weighted mean differences will be used. Second, for dichotomous outcomes we will extract relative risks. If only absolute risks are presented then we will calculate relative risks. Third, we will express the effects on dichotomous outcomes using risk ratio (RR). If RR was not reported, we will contact the authors to provide it to us. If the requested data was not be obtained, we will convert odds ratio (OR) to RR using the following formula:

$$RR = \frac{OR}{(1 - PEER) + (PEER \times OR)}$$

, where $PEER$ refers to the patient-expected event rate in the control group (26). In case $PEER$ is not available, we will approximate this measure using the overall event rate across the entire study population. We will use incidence rate ratios (IRRs) instead of RR for outcomes that could occur repeatedly with the same individual (i.e., hospitalisation). To make different TSE measures comparable, we will convert them into a binary variable with a value of 1 if the child was exposed to (detectable amounts of) TSE and 0 if unexposed to (detectable amounts of) TSE. Finally, two researchers will check independently whether the definitions of intervention, population, and outcomes are sufficiently comparable across the selected studies to allow meta-analysis.

The eligible studies are likely to be heterogeneous in various important characteristics which will be taken into account in the data synthesis. Therefore we will apply random-effects models (as opposed to fixed-effect) for data analysis of two or more studies with similar policies and outcome measure. Furthermore, it might be needed to take into account the dependency of observations, for example, if multiple effect estimates for similar policies across various regions were provided within a single study. A three-level meta-analysis will be considered if multiple estimations are extracted from the same study. The decision to perform a 2-level or 3-level meta-analysis will be based on the model performance using a one-sided log-likelihood-ratio test (27, 28). We will separately pool step changes (immediate changes) and slope changes (gradual changes) for each outcome in different models. If unspecified, the result will be considered a step change. We will conduct separate meta-analyses for various subtypes of smoke-free policies, according to the locations that they cover. Thus we will separately pool studies for (1) smoke-free policies covering indoor private places (e.g., cars), (2) smoke-free policies covering indoor semiprivate places (e.g., multi-unit housing), (3) smoke-free policies covering outdoor private or semiprivate spaces (e.g., shared gardens), and (4) smoke-free policies covering outdoor public places (e.g., parks, school grounds, beaches, and hospital grounds). Finally, regarding TSE we will pool separate models to assess unintended effects of the policy in places not covered by the policy.

The estimations of our meta-analysis will be displayed in forest plots. Heterogeneity will be quantified by the I^2 statistic for each pooled model.

We anticipate that most studies will have assessed the impact of novel smoke-free policies on TSE rather than on health outcomes. As the relationship between TSE and respiratory outcomes among children is well established (29, 30), we will use data from studies estimating the impact of novel smoke-free policies on child TSE to quantify the potential impact of these changes on our primary health outcomes. Thus, we will perform a health impact assessment and quantify the proportion of children with respiratory disease that can be prevented by the change in TSE following smoke-free policies. First, we will extract the RR of respiratory disease for those children who were exposed to TSE versus those who were not either from the studies included in our systematic review or from previous meta-analyses that quantified this association (29, 30). Preferably, this RR will be retrieved for the specific locations that are covered by the smoke-free policy and for distinct respiratory diseases. Second, based on the RRs and the difference in the proportion of children exposed to TSE before and after the policy implementation, we will calculate two population attributable fractions

(PAFs). One quantifies the proportion of each outcome that can be attributed to TSE before the smoke-free policy and the other after the policy's implementation. The difference of these PAFs will give us the potential impact fraction (PIF) capturing the change in the outcomes when the TSE level changes following a novel smoke-free policy's implementation.

II. Sensitivity analyses

Three sets of sensitivity analyses will be conducted. First, excluding results with a higher risk of bias might lead to different estimations. Thus, we will conduct meta-regression analyses using the overall risk of bias based on the ROBINS-I assessment to test if effect sizes differ according to the overall risk of bias. Second, the question may arise whether a less restrictive selection strategy regarding study design would lead to different results. We will check for the robustness of our results by additionally including studies applying not only methodologies considered as the state-of-the-art in the Cochrane EPOC guidelines, but also study designs with a larger risk of bias (20). Third, we will test the sensitivity of our results to the selection of studies when overlapping populations are analysed in multiple studies.

III. Subgroup analyses

We will conduct subgroup analyses to evaluate whether smoke-free policies have differential effects among certain subgroups. To assess whether policies introduced in high-income countries have a different effect than policies introduced in low- or middle-income countries, we will perform meta-regression based on the World Bank classification (31). Second, previous studies have shown that smoke-free policies have a different effect according to individuals' SES (21, 32-34). We will extract information from subgroup analyses by SES indicators where possible and perform separate meta-analyses for high and low SES groups.

Meta-bias assessment

In each meta-analysis, we will assess the extent of publication bias via funnel plots when at least ten studies are available. Furthermore, we will look for unpublished studies for which a protocol was registered, and we will contact the authors of these studies to obtain their results. Subsequently, we will compare the findings of these unpublished studies with those of the published reports. Finally, we will look for published reports that followed a registered protocol. We will compare these protocols with the final papers to check whether selective reporting was present at any stage.

Confidence in cumulative estimate

We will discuss the strength of the evidence on the association between implementation of smoke-free policies covering (semi)private and outdoor places, and the TSE and child respiratory health outcomes based on the effect sizes, the variances of the estimated effects, the estimated level of heterogeneity between studies, sensitivity analyses, subgroup analyses, risk of bias, elements supporting causal inference and publication bias.

ETHICS AND DISSEMINATION

Ethical approval will not be required because we will only be using aggregate-level published data from previous studies. We expect to complete the study until September 2020 and before submitting for publication we will update our search and include any additional eligible papers that may be identified. We will publish our results in a peer-reviewed international journal and report the research findings according to the PRISMA guideline (35). Finally, we will disseminate our results to the public and policy-makers following the acceptance of our paper (e.g., actively seek media-attention and interaction with policy-makers and other relevant stakeholders).

REFERENCES

1. Öberg M, Jaakkola MS, Woodward A, Peruga A, Prüss-Ustün A. Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. The Lancet. 2011;377(9760):139-46.
2. Faber T, Kumar A, Mackenbach JP, Millett C, Basu S, Sheikh A, et al. Effect of tobacco control policies on perinatal and child health: a systematic review and meta-analysis. The Lancet Public Health. 2017;2(9):420-37.
3. Toebes B, Gispen ME, Been JV, Sheikh A. A missing voice: the human rights of children to a tobacco-free environment. Tobacco Control. 2018;27(1):3-5.
4. Jarvis MJ, Sims M, Gilmore A, Mindell J. Impact of smoke-free legislation on children's exposure to secondhand smoke: cotinine data from the Health Survey for England. Tobacco Control. 2012;18(1):749-61.
5. Akhtar PC, Currie DB, Currie CE, Haw SJ. Changes in child exposure to environmental tobacco smoke (CHETS) study after implementation of smoke-free legislation in Scotland: national cross sectional survey. BMJ. 2007;335(7619):545-9.
6. Holliday JC, Moore GF, Moore LA. Changes in child exposure to secondhand smoke after implementation of smoke-free legislation in Wales: a repeated cross-sectional study. BMC Public Health. 2009;9(1):430.
7. Sims M, Bauld L, Gilmore A. England's legislation on smoking in indoor public places and work-places: impact on the most exposed children. Addiction. 2012;107(11):2009-16.
8. Nanninga S, Lhachimi SK, Bolte G. Impact of public smoking bans on children's exposure to tobacco smoke at home: a systematic review and meta-analysis. BMC Public Health. 2018;18(1):749.
9. Hone T, Szklo AS, Filippidis FT, Laverty AA, Sattamini I, Been JV, et al. Smoke-free legislation and neonatal and infant mortality in Brazil: longitudinal quasi-experimental study. Tobacco Control. 2020(29):312-9.
10. Byron MJ, Cohen JE, Frattaroli S, Gittelsohn J, Drope JM, Jernigan DH. Implementing smoke-free policies in low-and middle-income countries: A brief review and research agenda. Tobacco induced diseases. 2019;17:60.
11. Painter K. Outdoor smoking: fair or foul? BMJ. 2019;366:l5754.
12. Laverty AA, Hone T, Vamos EP, Anyanwu PE, Taylor-Robinson D, de Vocht F, et al. Impact of banning smoking in cars with children on exposure to second-hand smoke: a natural experiment in England and Scotland. Thorax. 2020;75(4):345-347.

13. Faber T, Mizani MA, Sheikh A, Mackenbach JP, Reiss IK, Been JV. Investigating the effect of England's smoke-free private vehicle regulation on changes in tobacco smoke exposure and respiratory disease in children: a quasi-experimental study. The Lancet Public Health. 2019;4(12):607-17.

14. Been JV, Nurmatov UB, Cox B, Nawrot TS, Van Schayck CP, Sheikh A. Effect of smoke-free legislation on perinatal and child health: a systematic review and meta-analysis. The Lancet. 2014;383(9928):1549-60.

15. World Health Organization. Report on the global tobacco epidemic: offer help to quit tobacco use: executive summary. 2019.

16. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;4(1):1-9.

17. Higgins JP, Green S. Defining the review question and developing criteria for including studies. In: Cochrane handbook for systematic reviews of interventions. version 5. 2008.

18. Croghan IT, Ebbert JO, Hays JT, Schroeder DR, Chamberlain AM, Roger VL, et al. Impact of a countywide smoke-free workplace law on emergency department visits for respiratory diseases: a retrospective cohort study. BMC Pulmonary Medicine. 2015;15(1):6.

19. World Health Organization. WHO report on the global tobacco epidemic, 2017: monitoring tobacco use and prevention policies. Available from: https://www.who.int/tobacco/global_report/2017/en/

20. Cochrane Effective Practice and Organisation of Care (EPOC). EPOC Resources for review authors, 2017. Available from: https://epoc.cochrane.org/resources/epoc-resources-review-authors

21. Been JV, Millett C, Lee JT, van Schayck CP, Sheikh A. Smoke-free legislation and childhood hospitalisations for respiratory tract infections. European Respiratory Journal. 2015;46(3):697-706.

22. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. Journal of the Medical Library Association. 2016;104(3):240-3.

23. Effective Public Health Practice Project (EPHPP). Quality Assessment Tool for Quantitative Studies. 2010.

24. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

25. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health. 2012;66(12):1182-6.

26. Cummings P. The relative merits of risk ratios and odds ratios. Archives of Pediatrics & Adolescent Medicine. 2009;163(5):438-45.

27. Assink M, Wibbelink CJ. Fitting three-level meta-analytic models in R: A step-by-step tutorial. The Quantitative Methods for Psychology. 2016;12(3):154-74.

28. Cheung MW-L. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. Psychological Methods. 2014;19(2):211-29.

29. Vardavas CI, Hohmann C, Patelarou E. The independent role of prenatal and postnatal exposure to active and passive smoking on the development of early wheeze in children. Eur Respir J. 2016;48(1):115-24.

30. Jones LL, Hashim A, McKeever T, Cook DG, Britton J, Leonardi-Bee J. Parental and household smoking and the increased risk of bronchitis, bronchiolitis and other lower respiratory infections in infancy: systematic review and meta-analysis. Respiratory research. 2011;12(1):5.

31. World Bank. World Bank Open Data, 2019. Available from: https://data.worldbank.org/

32. Hajdu T, Hajdu G. Smoking ban and health at birth: Evidence from Hungary. Economics & Human Biology. 2018;30:37-47.

33. Vicedo-Cabrera AM, Schindler C, Radovanovic D, Grize L, Witassek F, Dratva J, et al. Benefits of smoking bans on preterm and early-term births: a natural experimental design in Switzerland. Tobacco Control. 2016;25(2):135-41.

34. Bakolis I, Kelly R, Fecht D, Best N, Millett C, Garwood K, et al. Protective effects of smoke-free legislation on birth outcomes in England-a regression discontinuity design. Epidemiology. 2016;27(6):810-8.

35. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Medicine. 2009;6:e1000100.

SUPPLEMENTAL MATERIAL

Appendix 1: Bias due to confounding and due to deviations from intended interventions

Level of bias	Bias due to confounding	Bias due to deviations from intended interventions
Low	(I) The pre-implementation temporal trends in the outcomes are controlled (II) Individual characteristics are controlled (III) For health outcomes: Contextual characteristics are controlled if reported to be an issue in the study	The percentage of the units affected by any of the co-interventions is below 10%
Moderate	(I) The pre-implementation temporal trends in the outcomes are controlled (II) Individual characteristics are not all controlled (III) For health outcomes: Contextual characteristics are controlled if reported to be an issue in the study	The percentage of the units affected by any of co-interventions is between 10% and 20%
Serious	(I) The pre-implementation temporal trends in the outcomes are not controlled (II) Some of the individual characteristics are controlled (III) For health outcomes: Contextual characteristics are controlled if reported to be an issue in the study	The percentage of the units affected by any of co-interventions is between 20% and 40%
Critical	(I) The pre-implementation temporal trends in the outcomes are not controlled (II) Some of individual characteristics are controlled OR (III) For health outcomes: Contextual characteristics are not controlled while being reported to be an issue in the study	The percentage of the units affected by any of co-interventions is above 40%
No information	No information on the controlled confounding variables	No information on the controlled co-interventions

Appendix 2: Bias assessment criteria (confounders and co-interventions)

ROBINS-I requires to pre-specify certain risk-of-bias criteria that are relevant for the given study. For assessing the risk of bias due to confounding, we consider the following potential confounders relevant: age, gender, socio-economic status (SES) and ethnicity, and the pre-implementation temporal trend in the given outcome variable. We will score results as having a low risk of bias due to confounding if it controls for all of these variables, moderate risk of bias if not all confounders are controlled for but the pre-implementation trend is, high risk of bias if the pre-implementation trend is not controlled for but some of the other confounders are, and critical risk of bias if none of the listed confounders are controlled for. Other potential contextual confounders specifically for the child health outcomes are air pollution, quality of the health care system, access to health care facilities, and the vaccination programme in the country. Therefore, additionally to the criteria noted above, studies will also be considered to have a critical risk of bias due to confounding if contextual confounders were unaccounted for despite these being reported to be an issue in the study.

For assessing bias due to co-interventions (i.e., part of 'bias due to deviations from intended interventions'), the following co-interventions are considered: policies aimed at reducing ambient air pollution, introduction of relevant health-care reforms and public health programmes, introduction of other tobacco control policies, change in vaccination programmes, and wars, epidemics, or economic crises which influenced the health of children. We will approximate the percentage of the observation units affected by any of these uncontrolled co-interventions. When this percentage is below 10% a result will be considered to have a low risk of bias, when it ranges between 10% and 20% moderate risk of bias, between 20% and 40% high risk of bias, and above 40% it will be considered serious risk of bias.

6

Impact of smoke-free policies in outdoor areas and (semi-) private places on children's tobacco smoke exposure and respiratory health: a systematic review and meta-analysis

Martá K. Radó*, Famke J.M. Mölenberg*, Lauren E.H. Westenberg, Aziz Sheikh, Christopher Millett, Alex Burdorf, Frank J. van Lenthe, Jasper V. Been

* Both authors contributed equally to this work

The Lancet Public Health 2021; In press

ABSTRACT

Introduction

Smoke-free policies in outdoor areas and (semi-)private places (e.g., cars) may lower health harms caused by tobacco smoke exposure (TSE). We aimed to review studies reporting the impact of such policies on children's TSE and respiratory health.

Methods

We conducted a systematic review and meta-analysis and searched 13 electronic databases until January 29, 2021. Eligible for the main analysis were (non)-randomised trials, interrupted time series and controlled before-after studies. Primary outcomes were: TSE in places covered by the policy, and hospital attendance for wheezing/asthma, and for respiratory tract infections (RTIs) in children aged <17 years. Risk-of-bias was assessed on a 4-point scale ranging from low to critical using ROBINS-I. Random-effects meta-analysis was conducted where appropriate. PROSPERO: CRD42020190563.

Results

Seven of the 11 identified studies fit pre-specified robustness criteria. These assessed smoke-free cars (n=5), schools (n=1), and a comprehensive policy covering multiple areas (n=1). Risk-of-bias was low to moderate in six studies and critical in one. In meta-analysis of ten effect estimates from four studies, smoke-free car policies were associated with an immediate TSE reduction in cars (risk ratio [RR] 0.69, 95% CI: 0.55 to 0.87; n=161,466). One study reported a gradual TSE decrease in cars. Individual studies found TSE reductions in school grounds following a smoke-free school policy and in hospital attendances for RTI following a comprehensive smoke-free policy.

Conclusion

Smoke-free car policies are associated with reductions in reported child TSE in cars, which could translate in health benefits. Very few studies assessed policies regulating smoking in outdoor areas and semiprivate places.

INTRODUCTION

Environmental tobacco smoke exposure (TSE) resulting from second-hand smoke (i.e., inhalation of emitted smoke) and third-hand smoke (i.e., the uptake of tobacco smoke residuals from polluted surfaces) is known to cause a major burden on children's health. Each year, second-hand smoke is responsible for an estimated 56,000 deaths in children under ten years of age globally (1) and 35,633 disability-adjusted life years (DALYs) lost among children in the European Union (2). TSE has been linked to a range of adverse respiratory health outcomes in children, including respiratory tract infections (RTIs), wheezing, and asthma (3, 4). Governmental action to protect children from these deleterious effects of TSE is urgently needed, as children are not able to control their level of exposure. Smoke-free policies have been identified as a key instrument to achieve the United Nations Sustainable Development Goal 3, to improve child health and well-being (5, 6).

Solid evidence indicates that smoke-free policies covering enclosed public places can effectively reduce adverse respiratory health outcomes in children, including decreasing hospital attendance for asthma exacerbations by -10% [95% confidence interval (CI): -17 to -3] and hospital attendance for lower RTIs by -18% (95% CI:-33 to -4) (4). These health benefits are likely mediated via the reduction in TSE from second- and third-hand smoke. Smoke-free policies covering enclosed public areas can decrease child TSE in places covered by the policy, but also – via norm spreading – in places not covered including private places, such as cars or homes (7-11).

In recent years, an increasing number of jurisdictions have expanded smoke-free policies to encompass outdoor areas (e.g., school grounds, playgrounds and parks) (12), semi-private places (e.g., shared housing), and private places (e.g., cars) (13). Many of these places are frequented by children and therefore contribute to TSE during childhood. Estimates of the effectiveness of these "novel smoke-free policies" cannot be easily extrapolated from earlier evidence on smoke-free indoor areas, for example, given the dilution of TSE in outdoor places, and enforcement issues regarding policies covering private areas (13). Besides, it is unclear whether such policies affect TSE exposure in other places either negatively (i.e., via displacement of smoking) or positively (i.e., via norm spreading) (13).

In order to inform policy, we sought to systematically review evidence on the impact of smoke-free policies covering outdoor areas or (semi-)private places on TSE and respiratory health in children.

METHODS

Protocol and registration

This systematic review and meta-analysis was conducted according to a published peer-reviewed protocol with PROSPERO registration (CRD42020190563) (14). We used PRISMA guidelines to report our findings (15).

Eligibility criteria

Studies were eligible for inclusion if they evaluated the association between implementation of policies restricting smoking in designated (semi-)private places or any outdoor areas, and TSE or respiratory health outcomes in children. In line with earlier systematic reviews (4, 16), we included studies in which at least 50% of the study population was aged <17 years to ensure that we would not exclude studies that included a high proportion of children. Studies reporting the effect of a smoke-free policy covering indoor private places (e.g., cars), indoor semi-private places (e.g., multi-unit housing), outdoor (semi-)private places (e.g., shared gardens), and outdoor public places (e.g., parks, school grounds, beaches, hospital grounds) introduced at any governmental or institutional level were considered eligible. Primary outcomes included: 1) TSE in places covered by the policy; 2) unplanned hospital attendance for wheezing/asthma; and 3) unplanned hospital attendance for RTIs. Following the methodological recommendations of the Cochrane Effective Practice and Organization of Care (EPOC) group (17), we selected studies with the most robust study designs for our main analyses, namely: (non-)randomised trials, interrupted time series, and controlled before-after studies.

Information sources and study selection

To identify eligible studies we searched 13 database from date of inception until 29 January 2021. No restrictions were imposed for the observational period, publication date, or language. To identify additional relevant studies including grey literature, we hand-searched reference lists and citations of included studies and consulted with experts in the field. Two out of the three reviewers (MKR, FJMM, and LEHW) independently screened each record's title and abstract, and thereafter the full-texts, to identify eligible studies. Disagreement was resolved through discussion or by involving an adjudicator (JVB).

Data collection and risk of bias assessment

Two of the three reviewers (MKR, FJMM, and LEHW) independently extracted data with a pre-specified extraction form and assessed risk of bias for effect estimates from each study using the Risk of Bias in Non-Randomised Studies of Interventions (ROBINS-I) tool (18). Again, disagreements were resolved through discussion or by involving an adjudicator (JVB). If relevant data were missing, we contacted the corresponding author. We extracted from the included studies any additional information that could provide further insight on the robustness of the findings (i.e., use of an alternative comparison group, neutral outcomes, or alternative method) and on the mechanism of how policies might have impacted the outcomes of interest, following the United Kingdom (UK) Medical Research Council guidance on natural experiments (19).

Summary measures

Point estimates and 95% CI are reported in tabular form. For dichotomous outcomes, we harmonised effect estimates into relative risk (RR). When a study reported odds ratios (ORs) instead, we contacted the corresponding author to request the RR. In case this could not be provided, we applied the following formula to approximate RR based on OR as described in Equation 1:

$$RR = \frac{OR}{(1 - EER) + (EER \times OR)} \tag{1}$$

where EER is the expected event rate or prevalence in the control group (20). If EER was not available in interrupted time series studies, the overall event rate of the study population was used instead. Regarding outcomes that could occur multiple times with the same individual (e.g., hospital attendance), we analysed incidence rate ratios (IRRs).

Methods of analysis

We performed random-effects meta-analyses to derive pooled effect estimates when at least two studies evaluated policies that regulated smoking in similar places and reported on the same outcome. Based on a one-sided log-likelihood-ratio test, we assessed whether a three-level meta-analysis instead of a two-level meta-analysis would be needed to account for dependency of observations for estimates of similar policies implemented in different regions within a country. Step changes (immediate changes) and slope changes (gradual changes) were pooled in separate models.

We conducted sensitivity analysis including studies with a less robust design (14, 17). A priori, we planned a number of other sensitivity and subgroup analyses that we could not conduct due to the low number of eligible studies (14). We presented findings on effect modification by socioeconomic status when it was reported. In addition to quantitative study findings, we narratively described additional elements from individual studies that may support causal inference.

As we anticipated that most studies would have evaluated TSE rather than health outcomes, we planned an additional health impact assessment to quantify the potential impact of any observed changes in TSE following the introduction of smoke-free policies. We calculated the potential impact fraction (PIF) which captures the change in health outcomes attributable to the change in TSE following the policy implementation (21), as described in Equation 2:

$$PIF = \frac{(P_0 - P_1) \times (RR - 1)}{P_0 \times (RR - 1) + 1} \tag{2}$$

where P_0 is the prevalence of TSE before policy implementation, P_1 is the prevalence of TSE after policy implementation, and RR the relative risk of respiratory disease of children exposed to TSE over unexposed children. To capture the sensitivity of PIFs to varying parameters, we calculated PIFs given a plausible range of TSE baseline levels, and associations between exposure and outcome. Based on studies identified in the review, this analysis was only possible for smoke-free car policies. In these studies, TSE before implementation ranged from 6% in the UK (22) to 43% in Canadian provinces (23). Therefore, we modelled scenarios varying the baseline TSE levels between 5 and 45%. To account for location-specific strength of associations, we used the associations between TSE in cars and asthma diagnosis presented by one included study (i.e., RR 1.12 (95% CI: 0.98 to 1.28) for children with 1 to 2 days per week of TSE in cars as compared to children with no TSE in cars, and RR 1.19 (95% CI: 1.02 to 1.38) for children with 3 to 7 days per week of TSE in cars (24)).

Role of the funding source

The funders of this study had no role in study design, data collection, analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

We identified 5,745 records and after deduplication, 2,831 records were screened on title and abstract. From 204 full-text articles assessed, 11 studies were identified (Figure 1). No ongoing or unpublished studies were found.

Seven studies meeting EPOC criteria were included: four controlled before-after studies (22, 23, 25, 26), and three interrupted time series studies (13, 24, 27) (Supplemental Table 1). Overall, risk of bias was low in three studies (13, 23, 24), moderate in three (22, 25, 26), and critical in one study (26) (Supplemental Table 2). Table 1 shows the primary and secondary outcomes that were assessed in the identified studies. Information on the evaluated policies and enforcement is presented in Table 2. Five studies evaluated smoke-free car policies, of which two focused on the national policy in England (UK) (13, 22), two on various policies across Canadian provinces (23, 25), and one in California (US) (24). One study assessed a smoke-free school policy in Canada (26), and another evaluated a comprehensive smoke-free policy covering enclosed public, (semi-)private, and outside areas in Hong Kong (27). Four of the 11 eligible studies did not meet EPOC criteria, and are discussed separately (28-31).

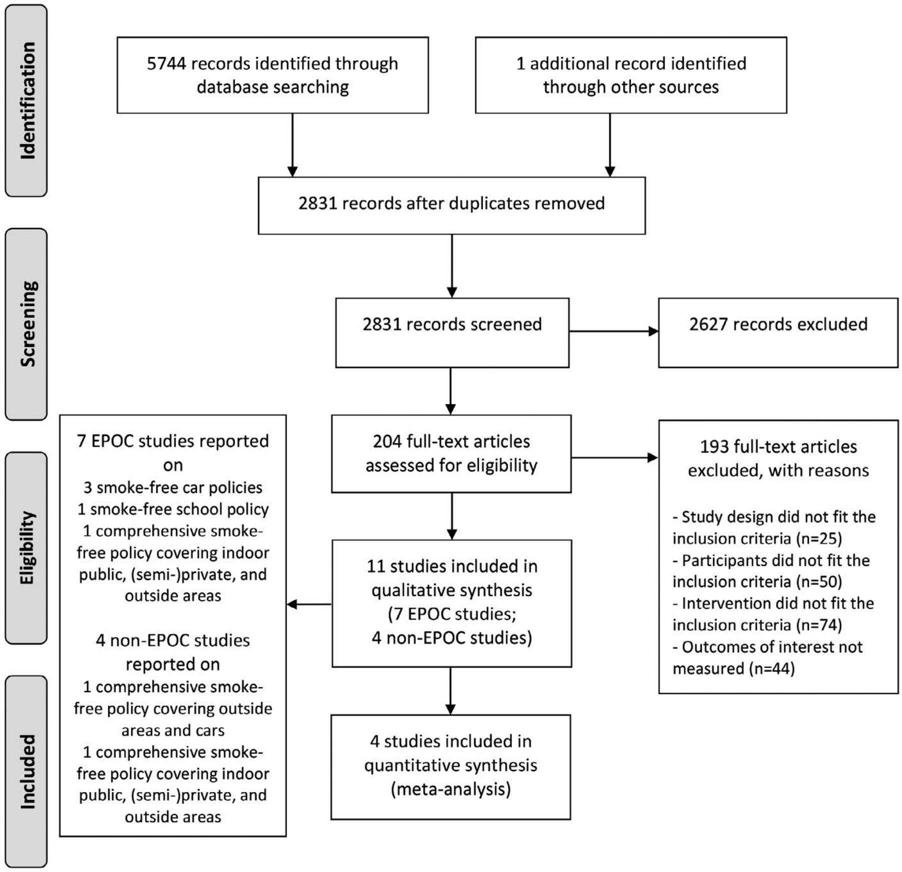


Figure 1: PRISMA flow diagram

Abbreviation: EPOC = Effective Practice and Organization of Care

In two-level meta-analysis, smoke-free car policies were associated with an immediate risk reduction in child TSE in cars (RR 0.69, 95% CI: 0.55 to 0.87; ten effect estimates from four studies, 161,466 participants) (Figure 2). A one-sided log-likelihood-ratio test favoured two-level over three-level meta-analysis (P-value: 0.38) (Supplemental Figure 1). One additional study from California (US) found that the smoke-free car policy was followed by an annual reduction in child TSE in cars (RR 0.95/year, 95% CI: 0.94 to 0.97; 151,074 participants), with no significant temporal trend in TSE in the pre-intervention period (24).

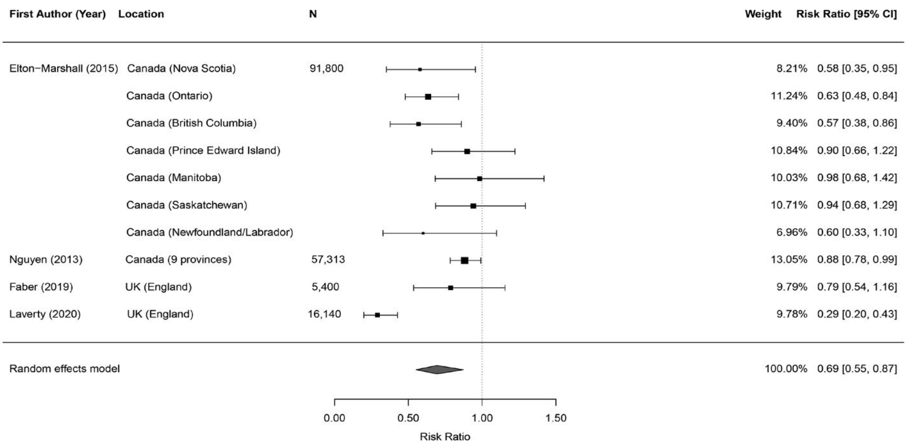


Figure 2: Meta-analysis of relative risk of child tobacco smoke exposure in cars before and after implementation of smoke-free car policy

Black squares indicate the point estimate for each policy, with the horizontal lines representing the 95% confidence intervals. The pooled effect of the ten estimates was obtained using a random-effects meta-analysis, with the effect of each event weighted by the inverse of its variance. The pooled effect is indicated with the diamond, and the width of the diamond corresponds with the 95% confidence intervals of the pooled effect. N refers to the number of participants in the given study.

Two studies reported on secondary TSE outcomes (Supplemental Table 3) (13, 23). One study from England found a relative increase in the proportion of children having detectable salivary cotinine levels (RR 1.22, 95% CI: 1.06 to 1.38; 7,858 participants) (13). Although TSE at home or other people's homes appeared to increase following the policy in this study, this was statistically insignificant. One Canadian study found that the smoke-free car policy was not associated with significant changes in TSE at places other than cars, including bus stops and shelters, parks and sidewalks, and inside restaurants (23).

Regarding health outcomes, one study from England found no significant change in the incidence of childhood wheezing or asthma following the smoke-free car policy (RR 0.82, 95% CI: 0.63 to 1.05; 13,369 participants) (Supplemental Table 3) (13).

Among studies assessing other policies, a controlled before-after study from Canada found a reduction in TSE among high school students on school property following a smoke-free school policy (RR 0.89, 95% CI: 0.83 to 0.95; 20,388 participants) (26). No health outcomes were assessed.

Table 1: Primary and secondary outcomes that were reported in included studies on smoke-free policies

First author (year of publication)	Meeting EPOC criteria	TSE outcome				Health outcome		
		Primary: TSE in places covered by the policy	Secondary: TSE in unspecified places	Secondary: TSE in places not covered by the policy	Secondary: biomarker of TSE	Primary: unplanned hospital attendance for RTIs	Secondary: incidence of wheezing/ asthma	Secondary: chronic cough
Smoke-free car policy								
Elton-Marshall (2015)	Yes	X						
Faber (2019)	Yes	X		X	X		X	
Laverty (2020)	Yes	X						
Nguyen (2013)	Yes	X		X				
Patel (2018)	Yes	X						
Comprehensive smoke-free policy covering outside areas and cars								
Gagné (2020)	No	X		X				
Smoke-free school ground policy								
Azagba (2016)	Yes	X						
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas								
Lee (2016)	Yes					X		
Chan (2011)	No		X	X			X	
Chan (2014)	No		X	X	X			
Ho (2010)	No	X	X	X				X

Abbreviations: tobacco smoke exposure (TSE); respiratory tract infection (RTI); Cochrane Effective Practice and Organization of Care (EPOC). Note: 'X' denotes that the certain study included the given outcome.

Table 2: Description of the novel smoke-free policies evaluated in eligible studies[a]

First author (year)	Country (region)	Description	Level of enforcement	Timing of implementation	Enforcement	Actual enforcement
Smoke-free car policy						
Nguyen (2013); Elton-Marshall (2015)	Canada (nine provinces)	Smoke-free car policy prohibiting smoking in private vehicles with anyone aged ≤15 years present (≤18 years for Nova Scotia and Prince Edward Island)	Province	Nova Scotia April 1, 2008; Ontario January 21, 2009; British Columbia April 7, 2009; Prince Edward Island September 1, 2009; New Brunswick January 1, 2010; Manitoba July 15, 2010; Saskatchewan October 1, 2010; Newfoundland May 31, 2011; Alberta January 1, 2013.	Law enforcement agencies were authorised to issue fines or warnings to those who do not comply with the ban. Fines vary across provinces (maximum fines stated in the provincial Tobacco Control Act for any offences, but no specific guideline for violating smoke-free car policy).	A few fine tickets were issued in the initial periods which mainly relied on the deterrence effect and educating the value of the policy.
Faber (2019); Laverty (2020)	United Kingdom (England)	Smoke-free car policy prohibiting smoking in private vehicles with anyone aged ≤18 years present. Exceptions apply for convertible cars with the roof completely down and for e-cigarettes	Country	October 1, 2015	The driver and smokers who break this policy risk a £50 (i.e., US$60) fine each. Before the policy came into force, police had announced that they were not planning to actively enforce the policy.	One year after the policy was imposed, only one single penalty was issued in England. Other cases were dealt with by verbal warnings.
Patel (2018)	United States of America (California)	Smoke-free car policy prohibiting smoking in a motor vehicle with anyone aged ≤17 years present	State	January 1, 2008	Police was not authorised to stop a vehicle for a smoking violation alone; it must have been secondary to another infraction. Violators of the policy can be fined up to US$100.	Not reported

Table 2: (Continued) Description of the novel smoke-free policies evaluated in eligible studies[a]

First author (year)	Country (region)	Description	Level of enforcement	Timing of implementation	Enforcement	Actual enforcement
Comprehensive smoke-free policy covering outside areas and cars						
Gagné (2020)	Canada (Quebec)	A comprehensive smoke-free policy covering bar and restaurant patios, playgrounds, within 9 metres from building entrances and in vehicles with youth under the age of 16. Furthermore, it permitted landlords to enforce a smoke-free policy in multi-unit apartment buildings	Province	November 2015	Police services may stop a motor vehicle if the member has reasonable grounds to believe that a person is smoking in the vehicle while a minor under 16 years of age is present in it. Furthermore, smoking in a prohibited place is fined (US$40 to US$600), and repeated offences are higher (US$80 to $1200).	Not reported
Smoke-free school policy						
Azagba (2016)	Canada (four provinces)	Smoke-free school policy (not further specified)	Province	Quebec May 2006; British Colombia March 2008; Prince Edward Island September 2009; Saskatchewan August 2010.	Law enforcement agencies were authorised to issue fines or warnings for any smoke-free policy offences. Fines vary across provinces (maximum fines stated in the provincial Tobacco Control Act for any smoke-free policy offences, but no specific guideline for violating smoke-free school policy).	Not reported

Table 2: (Continued) Description of the novel smoke-free policies evaluated in eligible studies[a]

First author (year)	Country (region)	Description	Level of enforcement	Timing of implementation	Enforcement	Actual enforcement
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas						
Ho (2010); Chan (2011); Chan (2014); Lee (2016)	Hong Kong	A comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas (i.e., public playgrounds, parks, beaches, barbecue sites, public swimming pools and areas of public housing estates)	City	January 1, 2007	Policy was enforced by the Tobacco Control Office. The budget for policy enforcement increased from the US $0.9 million in 2006 (pre-legislation) to US$3 million in 2007 (post legislation). Penalty points are allotted to households for smoking and other offences, with the ultimate punishment being the termination of tenancy.	The policy was effectively enforced. In two years, 11085 penalties were issued against smoking offences (outside or inside areas).

[a] A list of additional sources used to extract information on the policies is available in Appendix 1

Table 3: Findings from studies meeting EPOC criteria reporting the association between implementation of novel smoke-free policies and primary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Outcome	Reported intervention effect	Summary of findings
Smoke-free car policy								
Elton-Marshall (2015)	Canada (Nova Scotia)	Children aged 11-14 years	91800 (without missing values 83331)	2004-2012	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days (any vs none)	OR: 0.47 (0.25 to 0.89)	Smoke-free car policy in Nova Scotia was associated with a reduction in TSE in cars
	Canada (Ontario)				Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days (any vs none)	OR: 0.58 (0.42 to 0.80)	Smoke-free car policy in Ontario was associated with a reduction in TSE in cars
	Canada (British Columbia)				Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days (any vs none)	OR: 0.51 (0.32 to 0.82)	Smoke-free car policy in British Colombia was associated with a reduction in TSE in cars
	Canada (Prince Edward Island)				Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days (any vs none)	OR: 0.87 (0.59 to 1.30)	Smoke-free car policy in Prince Edward Island was not associated with significant changes in TSE in cars
	Canada (Manitoba)				Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days (any vs none)	OR: 0.98 (0.62 to 1.54)	Smoke-free car policy in Manitoba was not associated with significant changes in TSE in cars
	Canada (Saskatchewan)				Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days (any vs none)	OR: 0.91 (0.56 to 1.48)	Smoke-free car policy in Saskatchewan was not associated with significant changes in TSE in cars

Table 3: (Continued) Findings from studies meeting EPOC criteria reporting the association between implementation of novel smoke-free policies and primary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Outcome	Reported intervention effect	Summary of findings
	Canada (Newfoundland/ Labrador)				Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days (any vs none)	OR: 0.53 (0.26 to 1.09)	Smoke-free car policy in Newfoundland/Labrador was not associated with significant changes in TSE in cars
Faber (2019)	United Kingdom (England)	Children aged 8-15 years	5400 (without missing values 5399)	2008-2017	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past days (any vs none)	OR: 0.77 (0.51 to 1.17)	Smoke-free car policy in England was not associated with significant changes in TSE in cars
Laverty (2020)	United Kingdom (England)	Children aged 13-15 years	16140 (missing values unknown)	2012-2016	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days (regular exposure vs non/ occasional)	OR: 0.28 (0.21 to 0.37) Absolute difference (percentage points): -4 (-3 to -2)	Smoke-free car policy in England was associated with a reduction in TSE in cars
Nguyen (2013)	Canada (Prince Edward Island, Nova Scotia, New Brunswick, Ontario, Manitoba, Saskatchewan, British Columbia)	Children aged 15-16 years, except for Nova Scotia and Prince Edward Island where children were aged 15-19 years	57313 (without missing values 56596)	2005-2010	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past month (any vs none)	Absolute difference (percentage points): -5.1 (-9.8 to -1.0)	Smoke-free car policy in seven Canadian provinces was associated with an overall reduction in TSE in cars

Table 3: (Continued) Findings from studies meeting EPOC criteria reporting the association between implementation of novel smoke-free policies and primary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Outcome	Reported intervention effect	Summary of findings
Patel (2018)	United States of America (California)	Children aged 11-18 years	151074 (missing values unknown)	2001-2012	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past month (any vs none)	Pre-intervention annual trend (percentage points): -0.3/year (-0.6 to 0.7) Post-intervention annual trend, (percentage points): -1.2/year (-1.5 to -0.8)	The smoke-free car policy in California was followed by an annual decline in TSE in cars, whereas there was no significant temporal trend in the pre-intervention period. Step or slope changes were not formally tested.
Smoke-free school policy								
Azagba (2016)	Canada (Quebec, British Columbia, Prince Edward Island, Saskatchewan)	Children aged 15-18 years	20388 (missing values unknown)	2005-2012	Primary: TSE in places covered by the policy	Child-reported TSE on a school property in the past month (any vs none)	Absolute difference (percentage points): -6.5 (-10.0 to -3.0)	Smoke-free school policy in four Canadian provinces was associated with an overall reduction in TSE on a school property

Table 3: (Continued) Findings from studies meeting EPOC criteria reporting the association between implementation of novel smoke-free policies and primary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Outcome	Reported intervention effect	Summary of findings
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas								
Lee (2016)	Hong Kong	Children aged 0-18 years	75870 (missing values unknown)	2004-2012	Primary: unplanned hospital attendances for RTIs	Unplanned hospital admissions for lower RTI (yes vs no)	Immediate change: OR: 0.66 (0.63 to 0.69) Gradual change: OR: 0.86/year (0.84 to 0.88)	Comprehensive smoke-free policy in Hong Kong was associated with an immediate reduction in hospital admissions for lower RTI, followed by an additional reduction per year

Abbreviations: odds ratio (OR), 95% confidence interval between brackets; tobacco smoke exposure (TSE); respiratory tract infection (RTI).

6

An interrupted time series study from Hong Kong found that a comprehensive smoke-free policy covering (semi-)private, and outside areas in addition to enclosed public spaces was associated with an immediate drop in unplanned hospital attendances for RTIs among children (RR 0.66, 95% CI: 0.63 to 0.69; n=75,870 hospital attendance) and an additional annual decrease (RR 0.86/year, 95% CI: 0.84 to 0.88) (27).

Four studies did not meet EPOC criteria (Supplemental Table 4), and all had critical risk of bias (Supplemental Table 5). One uncontrolled before-after study evaluated a comprehensive smoke-free policy covering cars and outside areas in Quebec (Canada) (31). The results indicated an immediate reduction in child TSE in cars (RR 0.42, 95% CI: 0.30 to 0.57), and a decline in child TSE at home (RR 0.55, 95% CI: 0.41 to 0.73) (Supplemental Table 6). Including this estimate in the meta-analysis did not materially change the overall effect estimate of smoke-free car policies on child TSE (RR 0.66, 95% CI: 0.53 to 0.83) (Supplemental Figure 2).

We identified three uncontrolled before-after studies that evaluated a comprehensive smoke-free city-wide policy in Hong Kong (Supplemental Table 4) (28-30). Meta-analysis of these studies was not possible since study populations overlapped, or outcomes could not be harmonised. One study found that the policy was followed by a significant increase in TSE in places covered by the policy and in TSE overall (Supplemental Table 6) (28). Two other studies using parental-reported outcomes did not assess TSE in places covered by the policy, but found a significant decrease in child TSE at home (29, 30).

Using effect estimates from the meta-analysis, we estimated the proportion of respiratory disease that could potentially be prevented by the observed TSE reductions following smoke-free car policies. Assuming a baseline level of TSE in cars of 20% and using the association between TSE for 1-2 days a week and asthma (RR=1.12), we estimated that 2.3% (95% CI: -0.4 to 5.3) of asthma diagnoses are attributable to TSE in cars. Based on the effect estimate of our meta-analysis, the PIF indicated that the proportion of asthma diagnoses would decrease by -0.7 percent (95% CI: -1.1 to -0.3) by implementing a smoke-free car policy (Figure 3). In Supplemental Table 7, we show different scenarios indicating that the proportion of asthma cases among children that could potentially be prevented by smoke-free car policies ranged between -0.2% and -2.4%. PIFs were not calculated for policies covering outdoor areas, since meta-analysis on the effect of these policies on TSE was not possible.

One study assessed socioeconomic inequalities in TSE following the implementation of a comprehensive smoke-free policy including outside areas and cars in Quebec (Canada) (31). Their findings suggested that child TSE in cars and at home decreased in each education and income group, but that the relative inequalities remained unchanged.

Some studies provided further information supporting robustness of the findings (Supplemental Table 8 and Table 9) (32). Three studies found that their results were robust to different specifications of comparison groups (13, 22, 25). A Canadian study did not correct for a pre-legislation trend in their main analysis, but additionally showed that TSE in cars did not decline before the policy was introduced (23). The US study on smoke-free cars showed that the change in TSE in cars in California could not be explained by secular trends at the national level (24). Some studies reported further information on the underlying mechanism that may explain the change in outcomes. A study from England found that the implementation of the smoke-free car policy did not significantly change active smoking or TSE in cars among adults, possibly explaining the null findings for TSE in cars among children in that study (13). A study evaluating the smoke-free car policy in Canada did not find a significant change in smoking at home among smokers, suggesting that no significant displacement of TSE or norm spreading towards other private areas occurred (23). Furthermore, the policy impact was restricted to children whose parents had a car, supporting causality of the findings (23).

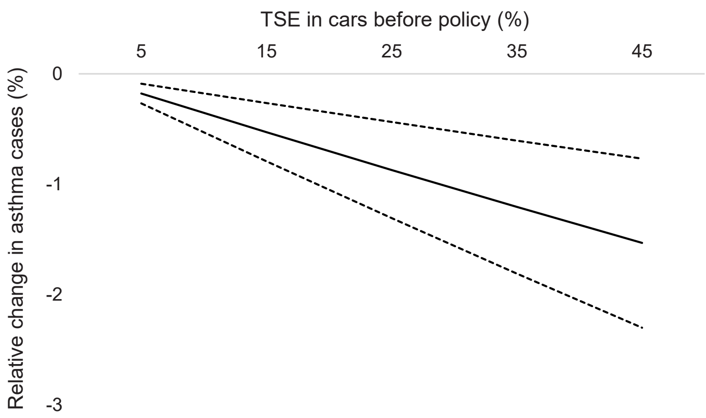


Figure 3: Estimated proportion of asthma cases in children that could be prevented by introducing smoke-free car policies for varying baseline levels of tobacco smoke exposure in cars

The relative change in asthma cases was estimated by calculating the potential impact fraction, which captures the change in asthma cases attributable to the change in tobacco smoke exposure following the implementation of smoke-free car policies. The solid line represents the average effect, dotted lines represent 95% confidence interval
Abbreviation: tobacco smoke exposure (TSE).

DISCUSSION

This systematic review and meta-analysis found that smoke-free car policies were associated with substantial reductions in TSE among children in cars. We estimate that such changes may translate into an estimated -0.2 to -2.4% decrease in asthma diagnoses. Additionally, a very limited number of studies indicated that smoke-free policies covering school grounds and a comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas may reduce TSE and improve health outcomes in children. Although based on a small number of studies, the evidence identified suggests that extending smoke-free policies to private and outdoor settings may help protect children from TSE-related harms and may provide additional health benefits.

To our knowledge, this is the first systematic assessment of the impact of smoke-free policies covering outdoor and (semi-)private places on children's TSE and respiratory health. Whereas the link between smoke-free policies in enclosed public places and child health benefits is already well established (4, 16), the effect of smoke-free policies in other locations was unclear. To ensure that all relevant studies were identified, we used a comprehensive search strategy including screening 13 electronic databases, checking references and citations, and consulting experts to identify additional studies. Moreover, we followed EPOC guidelines for including studies using methodologically robust designs in the main analysis (17). We also extracted supportive information from the included studies to facilitate causal reasoning.

Evidence on the effectiveness of smoke-free policies, similar to most large-scale public health interventions, was derived from quasi-experimental studies. Although such methodologies have a risk of bias (32), this was assessed to be low to moderate in six out of seven studies in our primary analysis, strengthening confidence in our results. Due to the small number of eligible studies, we could not conduct our pre-planned meta-regression and subgroup analyses, nor assess potential publication bias. Our findings need to be supported with future additional studies, and at present need to be interpreted with caution. Further, our findings must be interpreted in the light of the observational nature of the available evidence, the fact that findings from systematic reviews and meta-analyses are inherently dependent on the quality of the underlying studies (which is why we limited our primary analyses to EPOC-approved designs), and on heterogeneity of the estimated effects (33, 34). All TSE outcomes were child-reported or parent-reported. Although these measures might be subject to desirability bias, previous studies support their validity in quantifying actual exposure (35, 36). Biomarkers for exposure were evaluated in some studies (13, 30), but these cannot discern between TSE in various locations. Further, TSE presents different risks in different settings. For example, TSE in cars will likely be more harmful to child health than TSE in outside areas (37, 38). A formula suggested by the Cochrane Handbook was used to compute RR when only OR was available, however, the conversion may be biased in situations with a high level of confounding (39).

Our review builds on solid existing evidence indicating child health benefits of smoke-free policies (4). Based on the current meta-analysis, we estimated that smoke-free car policies may contribute to a moderate reduction in the number of asthma diagnoses in children, ranging from -0.2% in the most conservative scenario with low baseline levels of TSE in cars, to -2.4% in the most favourable scenario. It is important to note that these calculations assumed that there was no change in TSE in other places than cars, which needs further substantiation in future research as the current evidence-base on this is rather limited (13). Despite the relatively modest reductions, more widespread implementation of smoke-free car policies might translate to important health benefits given the substantial global burden of asthma (40, 41).

All evidence in this review was derived from countries with an already existing and well-enforced comprehensive smoke-free legislation covering enclosed public places. Thus, countries may derive substantial additional benefits by implementing an even more comprehensive measure covering indoor enclosed places (4), as well as private and outdoor areas. Moreover, the comprehensive smoke-free policy in Hong Kong covering indoor public, (semi-)private, and outside areas was associated with large reductions in hospital attendances for lower RTIs in children. Although it is not possible to disentangle the relative contributions of the various spaces covered by this smoke-free policy, the effect sizes were much larger than those from other studies which assessed the impact of policies covering enclosed public places only on child RTIs (42-44), suggesting that part of the impact may be from its additional coverage of (semi-)private and outdoor areas.

Previous studies indicated that part of the positive health impact of smoke-free policies covering indoor public places may be mediated via reducing TSE in cars and homes through norm spreading (11). At present, there is limited insight on whether norm spreading also occurs following smoke-free policies covering outdoor or private places, or whether displacement of smoking to areas not covered by the policy may occur (45). In our review, there was mixed evidence on the impact of smoke-free policies covering private or outdoor areas on TSE in areas not covered by the policy. One study found that salivary cotinine levels increased in children following the smoke-free car policy in England, indicating that overall TSE may have increased (13). In contrast, there was no evidence of displacement of smoking to outside areas or restaurants using surveys following the Canadian smoke-free car policy

(23), Another Canadian study showed that TSE at home was reduced following a comprehensive policy including outdoor areas and cars (31). Two studies found that the comprehensive city-wide smoke-free policy in Hong Kong was associated with a reduction in TSE in areas which were not covered by the policy (29, 30), however, one study found the opposite (28). This inconsistency could derive from the fact that the latter study was based on child reports, whereas the other two studies relied on parent-reported outcomes which could be subject to higher desirability bias. Further, substantial displacement of TSE is unlikely in Hong Kong given the observed reduction in hospital attendances for RTIs (27). We are unaware of any evidence from other studies on whether smoke-free policies in (semi-)private or outdoor areas had an impact on smoking behaviour in places not covered by the policy.

Compliance is essential for smoke-free policies to be effective, and enforcement of policies covering private or outdoor places can be challenging (46-49). Despite these challenges which were also noted in the included studies, health benefits could be demonstrated, and these may increase with more widespread adoption and acceptability. Several measures could be taken to improve policy compliance such as penalties (46, 49), smoke-free signage's (50, 51), or information campaigns (47, 52), which can carry additional child health benefits (53). In general, the reviewed policies were positively perceived by the general population which can foster effective policy implementation (54). Smoke-free policies introduced in public enclosed areas often gain support after implementation as they become customary (55-58).

Our review provides the first meta-analysis assessing the effectiveness of smoke-free car policies on TSE among children. More studies are needed to further substantiate findings for smoke-free policies in outdoor areas and should cover a wider range of areas. No eligible studies assessing policies for private homes, outdoor hospital grounds, or parks were identified even though these policies could potentially have a great impact on children's health. Previous studies found that such policies can decrease TSE among adults and cross-sectional studies also supported the potential child health benefits of these policies (38, 59-63). We did not identify any study exploring the differential effects of novel smoke-free policies in low-income countries. This is worthy of future investigation as previous studies have observed that population-level tobacco control policies might produce greater health benefits in low-income populations (13, 64).

CONCLUSION

Although the health burden associated with TSE has declined over past decades around the world, there is still considerable scope to further reduce this preventable harm to children (65). To realise this goal, comprehensive smoke-free policies are needed and our review, albeit based on a small number of studies, suggests that including private and outdoor places in national tobacco control policies may produce additional benefits. The majority of studies identified evaluated smoke-free car policies, and when taken together, these suggest that such policies can help reduce children's TSE in cars. Based on informed estimations, we demonstrate that this may translating into small improvements in respiratory health. We found limited evidence indicating that policies covering other private or outdoor areas may also reduce TSE and offer additional respiratory health benefits. All children around the world should have the right to breathe clean air in private, public, indoor and outdoor areas.

REFERENCES

1. Alam K, Hankey GJ, Collaborators GRF. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1345-422.
2. Carreras G, Lachi A, Cortini B, Gallus S, López MJ, López-Nicolás Á, et al. Burden of disease from exposure to secondhand smoke in children in Europe. Pediatric Research. 2020:1-7.
3. Öberg M, Jaakkola MS, Woodward A, Peruga A, Prüss-Ustün A. Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. The Lancet. 2011;377(9760):139-46.
4. Faber T, Kumar A, Mackenbach JP, Millett C, Basu S, Sheikh A, et al. Effect of tobacco control policies on perinatal and child health: a systematic review and meta-analysis. The Lancet Public Health. 2017;2(9):420-37.
5. Sachs JD. From millennium development goals to sustainable development goals. The Lancet. 2012;379(9832):2206-11.
6. Lee BX, Kjaerulf F, Turner S, Cohen L, Donnelly PD, Muggah R, et al. Transforming our world: implementing the 2030 agenda through sustainable development goal indicators. Journal of public health policy. 2016;37(1):13-31.
7. Jarvis MJ, Sims M, Gilmore A, Mindell J. Impact of smoke-free legislation on children's exposure to secondhand smoke: cotinine data from the Health Survey for England. Tobacco Control. 2012;18(1):749-61.
8. Akhtar PC, Currie DB, Currie CE, Haw SJ. Changes in child exposure to environmental tobacco smoke (CHETS) study after implementation of smoke-free legislation in Scotland: national cross sectional survey. BMJ. 2007;335(7619):545-9.
9. Holliday JC, Moore GF, Moore LA. Changes in child exposure to secondhand smoke after implementation of smoke-free legislation in Wales: a repeated cross-sectional study. BMC Public Health. 2009;9(1):430.
10. Sims M, Bauld L, Gilmore A. England's legislation on smoking in indoor public places and work-places: impact on the most exposed children. Addiction. 2012;107(11):2009-16.
11. Nanninga S, Lhachimi SK, Bolte G. Impact of public smoking bans on children's exposure to tobacco smoke at home: a systematic review and meta-analysis. BMC Public Health. 2018;18(1):749.
12. Painter K. Outdoor smoking: fair or foul? BMJ. 2019;366:l5754.

13. Faber T, Mizani MA, Sheikh A, Mackenbach JP, Reiss IK, Been JV. Investigating the effect of England's smoke-free private vehicle regulation on changes in tobacco smoke exposure and respiratory disease in children: a quasi-experimental study. The Lancet Public Health. 2019;4(12):607-17.

14. Radó MK, Mölenberg FJ, Sheikh A, Millett C, Bramer WM, Burdorf A, et al. Impact of expanding smoke-free policies beyond enclosed public places and workplaces on children's tobacco smoke exposure and respiratory health: protocol for a systematic review and meta-analysis. BMJ open. 2020;10(10):e038234.

15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Medicine. 2009;6(7) e1000100.

16. Been JV, Nurmatov UB, Cox B, Nawrot TS, Van Schayck CP, Sheikh A. Effect of smoke-free legislation on perinatal and child health: a systematic review and meta-analysis. The Lancet. 2014;383(9928):1549-60.

17. Cochrane Effective Practice and Organisation of Care (EPOC). EPOC Resources for review authors, 2017. Available from: https://epoc.cochrane.org/resources/epoc-resources-review-authors.

18. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

19. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health. 2012;66(12):1182-6.

20. Cummings P. The relative merits of risk ratios and odds ratios. Archives of Pediatrics & Adolescent Medicine. 2009;163(5):438-45.

21. Barendregt JJ, Veerman JL. Categorical versus continuous risk factors and the calculation of potential impact fractions. Journal of Epidemiology & Community Health. 2010;64(3):209-12.

22. Laverty AA, Hone T, Vamos EP, Anyanwu PE, Taylor-Robinson D, de Vocht F, et al. Impact of banning smoking in cars with children on exposure to second-hand smoke: a natural experiment in England and Scotland. Thorax. 2020;75(4):345-347

23. Nguyen HV. Do smoke-free car laws work? Evidence from a quasi-experiment. Journal of Health Economics. 2013;32(1):138-48.

24. Patel M, Thai CL, Meng Y-Y, Kuo T, Zheng H, Dietsch B, et al. Smoke-free car legislation and student exposure to smoking. Pediatrics. 2018;141(Supplement 1):S40-S50.

25. Elton-Marshall T, Leatherdale ST, Driezen P, Azagba S, Burkhalter R. Do provincial policies banning smoking in cars when children are present impact youth exposure to secondhand smoke in cars? Preventive Medicine. 2015;78:59-64.

26. Azagba S, Kennedy RD, Baskerville NB. Smoke-free school policy and exposure to secondhand smoke: a quasi-experimental analysis. Nicotine & Tobacco Research. 2016;18(2):170-6.

27. Lee SL, Wong WHS, Lau YL. Smoke-free legislation reduces hospital admissions for childhood lower respiratory tract infection. Tobacco Control. 2016;25(e2):e90-e4.

28. Ho SY, Wang MP, Lo WS, Mak KK, Lai HK, Thomas GN, et al. Comprehensive smoke-free legislation and displacement of smoking into the homes of young children in Hong Kong. Tobacco Control. 2010;19(2):129-33.

29. Chan S, Leung D, Mak Y, Leung G, Leung S, Lam T. New anti-smoking legislation on second-hand smoke exposure of children in homes. Hong Kong medical journal. 2011;17(3 Suppl 3):38-42.

30. Chan SSC, Cheung YTD, Leung DYP, Mak YW, Leung GM, Lam TH. Secondhand smoke exposure and maternal action to protect children from secondhand smoke: pre-and post-smokefree legislation in Hong Kong. PLoS One. 2014;9(8):e105781.

31. Gagné T, Lapalme J, Ghenadenik AE, JL OL, Frohlich K. Socioeconomic inequalities in secondhand smoke exposure before, during and after implementation of Quebec's 2015 'An Act to Bolster Tobacco Control'. Tob Control. 2020:tobaccocontrol-2020-056010.

32. Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. Annual review of public health. 2017;38:39-56.

33. Greenland S. Can meta-analysis be salvaged? American Journal of Epidemiology. 1994;140(9):783-7.

34. Bailar JC, 3rd. The promise and problems of meta-analysis. The new England journal of medicine. 1997;337(8):559-61.

35. Jaakkola M, Jaakkola J. Assessment of exposure to environmental tobacco smoke. European Respiratory Journal. 1997;10(10):2384-97.

36. Hovell MF, Zakarian JM, Wahlgren DR, Matt GE, Emmons KM. Reported measures of environmental tobacco smoke exposure: trials and tribulations. Tobacco Control. 2000;9(suppl 3):iii22-iii8.

37. Sendzik T, Fong GT, Travers MJ, Hyland A. An experimental investigation of tobacco smoke pollution in cars. Nicotine & tobacco research. 2009;11(6):627-34.

38. Sureda X, Fernández E, López MJ, Nebot M. Secondhand tobacco smoke exposure in open and semi-open settings: a systematic review. Environmental Health Perspectives. 2013;121(7):766-73.

39. Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, Akl EA, Guyatt GH. Chapter 15: Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2. Cochrane, 2021.

40. Global Asthma Network. The Global Asthma Report 2014. Auckland, New Zealand: Global Asthma Network, 2014.

41. World Bank. World Bank Open Data. Available from: https://data.worldbank.org/.

42. Been JV, Millett C, Lee JT, van Schayck CP, Sheikh A. Smoke-free legislation and childhood hospitalisations for respiratory tract infections. European Respiratory Journal. 2015;46(3):697-706.

43. Been JV, Mackay DF, Millett C, Soyiri I, van Schayck CP, Pell JP, et al. Smoke-free legislation and paediatric hospitalisations for acute respiratory tract infections: national quasi-experimental study with unexpected findings and important methodological implications. Tob Control. 2018;27(e2):e160-e6.

44. Hawkins SS, Hristakeva S, Gottlieb M, Baum CF. Reduction in emergency department visits for children's asthma, ear infections, and respiratory infections after the introduction of state smoke-free legislation. Prev Med. 2016;89:278-85.

45. González-Salgado IL, Rivera-Navarro J, Sureda X, Franco M. Qualitative examination of the perceived effects of a comprehensive smoke-free law according to neighborhood socioeconomic status in a large city. SSM - population health. 2020;11:100597.

46. Zhou L, Niu L, Jiang H, Jiang C, Xiao S. Facilitators and barriers of smokers' compliance with smoking bans in public places: a systematic review of quantitative and qualitative literature. International journal of environmental research and public health. 2016;13(12):1228.

47. World Health Organization. Report on the global tobacco epidemic: offer help to quit tobacco use: executive summary. 2019.

48. Schreuders M, Nuyts PA, van den Putte B, Kunst AE. Understanding the impact of school tobacco policies on adolescent smoking behaviour: a realist review. Social Science & Medicine. 2017;183:19-27.

49. Kegler MC, Lea J, Lebow-Skelley E, Lefevre AM, Diggs P, Haardörfer R. Implementation and enforcement of smoke-free policies in public housing. Health Education Research. 2019;34(2):234-46.

50. Goel S, Ravindra K, Singh RJ, Sharma D. Effective smoke-free policies in achieving a high level of compliance with smoke-free law: experiences from a district of North India. Tobacco control. 2014;23(4):291-4.

51. Platter HN, Pokorny SB. Smoke-free signage in public parks: impacts on smoking behaviour. Tobacco control. 2018;27(4):470-3.

52. Rennen E, Nagelhout GE, van den Putte B, Janssen E, Mons U, Guignard R, et al. Associations between tobacco control policy awareness, social acceptability of smoking and smoking cessation. Findings from the International Tobacco Control (ITC) Europe Surveys. Health Education Research. 2014;29(1):72-82.

53. Turner S, Mackay D, Dick S, Semple S, Pell JP. Associations between a smoke-free homes intervention and childhood admissions to hospital in Scotland: an interrupted time-series analysis of whole-population data. The Lancet Public Health. 2020;5(9):e493-e500.

54. World Health Organization. Evaluating the effectiveness of smoke-free policies. Lyon: IARC Press, International Agency for Research on Cancer; 2009.

55. Fabian LE, Bernat DH, Lenk KM, Shi Q, Forster JL. Smoke-free laws in bars and restaurants: does support among teens and young adults change after a statewide smoke-free law? Public health reports. 2011;126(5):669-76.

56. Fong GT, Hyland A, Borland R, Hammond D, Hastings G, McNeill A, et al. Reductions in tobacco smoke pollution and increases in support for smoke-free public places following the implementation of comprehensive smoke-free workplace legislation in the Republic of Ireland: findings from the ITC Ireland/UK Survey. Tobacco control. 2006;15(suppl 3):iii51-iii8.

57. Thomson G, Wilson N. One year of smokefree bars and restaurants in New Zealand: Impacts and responses. BMC Public Health. 2006;6(1):1-9.

58. Verdonk-Kleinjan WM, Rijswijk PC, de Vries H, Knibbe RA. Compliance with the workplace-smoking ban in the Netherlands. Health Policy. 2013;109(2):200-6.

59. Zhou Y, Mak Y, Ho G. Effectiveness of Interventions to Reduce Exposure to Parental Secondhand Smoke at Home among Children in China: A Systematic Review. International Journal of Environmental Research and Public Health. 2019;16(1):107.

60. Martínez C, Guydish J, Robinson G, Martínez-Sánchez JM, Fernández E. Assessment of the smoke-free outdoor regulation in the WHO European Region. Preventive medicine. 2014;64:37-40.

61. Lupton JR, Townsend JL. A systematic review and meta-analysis of the acceptability and effectiveness of university smoke-free policies. Journal of American College Health. 2015;63(4):238-47.

62. Snyder K, Vick JH, King BA. Smoke-free multiunit housing: a review of the scientific literature. Tobacco control. 2016;25(1):9-20.

63. Licht AS, Hyland A, Travers MJ, Chapman S. Secondhand smoke exposure levels in outdoor hospitality venues: a qualitative and quantitative review of the research literature. Tobacco control. 2013;22(3):172-9.

64. Smith CE, Hill SE, Amos A. Impact of population tobacco control interventions on socioeconomic inequalities in smoking: a systematic review and appraisal of future research directions. Tob Control. 2020:tobaccocontrol-2020-055874.

65. Murray CJ, Aravkin AY, Zheng P, Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1223-49.

SUPPLEMENTAL MATERIAL

Appendix 1: Additional sources used to extract information on the policies

Canadian smoke-free car policies:

Nova Scotia. Smoke-free Places Act. 2020; https://nslegislature.ca/sites/default/files/legc/statutes/smoke-free%20places.pdf.

Ontario. Smoke-Free Ontario Act. 2019; https://www.ontario.ca/laws/statute/17s26.

British Columbia (Canada). Tobacco and Vapour Products Control Act [RSBC 1996] CHAPTER 451. Queen's Printer 2020; https://www.bclaws.ca/civix/document/id/complete/statreg/96451_01#section2.2.

Prince Edward Island. Smoke-Free Places Act. 2018; https://www.princeedwardisland.ca/sites/default/files/legislation/s-04-2-smoke-free_places_act.pdf.

New Brunswick. Smoke-free Places Act. 2011; http://laws.gnb.ca/en/showdoc/cs/2011-c.222.

Manitoba. The Smoking and Vapour Products Control Act. 2020; https://web2.gov.mb.ca/laws/statutes/ccsm/s150e.php.

Saskatchewan (Canada). The Tobacco and Vapour Products Control Act. 2019; https://pubsaskdev.blob.core.windows.net/pubsask-prod/1441/T14-1.pdf.

Newfoundland. Tobacco and Vapour Products Act. 2020; https://www.assembly.nl.ca/Legislation/sr/statutes/t04-1.htm.

Alberta. Tobacco and Smoking Reduction Act. 2018; https://www.qp.alberta.ca/documents/Acts/T03P8.pdf.

Quebec. Tobacco Control Act. 2020; https://www.quebec.ca/en/health/advice-and-prevention/healthy-lifestyle-habits/smoke-free-lifestyle/tobacco-control-act/.

England smoke-free car policy:

Independent. Law banning smoking in cars with children yields just one fine in first year since introduction. 2016; https://www.independent.co.uk/news/uk/home-news/law-ban-smoking-cars-children-one-fine-in-first-year-a7416186.html.

The Guardian. Just one person fined in UK for smoking with under-18 in car 2016; https://www.theguardian.com/society/2016/nov/13/just-one-person-fined-in-uk-for-smoking-with-under-18-in-car.

Canadian comprehensive smoke-free policy covering outdoor areas and cars:

National Assembly. An Act to bolster tobacco control. 2015;

http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=5&file=2015C28A.PDF

Québec. Tobacco Control Act. Last update: November 25, 2016;

https://www.quebec.ca/en/health/advice-and-prevention/healthy-lifestyle-habits/smoke-free-lifestyle/tobacco-control-act/#c437

Supplemental Table 1: Characteristics of studies meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention outcome prevalence	Statistical techniques	Control variables	Risk of bias
Smoke-free car policy										
Elton-Marshall (2015)	Controlled before-and-after study	Canada (Nova Scotia)	Youth Smoking Survey	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days	Any vs none	35.4%	Difference-in-Differences using weighted logistic regression with bootstrap sampling weights	Pre-intervention temporal trends in the outcomes, sex, school grade, parental smoking, survey module, survey year, and province	Moderate
		Canada (Ontario)		Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days	Any vs none	20.4%			Moderate
		Canada (British Columbia)		Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days	Any vs none	21.2%			Moderate
		Canada (Prince Edward Island)		Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days	Any vs none	23.8%			Moderate

Supplemental Table 1: (Continued) Characteristics of studies meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention outcome prevalence	Statistical techniques	Control variables	Risk of bias
		Canada (Manitoba)		Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days	Any vs none	19.4%			Moderate
		Canada (Saskatchewan)		Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days	Any vs none	36.3%			Moderate
		Canada (Newfoundland/ Labrador)		Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days	Any vs none	24.8%			Moderate
Faber (2019)	Interrupted time series	United Kingdom (England)	Health Survey for England	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past days	Any vs none	9.4%	Interrupted time series using survey-weighted logistic regression	Pre-intervention temporal trends in the outcomes, quarter of the year, sex, age, index of multiple deprivation quintile, and level of urbanisation	Low
				Secondary: TSE in places not covered by the policy	Child-reported TSE at home	Any vs none	15.9%			Low

6

Supplemental Table 1: (Continued) Characteristics of studies meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention outcome prevalence	Statistical techniques	Control variables	Risk of bias
				Secondary: TSE in places not covered by the policy	Child-reported TSE at other people's home in the past days	Any vs none	23.9%			Low
				Secondary: biomarker of TSE	Salivary cotinine	Detectable vs undetectable	33.1%			Low
				Secondary: biomarker of TSE	Salivary cotinine	<0.1 ng/mL 0.1-0.9 ng/mL ≥1.0 ng/mL	66.9% 23.9% 9.2%	Interrupted time series using survey-weighted ordinal logistic regression		Low
				Secondary: general incidence of wheezing/ asthma	Child-reported current wheezing or asthma	Yes vs no	6.2%	Interrupted time series using survey-weighted logistic regression		Low
				Secondary: general incidence of wheezing/ asthma	Child-reported respiratory condition	Yes vs no	5.4%			Low

Supplemental Table 1: (Continued) Characteristics of studies meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention outcome prevalence	Statistical techniques	Control variables	Risk of bias
Laverty (2020)	Controlled before-and-after study	United Kingdom (England)	England: Smoking Drinking and Drug Use Surveys Scotland: Scottish Health Survey	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past seven days	Regular exposure vs non/ occasional	5.9%	Difference-in-Differences using survey-weighted logistic regression	Age, sex, area deprivation	Moderate
Nguyen (2013)	Controlled before-and-after study	Canada (British Columbia, Manitoba, New Brunswick, Nova Scotia, Prince Edward Island, Ontario, Saskatchewan)	Canadian Tobacco Use Monitoring Survey	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past month	Any vs none	43%	Triple-Differences analysis with ordinary least squares regression	Education, sex, language, household size, marital status, month and year, age, province, opinion on smoking in restaurant, gas price, interaction terms between age group and year, age group and province, and between province and year	Low

6

Supplemental Table 1: (Continued) Characteristics of studies meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention outcome prevalence	Statistical techniques	Control variables	Risk of bias
				Secondary: TSE in places not covered by the policy	Child-reported TSE at bus stops/ shelters in the past month	Any vs none	Not reported			Low
				Secondary: TSE in places not covered by the policy	Child-reported TSE at parks/ sidewalks in the past month	Any vs none	Not reported			Low
				Secondary: TSE in places not covered by the policy	Child-reported TSE inside restaurants in the past month	Any vs none	Not reported			Low
Patel (2018)	Interrupted time series	United States of America (California)	California Student Tobacco Survey	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past month	Any vs none	24.0%	Interrupted time series using survey-weighted logistic regression	Pre-intervention temporal trends in the outcomes, sex, race and/ or ethnicity, and grade level	Low

Supplemental Table 1: (Continued) Characteristics of studies meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention outcome prevalence	Statistical techniques	Control variables	Risk of bias
Smoke-free school policy										
Azagba (2016)	Controlled before-and-after study	Canada (Quebec, British Columbia, Prince Edward Island, Saskatchewan)	Canadian Tobacco Use Monitoring Survey	Primary: TSE in places covered by the policy	Child-reported TSE on school property in the past month	Any vs none	58.5%	Difference-in-Differences using a linear probability model	Sex, child's smoking status, rural or urban residence, province of residence, and season	Moderate
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas										
Lee (2016)	Interrupted time series	Hong Kong	Hospital Authority hospital attendance records	Primary: unplanned hospital attendance for RTIs	Unplanned hospital attendances for lower RTI	Yes vs no	7 per 1000 (incidence of unplanned hospital attendance for RTIs before policy implementation)	Interrupted time series using negative binomial regression	Pre-intervention temporal trends in the outcomes, temperature, relative humidity, rainfall, population counts, daily hospital attendances for influenza, day of the study period, season, day of the week, holidays, PM_{10}, sulphur dioxide, nitrogen dioxide and ozone	Critical

Abbreviations: second-hand smoke (SHS); tobacco smoke exposure (TSE); respiratory tract infection (RTI); particulate matter with aerodynamic diameter <10 μm (PM_{10}).

Supplemental Table 2: Bias assessment for studies meeting the EPOC criteria based on ROBINS-I tool

First author (year of publication)	Outcome	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Smoke-free car policy									
Elton-Marshall (2015)	Child-reported TSE in cars in the past 7 days (Nova Scotia)	Moderate	Low	Low	Low	Low	Low	Low	Moderate
	Child-reported TSE in cars in the past 7 days (Ontario)	Moderate	Low	Low	Moderate	Low	Low	Low	Moderate
	Child-reported TSE in cars in the past 7 days (British Colombia)	Moderate	Low	Low	Low	Low	Low	Low	Moderate
	Child-reported TSE in cars in the past 7 days (Prince Edward island)	Moderate	Low	Low	Low	Low	Low	Low	Moderate
	Child-reported TSE in cars in the past 7 days (Manitoba)	Moderate	Low	Low	Low	Low	Low	Low	Moderate
	Child-reported TSE in cars in the past 7 days (Saskatchewan)	Moderate	Low	Low	Low	Low	Low	Low	Moderate
	Child-reported TSE in cars in the past 7 days (Newfoundland/ Labrador)	Moderate	Low	Low	Low	Low	Low	Low	Moderate

Supplemental Table 2: (Continued) Bias assessment for studies meeting the EPOC criteria based on ROBINS-I tool

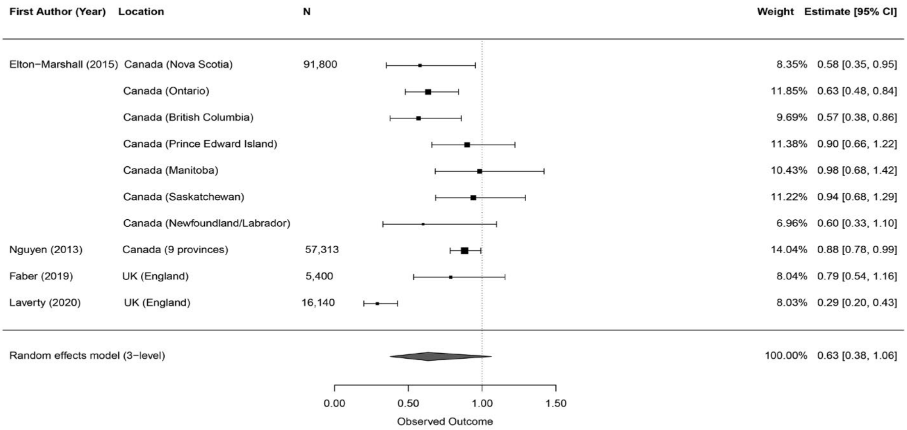
First author (year of publication)	Outcome	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Faber (2019)	Child-reported TSE in cars in the past days	Low	Low	Low	Low	Low	Low	Low	Low
	Child-reported TSE at home	Low	Low	Low	Low	Low	Low	Low	Low
	Child-reported TSE at other people's home in the past days	Low	Low	Low	Low	Low	Low	Low	Low
	Salivary cotinine	Low	Low	Low	Low	Low	Low	Low	Low
	Child-reported current wheezing or asthma	Low	Low	Low	Low	Low	Low	Low	Low
	Child-reported respiratory condition	Low	Low	Low	Low	Low	Low	Low	Low
Laverty (2019)	Child-reported TSE in cars in the past 7 days	Low	Low	Moderate	Low	No information	Moderate	Low	Moderate
Nguyen (2013)	Child-reported TSE in cars in the past month	Low	Low	Low	Low	Low	Low	Low	Low
	Child-reported TSE at bus stops/shelters in the past month	Low	Low	Low	Low	Low	Low	Low	Low
	Child-reported TSE at parks/sidewalks in the past month	Low	Low	Low	Low	Low	Low	Low	Low

6

Supplemental Table 2: (Continued) Bias assessment for studies meeting the EPOC criteria based on ROBINS-I tool

First author (year of publication)	Outcome	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
	Child-reported TSE inside restaurants in the past month	Low	Low	Low	Low	Low	Low	Low	Low
Patel (2018)	Child-reported TSE in cars in the past month	Low	Low	Low	Low	No information	Low	Low	Low
Smoke-free school policy									
Azagba (2016)	Child-reported TSE on school property in the past month	Moderate	Low	Low	Low	No information	Low	Low	Moderate
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas									
Lee (2016)	Unplanned hospital attendances for lower RTI	Critical	Low	Moderate	Low	No information	Low	Low	Critical

Abbreviations: tobacco smoke exposure (TSE); respiratory tract infection (RTI).



Supplemental Figure 1: Meta-analysis of relative risk of child tobacco smoke exposure after implementation of smoke-free car policy (3-level meta-analysis)

Black squares indicate the point estimate for each policy, with the horizontal lines representing the 95% confidence intervals. The pooled effect of the ten estimates was obtained using a random-effects 3-level meta-analysis, with the effect of each event weighted by the inverse of its variance. The pooled effect is indicated with the diamond, and the width of the diamond corresponds with the 95% confidence intervals of the pooled effect. N refers to the number of participants in the given study.

Supplemental Table 3: Findings from studies meeting EPOC criteria reporting the association between implementation of smoke-free policies and secondary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Definition of outcomes	Reported intervention effect	Summary of findings
Smoke-free car policy								
Faber (2019)	United Kingdom (England)	Children aged 8-15 years	5400 (without missing values 5399)	2008-2017	Secondary: TSE in places not covered by the policy	Child-reported TSE at home (any vs none)	OR: 1.17 (0.83 to 1.64)	Smoke-free car policy in England was not associated with significant changes in TSE at home
					Secondary: TSE in places not covered by the policy	Child-reported TSE at other people's home in the past days (any vs none)	OR: 0.93 (0.70 to 1.24)	Smoke-free car policy in England was not associated with significant changes in TSE at other people's homes
		Children aged 2-15 years excluding smokers	7999 (without missing values 7858)	2008-2017	Secondary: biomarker of TSE	Salivary cotinine (detectable vs undetectable)	OR: 1.36 (1.09 to 1.71)	Smoke-free car policy in England was associated with an increased risk of having detectable salivary cotinine levels
					Secondary: biomarker of TSE	Salivary cotinine (ordinal)	OR: 1.30 (1.04 to 1.62)	Smoke-free car policy in England was associated with an increased risk of having a higher salivary cotinine categorya
		Children aged 0-15 years	13371 (without missing values 13369)	2008-2017	Secondary: general incidence of wheezing/ asthma	Child-reported current wheezing or asthma (yes vs no)	OR: 0.81 (0.62 to 1.05)	Smoke-free car policy in England was not associated with significant changes in wheezing or asthma

Supplemental Table 3: (Continued) Findings from studies meeting EPOC criteria reporting the association between implementation of smoke-free policies and secondary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Definition of outcomes	Reported intervention effect	Summary of findings
			21096 (without missing values 17006)		Secondary: general incidence of wheezing/ asthma	Child-reported respiratory condition (yes vs no)	OR: 1.02 (0.80 to 1.29)	Smoke-free car policy in England was not associated with significant changes in respiratory conditions
Nguyen (2013)	Canada (Prince Edward Island, Nova Scotia, New Brunswick, Ontario, Manitoba, Saskatchewan, British Columbia)	Children aged 15-16 years, except for Nova Scotia and Prince Edward Island where children were aged 15-19 years	57313 (without missing values 56517)	2005-2010	Secondary: TSE in places not covered by the policy	Child-reported TSE at bus stops/shelters in the past month (any vs none)	Absolute difference (percentage points): -1.3 (-3.7 to 1.1)	Smoke-free car policy in seven Canadian provinces was not associated with significant changes in TSE at bus stops/shelters
			57313 (without missing values 56455)		Secondary: TSE in places not covered by the policy	Child-reported TSE at parks/sidewalks in the past month (any vs none)	Absolute difference (percentage points): -6.6 (-17.0 to 3.8)	Smoke-free car policy in seven Canadian provinces was not associated with significant changes in TSE at parks/sidewalks
			57313 (without missing values 56566)		Secondary: TSE in places not covered by the policy	Child-reported TSE inside restaurants in the past month (any vs none)	Absolute difference (percentage points): -2.3 (-5.2 to 0.6)	Smoke-free car policy in seven Canadian provinces was not associated with significant changes in TSE in restaurants

Abbreviations: odds ratio (OR), 95% confidence interval between brackets; tobacco smoke exposure (TSE)

Supplemental Table 4: Characteristics of studies not meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention prevalence	Statistical techniques	Control variables	Risk of bias
Comprehensive smoke-free policy covering outside areas and cars										
Gagné (2020)	Uncontrolled before-and-after study	Canada (Quebec)	Canadian Community Health Survey	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past month	Yes vs no	15.1%	Logistic regression	Age, sex	Critical
				Secondary: TSE in places not covered by the policy	Child-reported TSE at home, every day or almost every day	Yes vs no	16.2%	Logistic regression	Age, sex	Critical
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas										
Chan (2011)	Uncontrolled before-and-after study	Hong Kong	Data from a previous randomised controlled trial and a cross-sectional survey	Secondary: TSE in places not covered by the policy	Mother-reported child's TSE at home in the past seven days (hours/day)	Occasional, 1, 2-4, 5-7, 8-10	Occasional: 41.4% 1: 28.5% 2-4: 23.7% 5-7: 1.1% 8-10: 0.5%	Multivariate analysis of variance	Mother's education, mother's age, mother's employment status	Critical
				Secondary: TSE in unspecified places	Mother-reported paternal smoking around the child in the past seven days (cigarettes/day)	< 1, 1-4, 5-14, 15-24, >24	< 1: 36.0% 1-4: 41.9% 5-14: 10.8% 15-24: 1.1% >24: 0%	Chi-square	None	Critical
				Secondary: TSE in unspecified places	Father-reported frequency of smoking around the child in the past seven days	Sometimes, usually, always	Sometimes: 60.5% Usually: 4.4% Always: 4.4%			Critical

Supplemental Table 4: (Continued) Characteristics of studies not meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention prevalence	Statistical techniques	Control variables	Risk of bias
				Secondary: General incidence of wheezing/ asthma	Mother-reported child's coughing due to SHS	Yes vs no	Not reported			Critical
				Secondary: General incidence of wheezing/ asthma	Mother-reported child's shortness of breath due to SHS	Yes vs no	Not reported			Critical
Chan (2014)	Uncontrolled before-and-after study	Hong Kong	Data from a previous randomised controlled trial and a cross-sectional survey	Secondary: TSE in places not covered by the policy	Mother-reported child's TSE at home in the past seven days (hours/day)	<1, 1, 2-4, 5–7, 8–10	<1: 39.3% 1: 26.0% 2-4: 20.5% 5–7: 0.9% 8–10: 0.5%	Ordinal logistic regression	Age (father, mother, and child), father's education level, years of father's smoking, father's perceived health status, children's medical consultation in the past month, household income level and number of children in the home	Critical
				Secondary: TSE in unspecified places	Mother-reported paternal smoking around the child in the past seven days (cigarettes/day)	< 1, 1-4, 5-14, 15-14, >14	< 1: 34. 4% 1–4: 38.1% 5–14: 9.6% > 14: 0.9%			Critical
				Secondary: biomarker of TSE	Hair nicotine levels (ng/mg)	Continuous	Median (range): 0.36 (0.09–11.88)	Mann-Whitney test	None	Critical

Supplemental Table 4: (Continued) Characteristics of studies not meeting the EPOC criteria

First author (year)	Study design	Intervention country (region)	Data source	Outcome eligibility	Definition of outcome	Outcome categories	Pre-intervention prevalence	Statistical techniques	Control variables	Risk of bias
Ho (2010)	Uncontrolled before-and-after study	Hong Kong	Two cross-sectional school-based surveys were conducted specifically for this study	Primary: TSE in places covered by the policy	Child-reported TSE outside the home	Any vs none	19.8%	Logistic regression	Sex, age, school grade, place of birth, parental smoking status and the clustering effect of schools	Critical
				Secondary: TSE in places not covered by the policy	Child-reported TSE at home	Any vs none	10.2%			Critical
				Secondary: TSE in unspecified places	Child-reported TSE anywhere	Any vs none	23.2%			Critical
				Secondary: Chronic cough	Child-reported cough or phlegm	Yes vs no	36.6%	Chi-square test	None	Critical

Abbreviations: second-hand smoke (SHS); tobacco smoke exposure (TSE).

Supplemental Table 5: Bias assessment for studies not meeting the EPOC criteria based on ROBINS-I tool

First author (year of publication)	Outcome	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Comprehensive smoke-free policy covering outside areas and cars									
Gagné (2020)	Child-reported TSE in cars in the past month	Critical	Low	Low	Low	Unknown	Low	Moderate	Critical
	Child-reported TSE at home, daily	Critical	Low	Low	Low	Unknown	Low	Moderate	Critical
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas									
Chan (2011)	Mother-reported child's TSE at home in the past 7 days	Critical	Low	Low	Low	Low	Low	Moderate	Critical
	Mother-reported paternal smoking around the child in the past 7 days	Critical	Low	Low	Low	Low	Low	Moderate	Critical
	Father-reported frequency of smoking around the child in the past 7 days	Critical	Low	Low	Low	Low	Low	Moderate	Critical
	Mother-reported child's coughing due to SHS	Critical	Low	Low	Low	Unknown	Low	Moderate	Critical

6

Supplemental Table 5: (Continued) Bias assessment for studies not meeting the EPOC criteria based on ROBINS-I tool

First author (year of publication)	Outcome	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
	Mother-reported child's shortness of breath due to SHS	Critical	Low	Low	Low	Unknown	Low	Moderate	Critical
Chan (2014)	Mother-reported child's TSE at home in the past 7 days	Critical	Low	Low	Low	Low	Low	Moderate	Critical
	Mother-reported paternal smoking around the child in the past 7 days	Critical	Low	Low	Low	Low	Low	Moderate	Critical
	Hair nicotine levels	Critical	Low	Low	Low	Low	Low	Moderate	Critical
Ho (2010)	Child-reported exposure to smoking outside home	Critical	Moderate	Low	Low	No information	Low	Low	Critical
	Child-reported exposure to smoking at home	Critical	Moderate	Low	Low	No information	Low	Low	Critical
	Child-reported exposure to smoking anywhere	Critical	Moderate	Low	Low	No information	Low	Low	Critical
	Child-reported cough or phlegm	Critical	Moderate	Low	Low	No information	Low	Low	Critical

Abbreviations: second-hand smoke (SHS); tobacco smoke exposure (TSE).

Supplemental Table 6: Findings from studies not meeting the EPOC criteria reporting on the association between implementation of smoke-free policies and the primary and secondary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Outcome	Reported intervention effect	Summary of findings
Smoke-free car policy								
Gagné (2020)	Canada (Quebec)	Children aged 12-17 years	4975 to 5019 depending on the variables	2013-2018	Primary: TSE in places covered by the policy	Child-reported TSE in cars in the past month	OR: 0.38 (0.27 to 0.53)	Smoke-free car policy in Quebec was associated with a reduction in TSE in cars
					Secondary: TSE in places not covered by the policy	Child-reported TSE at home, every day or almost every day	OR: 0.55 (0.41 to 0.73)	Smoke-free car policy in Quebec was associated with a reduction in TSE at home
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas								
Chan (2011)	Hong Kong	Children aged 0-12 years	928 (without missing values 918)	2005-2008	Secondary: TSE in places not covered by the policy	Mother-reported child's TSE at home in the past seven days (ordinal)	P-value: <0.001	Comprehensive smoke-free policy in Hong Kong was associated with a significant reduction in TSE at home
			928 (without missing values 891)		Secondary: TSE in places not covered by the policy	Mother-reported paternal smoking around the child in the past seven days (ordinal)	P-value: <0.001	Comprehensive smoke-free policy in Hong Kong was associated with a significant reduction in mother-reported paternal smoking around the child
			794 (without missing values 776)		Secondary: TSE in places not covered by the policy	Father-reported paternal smoking around the child in the past seven days (ordinal)	P-value: <0.001	Comprehensive smoke-free policy in Hong Kong was associated with a significant reduction in father-reported paternal smoking around the child

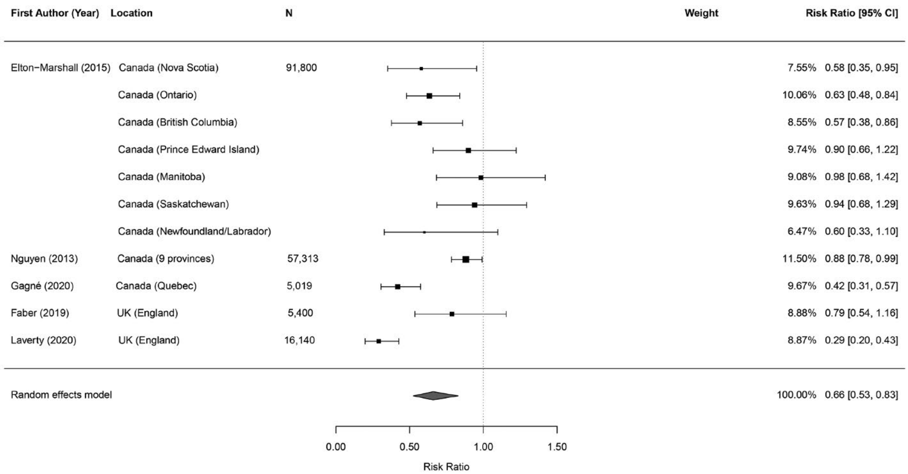
Supplemental Table 6: (Continued) Findings from studies not meeting the EPOC criteria reporting on the association between implementation of smoke-free policies and the primary and secondary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Outcome	Reported intervention effect	Summary of findings
			928 (missing values unknown)		Secondary: general incidence of wheezing/ asthma	Mother-reported child's coughing due to SHS (yes vs no)	P-value: <0.001	Comprehensive smoke-free policy in Hong Kong was associated with a significant reduction coughing due to SHS
			928 (missing values unknown)		Secondary: general incidence of wheezing/ asthma	Mother-reported child's shortness of breath due to SHS (yes vs no)	P-value: <0.001	Comprehensive smoke-free policy in Hong Kong was associated with a significant reduction in shortness of breath due to SHS
Chan (2014)	Hong Kong	Children aged 0-18 months	402 (without missing values 400)	2005-2008	Secondary: TSE in places not covered by the policy	Mother-reported child's TSE at home in the past seven days (ordinal)	OR: 0.08 (0.04 to 0.14)	Comprehensive smoke-free policy in Hong Kong was associated with a reduction TSE at home
			402 (without missing values 396)		Secondary: TSE in places not covered by the policy	Mother-reported paternal smoking around the child in the past seven days (ordinal)	OR: 0.05 (0.02 to 0.09)	Comprehensive smoke-free policy in Hong Kong was associated with an 83% reduction in mother-reported paternal smoking around the child
			68 (without missing values 68)		Secondary: biomarker of TSE	Hair nicotine levels (ng/mg)	P-value: <0.01	Comprehensive smoke-free policy in Hong Kong was associated with a significant reduction in hair nicotine levels

Supplemental Table 6: (Continued) Findings from studies not meeting the EPOC criteria reporting on the association between implementation of smoke-free policies and the primary and secondary outcomes

First author (year)	Country (region)	Intervention population	Sample size	Observational period	Outcome eligibility	Outcome	Reported intervention effect	Summary of findings
Ho (2010)	Hong Kong	Children in US grades 2-4	8208 (missing values unknown)	2006-2008	Primary: TSE in places covered by the policy	Child-reported TSE outside the home (any vs none)	OR: 1.60 (1.26 to 2.03)	Comprehensive smoke-free policy in Hong Kong was associated with an increase in TSE outside the home
			8208 (missing values unknown)		Secondary: TSE in places not covered by the policy	Child-reported TSE at home (any vs none)	OR: 1.56 (1.25 to 1.92)	Comprehensive smoke-free policy in Hong Kong was associated with an increase in TSE at home
			8208 (missing values unknown)		Secondary: TSE in places not covered by the policy	Child-reported TSE anywhere (any vs none)	OR: 1.54 (1.25 to 1.89)	Comprehensive smoke-free policy in Hong Kong was associated with an increase in TSE anywhere
			8208 (missing values unknown)		Secondary: Chronic cough	Child-reported cough or phlegm (yes vs no)	Absolute difference (percentage points): 2 (CI unknown) P-value: <0.001	Comprehensive smoke-free policy in Hong Kong was associated with an increase in cough or phlegm

Abbreviations: odds ratio (OR), 95% confidence interval between brackets; tobacco smoke exposure (TSE).



Supplemental Figure 2: Meta-analysis of relative risk of child tobacco smoke exposure in cars after implementation of smoke-free car policy (including study not meeting EPOC criteria)

Black squares indicate the point estimate for each policy, with the horizontal lines representing the 95% confidence intervals. The pooled effect of the ten estimates was obtained using a random-effects meta-analysis, with the effect of each event weighted by the inverse of its variance. The pooled effect is indicated with the diamond, and the width of the diamond corresponds with the 95% confidence intervals of the pooled effect. N refers to the number of participants in the given study.

Supplemental Table 7: Asthma cases that could be prevented by introducing smoke-free car policies for various scenarios

Association between TSE in cars and respiratory disease	Prevalence of TSE in cars before policy (%)	Relative reduction in respiratory disease (%)
RR = 1.12 Association with asthma for children with 1 to 2 days per week of TSE in cars as compared to children with no TSE in cars	5	-0.2 (-0.1 to -0.3)
	15	-0.5 (-0.2 to -0.8)
	25	-0.9 (-0.4 to -1.3)
	35	-1.2 (-0.5 to -1.8)
	45	-1.6 (-0.7 to -2.3)
RR = 1.19 Association with asthma for children with 3 to 7 days per week of TSE in cars as compared to children with no TSE in cars	5	-0.3 (-0.1 to -0.4)
	15	-0.9 (-0.4 to -1.2)
	25	-1.4 (-0.6 to -2.0)
	35	-1.9 (-0.8 to -2.8)
	45	-2.4 (-1.0 to -3.5)

Abbreviations: tobacco smoke exposure (TSE); risk ratio (RR); respiratory tract infection (RTI).

Supplemental Table 8: Studies meeting EPOC criteria reporting on elements supporting causal inference

First author (year)	Alternative comparison groups	Neutral outcome	Alternative methods
Smoke-free car policy			
Elton-Marshall (2015)	Findings were robust to adding participants to the analysis who had not been in a car or did not know if they were exposed to TSE in a car.	NR	NR
Faber (2019)	Findings were robust to adding participants from the additional booster samples to the analysis. Booster samples were included in some years to oversample certain population subgroups. Accordingly, the combined sample (i.e., original and the boost sample) is not generalisable and therefore not recommended to be used in the main analyses.	NR	Additional analysis showed that the implementation of the smoke-free car policy was not associated with significant changes in active smoking or TSE in cars among adults. However, the smoke-free car policy was associated with an increase in smoking at home for smokers older than 24 years of age, but a decrease in smoking in other people's homes for smokers aged 16–24 years.

Supplemental Table 8: (Continued) Studies meeting EPOC criteria reporting on elements supporting causal inference

First author (year)	Alternative comparison groups	Neutral outcome	Alternative methods
Laverty (2020)	Findings were robust to restricting the sample to children in England only, using the wider age range of 11 to 15 years. It showed that decreases in TSE were observed for all levels of baseline exposure after the introduction of the policy.	NR	NR
Nguyen (2013)	Findings were robust to adding a term to the model to examine whether there was a decline in TSE even before the policy was adopted. This indicated that TSE did not decline before the policy was introduced.	NR	Additional analysis showed that the implementation of the smoke-free car policy was not associated with an increase in smoking at home among smokers. Additional analysis using a subsample of participants who reside in urban areas and whose parents do not have a car confirmed the hypothesis that these children are unlikely to be affected by the policy.
Patel (2018)	California data were compared with national data for comparable time periods. The California trends in TSE in cars could not be explained by secular trends on the national level.	NR	NR
Smoke-free school policy			
Azagba (2016)	NR	NR	NR
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas			
Lee (2016)	NR	NR	NR

Abbreviations: tobacco smoke exposure (TSE); not reported (NR).

Supplemental Table 9: Studies not meeting EPOC criteria reporting on elements supporting causal inference

First author (year)	Alternative comparison groups	Neutral outcome	Alternative methods
Comprehensive smoke-free policy covering outside areas and cars			
Gagné (2020)	NR	NR	NR
Comprehensive smoke-free policy covering indoor public, (semi-)private, and outside areas			
Chan (2011)	NR	NR	Survey among fathers and mothers indicated that more than 95% of the parents were aware of the new policy. Mothers reported that they undertook more actions to protect their children from TSE, and that they were more active in helping their husband to quit smoking after the policy as compared to before the policy. Fathers reported that they had increased their motivation to quit smoking, recognition of the importance to quit smoking, confidence in quitting smoking, and less difficulty in quitting, following the implementation of the policy.
Chan (2014)	NR	NR	Survey among fathers and mothers to evaluate the impact of the policy. Mothers whose children were exposed to TSE were more likely to protect their children from TSE, take their children away from TSE, place a 'no-smoking' sign at home, and advise fathers to avoid smoking near the children after the policy implementation as compared to before the policy. There was no difference in the proportion of mothers who advised, took action, or supported fathers in quitting smoking after the policy implementation as compared to before the policy. Fewer mothers helped the fathers use substitutes for cigarettes after the policy implementation.
Ho (2010)	Findings were robust to restricting the population to children from schools that participated in both time periods before and after the policy was introduced.	NR	NR

Abbreviations: tobacco smoke exposure (TSE); not reported (NR).

Part II

Methods and approaches to evaluate natural experiments

7

A systematic review of the effect of infrastructural interventions to promote cycling: Strengthening causal inference from observational data

Famke J.M. Mölenberg, Jenna Panter, Alex Burdorf, Frank J. van Lenthe

International Journal of Behavioral Nutrition and Physical Activity 2019:16(1):93

ABSTRACT

Introduction

Previous reviews have suggested that infrastructural interventions can be effective in promoting cycling. Given inherent methodological complexities in the evaluation of such changes, it is important to understand whether study results obtained depend on the study design and methods used, and to describe the implications of the methods used for causality. The aims of this systematic review were to summarize the effects obtained in studies that used a wide range of study designs to assess the effects of infrastructural interventions on cycling and physical activity, and whether the effects varied by study design, data collection methods, or statistical approaches.

Methods

Six databases were searched for studies that evaluated infrastructural interventions to promote cycling in adult populations, such as the opening of cycling lanes, or the expansion of a city-wide cycling network. Controlled and uncontrolled studies that presented data before and after the intervention were included. No language or date restrictions were applied. Data was extracted for any outcome presented (e.g., bikes counted on the new infrastructure, making a bike trip, cycling frequency, cycling duration), and for any purpose of cycling (e.g., total cycling, recreational cycling, cycling for commuting). Data for physical activity outcomes and equity effects were extracted, and quality assessment was conducted following previous methodologies and the UK Medical Research Council guidance on natural experiments. The PROGRESS-Plus framework was used to describe the impact on subgroups of the population.

Studies were categorized by outcome, i.e., changes in cycling behavior, or usage of the cycling infrastructure. The relative change was calculated to derive a common outcome across various metrics and cycling purposes. The median relative change was presented to evaluate whether effects differed by methodological aspects.

Results

The review included 31 studies and all were conducted within urban areas in high-income countries. Most of the evaluations found changes in favour of the intervention, showing that the number of cyclists using the facilities increased (median relative

change compared to baseline: 62%; range: 4% to 438%), and to a lesser extent that cycling behavior increased (median relative change compared to baseline: 22%; range: -21% to 262%). Studies that tested for statistical significance and studies that used subjective measurement methods (such as surveys and direct observations of cyclists) found larger changes than those that did not perform statistical tests, and those that used objective measurement methods (such as GPS and accelerometers, and automatic counting stations). Seven studies provided information on changes of physical activity behaviors, and findings were mixed. Three studies tested for equity effects following the opening of cycling infrastructure.

Conclusion

Study findings of natural experiments evaluating infrastructural interventions to promote cycling depended on the methods used and the approach to analysis. Studies measuring cycling behavior were more likely to assess actual behavioral change that is most relevant for population health, as compared to studies that measured the use of cycling infrastructure. Triangulation of methods is warranted to overcome potential issues that one may encounter when evaluating environmental changes within the built environment.

Review registration

The protocol of this study was registered at PROSPERO (CRD42018091079).

INTRODUCTION

Promoting physical activity is one of the key strategies to combat the burden of many chronic diseases (1). Cycling can contribute to meeting the recommended daily physical activity levels (2, 3). A meta-analysis including 187,000 individuals and 2.1 million person-years showed that 2.5 hours per week of cycling at moderate intensity was associated with a 10% lower mortality risk, independent of overall levels of physical activity (4). In addition to this, a Danish study found that those who cycled *and*, those who started cycling after the age of 50 years had a lower risk of coronary heart disease and developing diabetes than those who did not cycle (5, 6). Modelling studies have also showed that the population health benefits of cycling outweigh the negative risks, such as exposure to air pollution and traffic accidents (7, 8). This indicates that promoting cycling can result in population-level health benefits.

Providing an infrastructure that supports the needs of cyclists has been considered as an important strategy to encourage more cycling in cities (9-11). However, designing studies to evaluate such infrastructural interventions is challenging. Although randomized controlled trials (RCTs) are regarded as the gold-standard for estimating causal effects of health interventions, to our knowledge no studies exist that used the RCT design to assess the impact of infrastructural interventions on cycling. This is not surprising, as changes in the built environment are often beyond control of the researcher and therefore difficult to randomize. Other analytical techniques are required to evaluate these so-called "natural experiments", in which variation in accessibility to new cycling infrastructure is used to assign intervention and control groups (12-14).

Two recent systematic reviews have been completed which examine the impact of infrastructure on levels of cycling (15, 16). Both reported that cycling increased following the introduction of new infrastructure, or upgrading of existing infrastructure. However, both reviews also noted that the methods in the included studies may have affected the study findings. Stappers and colleagues (15) noted variable quality in study designs across studies examining impacts on physical activity, active transport and sedentary behavior. They suggest that more refined designs may decrease the possibility of detecting intervention effects. Panter and colleagues (16) focused only on studies assessing walking and cycling, and examined the evidence for the effectiveness and mechanism of interventions. They found that higher quality studies were more likely to report intervention effects for cycling. Taken together, differences in methods may have impacted the overall conclusion (no changes vs positive changes), or the magnitude of the finding (small changes vs large changes). Ignoring methodological differences may wrongly lead to the conclusion that some interventions were more effective than others.

The current review builds on the main finding of previous reviews that interventions in the built environment may affect cycling (15, 16). We focused on the methodological approaches undertaken to evaluate the effects of infrastructural interventions. Both reviews did not quantitatively summarize the findings, thereby leaving the question unanswered if the magnitude of the findings changed when using different methodology. One review was unable to capture relevant literature published outside of health-related journals. The research questions are likely to be different between health researchers and transportation researchers, potentially leading to differences in study designs and findings.

Focusing on whether different methodological approaches produce different results, and assessing the strengths and limitations of different methods for causality, will provide greater understanding about the implications of findings from research and their utility for policy makers and practitioners. Therefore, the aims of this systematic review were to summarize the effects of infrastructural interventions on cycling and physical activity in the population, and to evaluate whether the effects varied by study design, data collection methods, or statistical approaches.

METHODS

The protocol of this study was registered in March 2018 at PROSPERO (CRD42018091079). Our systematic literature search followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (17).

Search strategy

Various electronic databases (Embase.com, Medline Ovid, Web of Science, PsycINFO Ovid, CINAHL EBSCOhost, Google scholar) were searched for literature published until February 2018 for any studies assessing infrastructural projects to promote cycling. We updated the initial search until June 2019 to additionally include most recent publications. Search terms were constructed of 3 parts, including synonyms for cycling infrastructure to identify exposures, synonyms for cycling behavior, active transport, physical activity and lifestyle changes to identify outcomes, and a term that excluded conference abstracts, letter to the editors, notes and editorials. No restrictions were made on language. Database searches were supplemented with searches of reference lists of included studies and key review papers.

Study selection and inclusion criteria

All titles and abstracts identified during the initial search were screened for inclusion by two independent researchers (FJMM, NB). Additional articles identified through the updated search were screened by a single author (FJMM). After screening titles and abstracts, full-text articles were screened according to predefined criteria. Articles obtained in full-text were reassessed for inclusion by the first two authors (FJMM, JP), and discrepancies were resolved after discussion with a third researcher (FJvL). Eligibility criteria included: 1) a study evaluating an infrastructural intervention to promote cycling, 2) any measure of cycling as outcome, 3) cycling measured before and after the intervention, and 4) reporting on a general adult population aged 16 years and above. Examples of interventions include the opening of cycling lanes, the installation of a city-wide cycling network, and the improvement of existing cycling infrastructure. We included papers that evaluated the same intervention, but reported on different outcomes or used different datasets or methods to collect outcome data. Controlled and uncontrolled studies were included to allow for a large variety of study designs. Studies were classified as controlled studies if data was collected in a different population that was selected based on comparable individual or neighbourhood characteristics, and if similar data collection methods were used. We also classified studies as controlled studies if a comparison was made within the study population between people who lived closer to an intervention and those who lived further away. Studies that presented city- or area-wide cycling trends as a comparison were considered uncontrolled, as the data collection methods used in routine monitoring surveys often differed from that used in the intervention group, and population characteristics often differed between areas.

Studies that evaluated the introduction of cycling infrastructure together with other environmental components were included (i.e., bike parking, showers, rental bikes), as long as the main goal of the intervention was to promote cycling. Environmental interventions that did not change the cycling infrastructure were excluded. We specifically aimed to study population-based approaches to change health behaviors, and therefore excluded infrastructural interventions that were part of a combined intervention with behavioral components targeting the behavior of individuals (i.e., cycling courses, safety lessons, or other approaches that target individual behaviors). Studies that included media campaigns along the intervention were included, as long as they aimed to target the population as a whole.

We excluded opinion articles, qualitative evaluations without quantitative assessment, studies retrospectively collecting data on cycling, and studies not directly linked to an infrastructural intervention. We also excluded studies in which the presented outcome measure was not specified for cycling, like active travel which combined walking and cycling together, or modal shifts where the shift in mode was not specified.

Data extraction

From the included studies, one researcher extracted data (FJMM) using a standardized data extraction form, and a second reviewer (JP) verified a 20% sample of the extracted data. The extracted data included publication details, description of the intervention, study design, data collection methods, analytical methodology, and study results.

Ideally, we would have extracted a single outcome related to cycling per study. However, most studies did not specify a primary outcome of cycling. Therefore, we extracted all cycling outcomes presented from the maximally adjusted model with the longest exposure time. We extracted all outcomes for various purposes of cycling (e.g., total cycling, recreational cycling, cycling for commuting), and all outcomes for various metrics of cycling (e.g., bike count data, cycling frequency, cycling duration). If the outcome was assessed in multiple populations or at multiple locations, we extracted the average change in cycling that was presented by the authors. If no summary measure was presented, we calculated an unweighted average effect. Some studies stratified the population by exposure status, and evaluated a possible exposure-outcome relationship by distance from home to the intervention or usage of the intervention. All available information was extracted for these studies and included in the descriptive part of the review. However, including all strata-specific outcomes in the quantitative analyses would mean that studies with multiple strata would have a much greater contribution to the findings than studies without stratification. Therefore, we only used the results from the group most likely to use the intervention in the quantitative summary (e.g., smallest distance or largest potential usage). We noted that various metrics were used for expressing data relevant to cycling. We distinguished outcomes that evaluated cycling behavior (e.g., making a bike trip, cycling frequency, cycling duration) from those that evaluated usage of cycling infrastructure (e.g., bikes counted in the city, bikes counted on the new infrastructure). We extracted data on both absolute change (no fixed unit, can refer to various metrics) and relative change (expressed as percentage change over time) in cycling between before and after measurements, and attempted to calculate outcomes for both where possible. We used a similar framework presented by Goodman (18) to compute measures of absolute and relative change. Outcomes expressed as ratios were interpreted as relative changes. For uncontrolled studies, the relative change was computed by dividing the absolute change by the baseline level of cycling in the study sample. For controlled studies, we first computed the relative change in the intervention and control group separately. Subsequently, the calculated relative change in the intervention group was divided by the calculated relative change in the control group. Likewise, to obtain an absolute change when only relative changes were presented, we multiplied the relative change by the baseline estimate in the study sample as a whole for uncontrolled studies, and by the baseline estimate in the control group for controlled studies. Examples of the data extracted and how outcomes were calculated are presented in Appendix 1. Authors were contacted if only the direction of the association was presented. For each study we extracted data on statistical tests performed, and if significant results were found ($P<0.05$). However, we focused on directions of the association rather than significance, since a substantive part of the studies did not test for significant changes in cycling outcomes that were of interest for this review.

We extracted data on the methodological quality, and on all design elements and additional analyses that may have supported causal inference following previous methodologies. The quality items described by Ogilvie et al. (19) were extracted, which used the criteria from the Community Guide of the US Task Force on Community Preventive Services to assess study design (20), and criteria developed for the Effective Public Health Practice Project in Hamilton, Ontario to score five items related to the quality of the research performed (21). The five items included representativeness, comparability, credibility of data collection instruments, attrition rate, and attributability of the effect to the intervention. The original instrument also assessed randomization, but this was not assessed as the allocation to the intervention and comparison group was not under control of the researcher. In addition, we extracted the results from additional analyses that may support causal inference identified by the UK Medical Research Council guidance on natural experiments (12), including multiple comparison groups, the inclusion of a neutral outcome that is not expected to change as a consequence of the new cycling infrastructure, and the use of complementing research methodologies.

The PROGRESS-Plus framework was used to describe the impact of the infrastructural interventions on subgroups of the population (22). The PROGRESS-Plus framework considers nine factors for which differences in effect may occur: 1) place of residence, 2) race, ethnicity, culture, language, 3) occupation, 4) gender, sex, 5) religion, 6) education, 7) socioeconomic status, 8) social capital, and 9) the 'Plus'-factor that could be other characteristics associated with social disadvantage. In our study we considered age, health status or BMI, bike ownership, and car ownership as Plus-factors, since these factors may have been relevant determinants of disadvantage given the context of the intervention.

Data synthesis

We provided a descriptive narrative synthesis of studies. There was no possibility to quantitatively summarize the results, because of the large variety of outcome metrics and purposes of cycling presented, the lack of a primary outcome, and the lack of a common outcome across studies. Therefore, we presented the median relative change for the umbrella-terms *cycling behavior* and *infrastructure usage* for all studies, and by study design (controlled vs uncontrolled; exposure time ≥1 year *vs* <1 year), data collection methods (objective *vs* subjective), and analytical approaches (tested *vs* not tested). We did not present units for the median relative change because it can refer to various metrics. For example, an increase in cycling behavior of 30% could refer to an increase in the proportion of cyclists, cycling frequency, or cycling duration.

An overview of studies with baseline characteristics or performed adjusted analyses by any of the PROGRESS-Plus factors was presented. We provided a descriptive narrative synthesis for the studies that formally tested for differential effects on PROGRESS-Plus factors.

RESULTS

Study characteristics

From the 3,542 potential records, 125 full-text articles were screened and this resulted in 31 studies (29 interventions) from 11 countries that met the eligibility criteria (Figure 1). Table 1 presents the characteristics of included studies categorized by the outcome of interest. 20 studies presented data on cycling behavior (23-42), and 16 studies assessed usage of the cycling infrastructure (23, 29, 31, 38, 42-53). All infrastructural interventions were conducted in urban areas in high-income countries. The interventions were very diverse in terms of design and scale, ranging from the introduction of a cycling bridge, single or multiple cycle paths or lanes, or a city-wide cycling network. Six studies (five interventions) described issues related to data collection due to delays in the construction work, resulting in shorter follow-up periods than planned (23, 31-34, 39). In addition to this, three studies (two interventions) mentioned that the intervention was not fully completed within the study time frame (31, 33, 34). Most studies used a similar analytical approach by comparing a single estimate before the intervention with a single estimate after the intervention, with or without comparing it to changes in a control group. One study used a fixed-effects approach to evaluate the within-person change over time (27), and three studies tested if there was a significant interaction between the intervention and time (29, 31, 35). One study conducted an interrupted time series analyses, whereby the date of the opening of the cycling track was used to set the time of interruption (47).

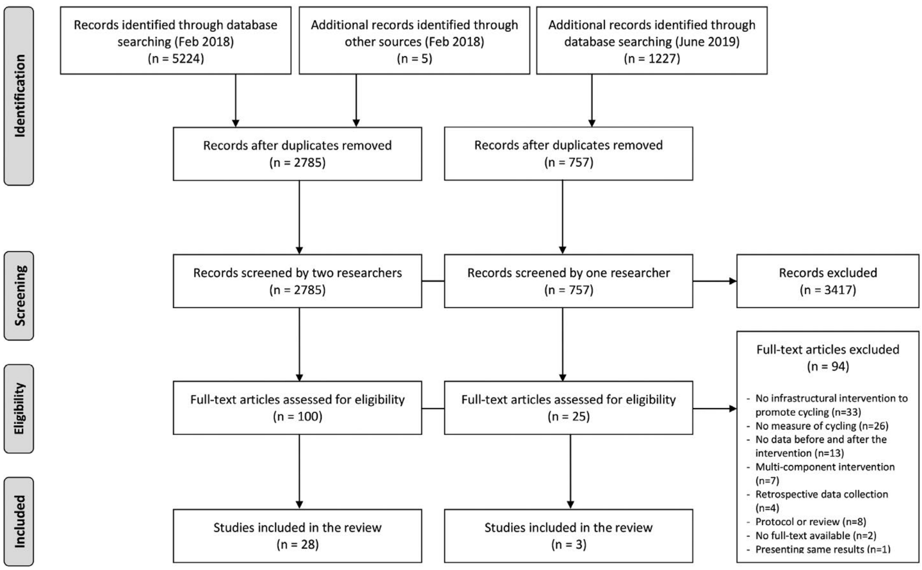


Figure 1: Flow diagram of study selection

Study results

Figure 2 presents an overview of median relative change for all outcomes reported, and according to approach to analysis, exposure time, method of assessment, and whether significance was tested. In general, studies reporting behavioral outcomes found smaller changes than studies presenting usage of the infrastructure. Larger changes were also found for studies that tested for statistical significance and studies that used subjective measurement methods (such as surveys and direct observations of cyclists), compared to studies that did not perform statistical tests, and used objective measurement methods (such as GPS and accelerometers, and automatic counting stations).

Supplemental Table 1 provides further details of the number of studies which assessed cycling behavior or usage of the infrastructure for cycling, and whether these were in favor of the intervention or not. 20 studies presented data on 52 cycling behavior outcomes. All but two (23, 32), found an increase in cycling for at least 1 outcome, and 73% (38/52) of all outcomes presented were in favor of the intervention. A total of 36 cycling behavior outcomes were used to quantitatively summarize the results. Together, studies found a median relative increase in cycling

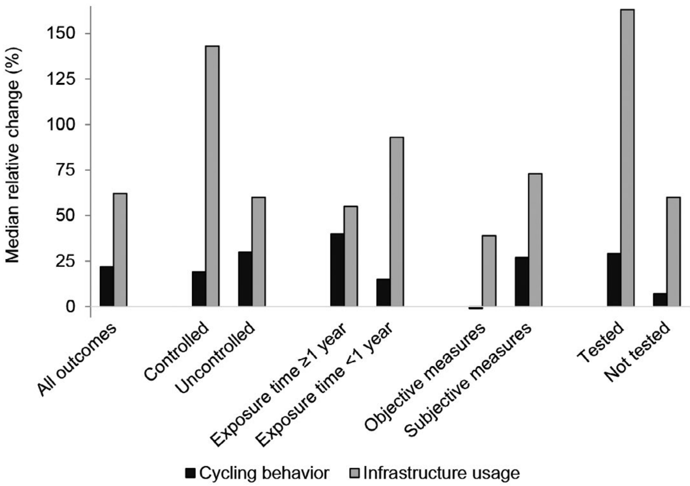


Figure 2: Summary of the results

behavior (median relative change: 23%; range: -21% to 262%). Changes in cycling did not essentially differ between controlled and uncontrolled studies. Studies with an exposure time shorter than 1 year found smaller changes when compared to those using a longer exposure time. Studies that used objective measures to assess cycling behavior found smaller changes than those that used self-reported measures, and studies that did not test for statistical significance found smaller changes than those that did.

Seven studies evaluated changes in physical activity patterns following cycling infrastructure interventions. Brown et al. showed that among cyclists, cycling time on intervention streets increased by 7 min/week and on other streets increased by 6 min/week. Daily energy expenditure increased in the study population by 0.19 kcal/min, which translates into 275 kcal/day (26). Goodman et al. found that living 1 km closer to the intervention increased cycling for recreation by 3 min/week, and total physical activity by 13 min/week (33). There was no evidence that compensation of physical activity behaviors took place, since physical activity excluding walking and cycling was not associated with the intervention. Burbidge et al. did not find changes in total physical activity time, but the number physical activity episodes seemed to have declined by 0.2 trips/day following the introduction of cycling infrastructure

(27). The other four studies did not find evidence that the introduction of cycling infrastructure affected physical activity (29, 31, 32, 39).

Usage of the infrastructure was presented in 16 studies with 21 outcomes, and all were in favor of the intervention (median relative change: 62%; range: 4% to 438%) (Table 2). Changes for infrastructure usage were smaller for studies that were uncontrolled, studies with longer exposure time, studies using automatic counters or GPS tracking information, and studies that did not test for statistical significance (Supplemental Table 1).

Quality assessment

Table 2 presents information on the quality of the studies. Nine out of 20 studies evaluating the impact of cycling infrastructure on cycling behavior presented data on participation, and nine on representativeness. Participation ranged between 2% and 49% for those that presented information. 13 studies collected data twice on the same individual, and retention ranged between 41% and 79%. Most studies used surveys to collect data, but the exact methodology and validity of the question items was often not reported.

When considering the quality of the studies for causal inference, studies reported that other changes in the physical and social environment might have affected or biased their results. Issues reported were the economic crisis, the rising cost of car transport, social marketing campaigns, and other infrastructural improvements during the same period. Authors were often unable to account for these and this could indicate that the changes observed could be partly attributable to other factors. Another problem mentioned is a spill-over effect, indicating that people from control areas might have used the facilities, which may have resulted in an underestimation of the effect. Some studies used multiple groups to test robustness of the findings by using different comparisons group or applying different cut-off values to define exposure or outcome. Some studies presented data for city- or nation-wide cycling trends (36, 37), or historical time trends (35). None of the studies included a neutral outcome which was hypothesized to be unaffected by the new infrastructure designed to promote cycling, thereby functioning as a control measure that captures time trends in transportation or physical activity behaviors. Complementing methodologies performed were surveys among residents (24-29, 31-34, 38, 39, 42) or employees (23), intercept surveys among infrastructure users (27), surveys among new residents who moved into the study area (27), and bike counts in the study area (23, 29, 31, 38, 42).

16 studies presented data on usage of the infrastructure. Five studies used automatic counting stations or mobile app data to objectively measure cyclist movements for periods between 5 months and 3 years. Others monitored the number of cyclist on selected hours and days using observation techniques. Issues that authors reported that may have partly contributed to the increase in infrastructure usage were tunneling of existing riders to the new infrastructure, other infrastructural changes, traffic conditions, rising cost of car transport, weather conditions and seasonality, demographic changes, social marketing, and changing methodology to collect data. One study indicated that improvements made to the cycling infrastructure could have been a consequence of high cycling levels in specific areas (51). Some studies presented data for city- or nation-wide cycling trends (29, 45, 50), or historical time trends (29). Additional methodologies included surveys among residents (31, 38, 42, 45, 50, 51) or employees (23), survey among infrastructure users (45-47, 51), and data collected on cycling behavior (23, 29, 31, 38, 42).

Equity effects

Figure 3 shows that studies assessing cycling behavior collected information on population characteristics more often than those assessing usage, thereby potentially providing insights in the population under study and characteristics of those engaging in cycling, and allowing a comparison of intervention and control groups according to baseline characteristics. The items that were most often used by behavioral studies to describe the population at baseline were age (75%), gender (70%) and a measure of socio-economic status (SES) (50%). Only three studies tested for differential effects on cycling by population subgroups. Aldred et al. did not find any differential effects by demographic and socio-economic characteristics (24). Goodman et al. showed that the change in cycling behavior was larger if there was no car in the household (33). Parker et al. showed that the increase in cyclists was larger among females than males (53).

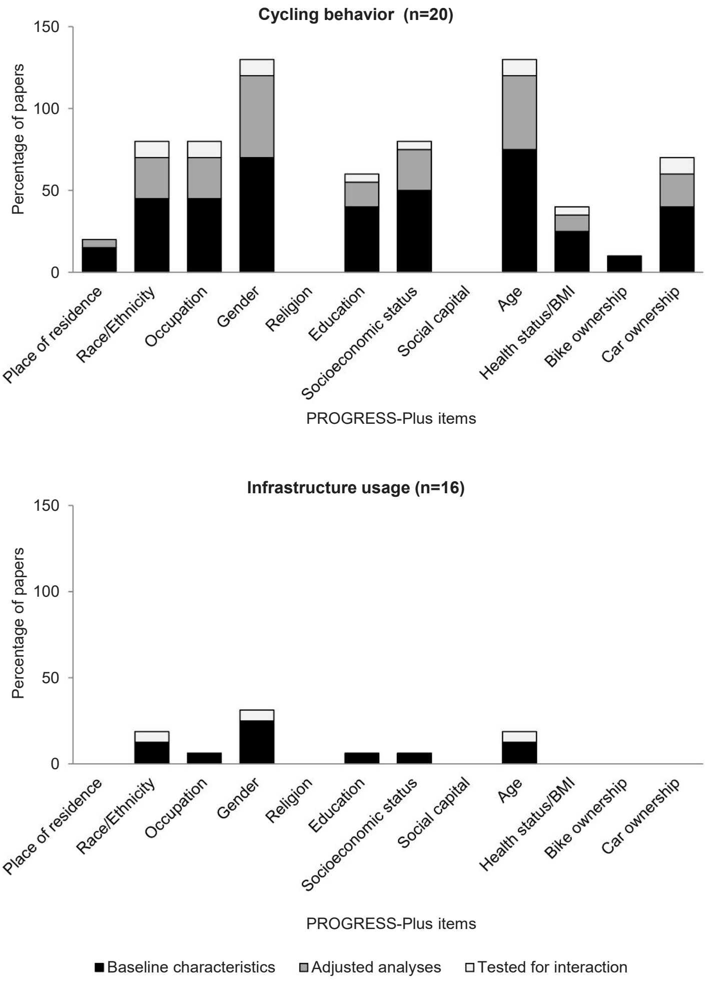


Figure 3: Percentage of studies that presented equity characteristics from the PROGRESS-Plus framework

Table 1: Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Cycling behaviour								
Aittasalo (23) (Finland)	Environmental improvements made to the main and connecting walking and cycling paths	No	Employees working at workplaces in the area where new infrastructure was introduced	Survey Time between measurements; 18-24 months Time exposed; 2 months	Cycling frequency as part of the journey to work (days/week)	Difference over time, tested by Wilcoxon Signed Rank Test	Unadjusted	Not in favor of the intervention, not significant A: Not applicable R: Not applicable
					Cycling distance as part of the journey to work (km/trip)			Not in favor of the intervention, not significant A: Not applicable R: Not applicable
					Cycling time as part of the journey to work (min/trip)			Not in favor of the intervention, not significant A: Not applicable R: Not applicable
					Cycling frequency as part of the journey from work (days/week)			Not in favor of the intervention, not significant A: Not applicable R: Not applicable

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
					Cycling distance as part of the journey from work (km/trip)			Not in favor of the intervention, not significant A: Not applicable R: Not applicable
					Cycling time as part of the journey from work (min/trip)			Not in favor of the intervention, not significant A: Not applicable R: Not applicable
Aldred (24) (UK)	Infrastructural interventions in 3 neighborhoods, transforming local environments for walking and cycling	Yes	Residents living in the intervention areas vs control areas	Travel diary Time between measurements; 12 months Time exposed; not specified but could range from 1-12 months	Made a bike trip in the past week (yes-no)	Difference-in-difference, tested by regression models	Demographic variables, socioeconomic variables, health indicator, and car ownership	In favor of the intervention, not significant A: 3.2%-point R: 16%
					Cycling time (min/week)			In favor of the intervention, not significant A: 4 min/week R: 14%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
		Yes	Residents living in low-dose or high-dose areas (defined by stakeholders involved in implementation) in the intervention areas vs control areas	Travel diary Time between measurements; 12 months Time exposed; not specified but could range from 1-12 months	Made a bike trip in the past week (yes-no)	Difference-in-difference, tested by regression models	Demographic variables, socioeconomic variables, health indicator, and car ownership	All comparisons in favor of the intervention Low-dose area: not significant A: 0.7%-point R: 10% High-dose area: significant A: 7.2%-point R: 24%
					Cycling time (min/week)			All comparisons in favor of the intervention: Low-dose area: not significant A: 1 min/week R: 5% High-dose area: not significant A: 9 min/week R: 30%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Brown (25) (US)	Complete street intervention including the completion of an incomplete bike lane (10.7 km), connecting the airport to down town districts	Yes	Residents living near (≤0.8 km) vs far (0.8-2 km) from the new infrastructure	GPS and accelerometers Time between measurements; 12 months Time exposed; 1-8 months	Made a bike trip on the intervention road (yes-no)	Difference-in-difference, but no statistical test conducted	Demographic and socioeconomic variables	Not in favor of the intervention, significance not tested A: 0%-point R: -11%
Brown (26) (US)	Same as above	No	Residents living within 2 km of the new infrastructure	GPS and accelerometers Time between measurements; 12 months Time exposed; 1-8 months	Cycling time on the intervention road among those who cycled (min/week)	Difference tested by paired t-test	Unadjusted	In favor of the intervention, not significant A: 7 min/week R: 38%
					Cycling time off the intervention road among those who cycled (min/week)			In favor of the intervention, not significant A: 6 min/week R: 15%

7

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Burbidge and Goulias (27) (US)	Installation of a multi-use trail, creating a 4-km loop connecting two currently existing sidewalks, serving as transportation and recreation facility	No	Residents living within 1.6 km of the new infrastructure	Travel diary Time between measurements; 12 months Time exposed; 5 months	Total cycling trips (trips/day)	Difference tested by fixed effects	By design: controlled for day of the week	In favor of the intervention, not significant A: 0.01 trips/day R: 33%
Chowdhury (28) (New Zealand)	Introduction of a 3 cycle ways linking suburbs with the central business district, and the associated promotional campaigns	No	Residents living in the city where the new infrastructure was introduced	Survey Time between measurements; 4 years Time exposed; 12 months	Cycling at least weekly (yes-no)	Difference, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested A: 10%-point R: 40%
Crane (29) (Australia)	A new cycle way (2.4 km) linking a new urban renewal area with the central business district	Yes	Residents living in the intervention area (suburbs surrounding the cycle way) vs a control area (matched for demographic characteristics)	Survey Time between measurements; 23-25 months Time exposed; 15-17 months	Cycling at least weekly (yes-no)	Difference-in-difference tested by regression models that included a two-way interaction term between time and proximity	Demographic variables	In favor of the intervention, not significant A: 44%-point R: 179%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
		Yes	Residents living closer (<1 km, 1-3 km) vs further (>3 km) from the new infrastructure	Travel diary Time between measurements; 23-25 months Time exposed; 15-17 months	Cycling duration (min/week)			Those living <1 km of the intervention: not in favor of the intervention, not significant A: -37 min/week R: -21% Those living 1-3 km from the intervention: in favor of the intervention, significant A: 96 min/week R: 54%
Deegan (30) (UK)	Extension of a city-wide cycling network aiming for 900 km, unfinished.	No	Residents living in 31 intervention areas. Area-wide cycling trends in 2 control areas are presented for comparison	Survey (census data) Time between measurements; 10 years Time exposed; not specified but could range from 1-10 years	Proportion of commuting trips made by bike (%)	Difference-in-difference, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested Average in 31 areas: A: not reported R: 87% Average in 2 control areas: A: not reported R: 75%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Dill (31) (US)	Installation of 8 bicycle boulevards (1.4 km to 6.7 km long)	Yes	Residents living within 0.3 km of the 8 intervention streets vs residents living within 0.3 km of the 11 control streets (selected to be similar in urban form and demographic characteristics)	GPS and accelerometers Time between measurements; 12 months Time exposed; 2-12 months	Cycling at least 10 minutes a day (yes-no)	Difference-in-difference tested by regression models that included a two-way interaction term between treatment and period	Demographic variables, weather conditions, distance to downtown, bike attitudes and car safety attitudes	In favor of the intervention, not significant A: 9%-point R: 22%
					Cycling time (min/day) for those cycling at least 10 min/ day			Not in favor of the intervention, significant A: -1 min/day R: -1%
					Made a bike trip (yes-no)			Not favor of the intervention, not significant A: -8 %-point R: -15%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
					Number of bike trips for those that made a bike trip			Not favor of the intervention, not significant A: -0.4 trips/day R: -9%
Evenson (32) (US)	Extension of an existing trail (4.5 km), along with a spur (3.2 km) passing by schools, shopping areas, apartment buildings, and residential areas	No	Residents living in census blocks that are crossed by the intervention	Telephone interview Time between measurements; 19-28 months Time exposed; 2 months	Median cycling time (min/week)	Difference tested by Wilcoxon nonparametric test for differences	Unadjusted	Not in favor of the intervention, not significant A: 0 min/week R: 0%
					Median cycling time for transportation (min/month)			Not in favor of the intervention, not significant A: 0 min/week R: 0%

7

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Goodman (33) (UK)	Construction of new walking and cycling infrastructure and improvement of existing routes in 3 cities plus a modest amount of promotion activities	Yes	Residents living within 5 km of the new infrastructure using proximity for comparison (per 1 km closer to the intervention)	7-day recall instrument Time between measurements; 24 months Time exposed; 7-21 months	Cycling time for transport (min/week)	Difference-in-difference tested by regression models	Demographic variables, socioeconomic variables, health indicator, and car ownership	Not in favor of the intervention, not significant A: -0.2 min/week R: not reported
				Survey	Cycling time for recreation (min/week)			In favor of the intervention, significant A: 2.5 min/week R: not reported
Song (34) (UK)	Same as above	No	Residents living within 5 km of the new infrastructure	7-day recall instrument Time between measurements; 24 months Time exposed; 7-21 months	Cycling time for utility purpose (min/week)	Difference over time tested by paired sample t-test	Unadjusted	In favor of the intervention, not significant A: 0.4 min/week R: 2%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
					Cycling distance for utility purpose (km/week)			In favor of the intervention, not significant A: 0.4 km/week R: 7%
Hirsch (35) (US)	Expansion of two trails (16.3 km), including a bicycle and pedestrian bridge connecting residential areas to employment centers downtown and at the university	No	Residents living in 116 areas of the city with the new infrastructure. Historical time trends are presented for comparison	Survey (census data) Time between measurements; 10 years Time exposed; not specified but could range from 3-10 years	Proportion of workers who commuted by bike (%)	Difference over time, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested A: 2.3%-point R: 130% Historical trend: A: 0.1%-point R: not reported

7

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
		Yes	Residents living in 116 areas of the city with the new infrastructure using distance to the intervention for comparison (results presented for the 25th, 50th and 75th percentiles)		Proportion of workers who commuted by bike (%)	Difference-in-difference tested by regression models that included a two-way interaction term between time and treatment	Demographic variables, socioeconomic variables, cycling infrastructure characteristics, total work-related trips, proportion of trips that cross the trail system	All comparisons in favor of the intervention, and all significant 25th percentile (1.1 km): A: 2.0%-point R: 115% 50th percentile (2.8km): A: 1.9%-point R: 107% 75th percentile (5.9 km): A: 1.6%-point R: 92%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
		Yes	Residents living in 116 areas of the city with the new infrastructure using proportion of commuting trips crossing the trial for comparison (results presented for the 25th, 50th and 75th percentiles)		Proportion of workers who commuted by bike (%)	Difference-in-difference tested by regression models that included a two-way interaction term between time and treatment	Demographic variables, socioeconomic variables, cycling infrastructure characteristics, total work-related trips, distance to the trail	All comparisons in favor of the intervention, and all significant 25th percentile (11%): A: 1.0%-point R: 54% 50th percentile (29%): A: 1.9%-point R: 107% 75th percentile (42%): A: 2.6%-point R: 146%
		Yes	Residents living in 116 areas of the city with the new infrastructure using the joined effect of distance and trips crossing the trial for comparison		Proportion of workers who commuted by bike (%)	Difference-in-difference tested by regression models that included a two-way interaction term between time and treatment	Demographic variables, socioeconomic variables, cycling infrastructure characteristics, total work-related trips, proportion of trips that cross the trail system, distance to the trail	In favor of the intervention The increase in bicycle commuting was restricted to tracts that were close to the intervention, and had a higher proportion of commuting trips that crossed the trials

7

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Krizek (36) (US)	Installation of multiple bicycle facilities and major bridge improvements to enhance accessibility to major employment centers	No	Residents living in areas within 1.6 km of the geographical centroids of a new facility. Area-wide cycling trends are presented for comparison	Survey (census data) Time between measurements; 10 years Time exposed; not specified but could range from 1-10 years	Bicycle mode share (%)	Difference tested by regression models	Not reported	In favor of the intervention, significant A: 0.2%-point R: 14% Whole area: A: 0.02%-point R: 5%
			Residents living in areas within 1.6 km of the geographical centroids of a new facility, or within 0.8 km from the endpoints of a facility					In favor of the intervention, significant A: 0.5%-point R: 46%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
		No	Bicycle mode share crossing the river. Cycling trends that remained on the same side of the river are presented for comparison	Survey (census data) Time between measurements; 10 years Time exposed; not specified but could range from 1-10 years	Bicycle mode share crossing the river (%)	Difference tested by regression models	Not reported	In favor of the intervention, significant Crossing river: A: 1.6%-point R: 52% Average that remained at the same side of the river: A: 0.6%-point R: 28%
Lanzendorf (37) (Germany)	Cycling infrastructure improvements and marketing campaigns in 4 cities	No	Residents living in cities with the new infrastructure. Cycling trends in big cities are presented for comparison	Survey enriched with regional data Time between measurements; 6 years Time exposed; not specified but could range from 1-6 years	Cycling frequency (trips/day)	Difference over time, tested by Mann-Whitney U-test	Not reported	In favor of the intervention, significant Average of 4 cities: A: 0.07 trips/day R: 27% Big cities: A: 0.09 trips/day R: 31%

7

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
					Bicycle mode share (%)	Difference over time, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested Average of 4 cities: A: 1.8%-point R: 21% Big cities: A: 2.4%-point R: 24%
Merom (38) (US)	The construction of a cycle way (16.5 km) and the associated promotional campaigns	Yes	Residents living near (<1.5 km) vs far (1.5-5 km) from the new infrastructure	Telephone interviews Time between measurements; 4 months Time exposed; 3 months	Cycling time among those who cycled (min/week)	Difference-in-difference tested by ANOVA	Not reported	In favor of the intervention, significant A: 26 min/week R: 147%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Panter (39) (UK)	New bus network and an adjacent traffic-free walking and cycling route (22 km)	Yes	Residents working in the city with the new infrastructure, and living within ~30 km of work using proximity for comparison (results presented comparing those living 4 km from the intervention vs 9 km)	7-day recall instrument Time between measurements; 3 years Time exposed; 9-14 months	Likelihood of an increase in cycling time for commuting (yes-no)	Difference-in-difference tested by regression models	Demographic variables, socioeconomic variables, health indicators, car ownership and work related variables	In favor of the intervention, significant A: 87 min/week (among those who reported more cycling for commuting at follow-up) R: 34%
				Survey	Likelihood of an increase in total cycling time (yes-no)			In favor of the intervention, significant A: 115 min/week (among those who reported more cycling at follow-up) R: 32%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Pedroso (40) (US)	Infrastructure expansion in bicycle lanes (147 km) and improvements in bicycle signage, parking, and cyclist awareness, and the addition of a bike share program	No	Residents living in the city with the new infrastructure	Survey (census data) Time between measurements; 9 years Time exposed; not specified but could range from 1-7 years	Proportion of workers who commuted by bike (%)	Difference over time tested by regression models	Not reported	In favor of the intervention, significant A: 1.5%-point R: 167%
Smith (41) (US)	Bicycle lane expansion (>160 km), and the introduction of bicycle share programs	No	Residents living in the city with the new infrastructure	Survey (census data) Time between measurements; 5 years Time exposed; 4 years	Number of cyclist	Difference over time tested by t-test	Not reported	In favor of the intervention, significant A: 4388 cyclist R: 262%
Wilmink and Hartman (42) (the Netherlands)	Improvements to an existing cycle route network, creating a comprehensive and interconnected network	No	Residents living in two neighborhoods with the new infrastructure	Postal home interview survey Time between measurements; 3 years Time exposed; not specified but could range from 1-3 years	Proportion of trips made by bike (%)	Difference-in-difference, no statistical test conducted	Not reported	In favor of the intervention, significance not tested A: 3%-point R: 7%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
		Yes	Residents living in two neighborhoods with the new infrastructure vs one control neighborhood without the new infrastructure		Cycling frequency (trips per person per day)			In favor of the intervention, significance not tested A: not reported R: 4%
					Cycling distance (distance per person per day)			In favor of the intervention, significance not tested A: not reported R: 8%
Usage of the infrastructure								
Aittasalo (23) (Finland)	Environmental improvements made to the main and connecting walking and cycling paths	No	4 locations in the study area	Automatic counters Time between measurements; 24 months Time exposed; 2 months	Bikes per day during afternoon peak hour	Difference, no statistical test conducted	Not reported	In favor of the intervention, significance not tested Average: A: 367 bikes/peak hour R: 57%

7

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Barnes (43) (US)	Complete street redesign of a gateway to the university to improve the conditions for non-motorized users	No	1 location on the study road	Direct observation Time between measurements; 6 months Time exposed; not specified but could range from 1-6 months	Bikes per hour	Difference, no statistical test conducted	Not reported	In favor of the intervention, significance not tested Average: A: 63 bikes/hour R: 83%
Crane (29) (Australia)	A new cycle way (2.4 km) linking a new urban renewal area with the central business district	No	2 locations along the new infrastructure. City-wide cycling trends and historic time trends are presented for comparison	Automatic counters Time between measurements; 36 months Time exposed; 16 months	Bikes per day during peak hours (6h/day)	Difference, no statistical test conducted	If adjusted, estimated were adjusted for population growth	In favor of the intervention, significance not tested Average: A: 144 bikes/peak hours (unadjusted) R: 4% (adjusted)

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
								City as a whole: A: -80 bikes/peak hours (unadjusted) R: -2% (adjusted) Historical trend: A: 300 bikes/peak hours (unadjusted) R: 126% (unadjusted) Historical trend, city as a whole: A: 300 bikes/peak hours (unadjusted) R: 111% (unadjusted)
Dill (31) (US)	Installation of 8 bicycle boulevards (1.4 km to 6.7 km long)	No	10 locations on the study roads	Method not described Time between measurements; 3 years Time exposed; 18 months	Number of bikes	Difference, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested A: not reported R: 22%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Fitzhugh (44) (US)	Retrofitting a neighborhood with an urban trail (4.6 km) that enhanced connectivity to retail and school destinations	Yes	1 location in the intervention neighborhood vs 2 locations in 2 control neighborhoods (matched along socioeconomic dimensions)	Direct observation Time between measurements; 2 years Time exposed; 14 months	Median number of bikes per 2 hours	Difference-in-difference tested by Wilcoxon rank sums test	Not reported	In favor of the intervention, significant A: 2.2 bikes/2 hour R: 224%
Goodno (45) (US)	The installation of two linked bicycle facilities serving downtown	No	4 locations on the study roads. City-wide cycling trends are presented for comparison	Methods not described; Time between measurements; 18-20 months Time exposed; 7-12 months	Bikes during peak hour	Difference, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested Average of 4 locations: A: 124 bikes/peak hour R: 438% City as a whole: A: 20 bikes/peak hour R: 32%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Hans (46) (Denmark)	Improvements made to two large, interconnected bicycle infrastructures (18 km and 15 km) in city suburbs to enhance connectivity	No	2 locations on the study roads	Automatic counters, calibrated by visual counts Time between measurements; 35 months Time exposed; 16-22 months	Bikes per hour on weekdays during the rush hour in day light	Difference over time, but no statistical test conducted	Seasonal, weather and temporal variables	In favor of the intervention, significance not tested Average of the 2 locations: A: 43 bikes/hour R: 47%
					Bikes per hour on weekdays during the rush hour in dark			In favor of the intervention, significance not tested Average of the 2 locations: A: 38 bikes/hour R: 72%
					Bikes per hour on weekdays during the non-rush hour in day light			In favor of the intervention, significant Average of the 2 locations: A: 11 bikes/hour R: 19%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
					Bikes per hour on weekend days in day light			In favor of the intervention, significant Average of the 2 locations: A: 10 bikes/hour R: 29%
Heesch (47) (Australia)	The opening of three new segments of a cycling lane (1.4 km, 0.9 km, 2.3 km) connecting the suburbs and the city center	No	1 location on the study road before the intervention, 2 locations on the study road after the intervention	Direct observation Time between measurements; 4 years and 1 month Time exposed; 3-38 months	Bikes per 2.5 hour	Difference over time, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested A: 376 bikes/2.5 hour R: 276%
	The opening of the last segment of a cycling lane (2.3 km) connecting the suburbs and the city center	Yes	GPS tracking information on the study road vs 3 other routes surrounding the intervention	Mobile phone application Time between measurements; 1 year Time exposed; 6 months	Trend in monthly bike trips on the intervention road	Interrupted time-series	Seasonal variables	In favor of the intervention, significant A: 225 bike trips / month R: not applicable

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
		No	GPS tracking information on the major routes between suburbs and city center, including the intervention		Trend in monthly bike trips between suburbs and the city center			In favor of the intervention, significant A: 90 bike trips / month R: 102%
Law (48) (UK)	The introduction of superhighways for cyclists creating continuous cycling routes in the city center, and a public bike sharing system	No	21 locations in the intervention area	Direct observations (before) and automatic counters (after) Time between measurements; 9 years Time exposed; not specified but could range from 1-9 years	Bikes per hour	Difference over time, test not described	Not reported	In favor of the intervention, significant A: 154 bikes/hour R: 432%

7

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Marques (49) (Spain)	Introduction of a cycling network in the city (164 km)	No	2000-2005: data from 2006 extrapolated 2006-2010: counts made in the city 2011-2013: algorithm based on count data and the number of rental bikes	Count data, changing methodology over time Time between measurements; 14 years Time exposed; not specified but could range from 1-7 years	Million bikes per year	Difference, no statistical test conducted	Seasonal variables	In favor of the intervention, significance not tested A: 13.3 million trips/ year R: 435%
McCartney (50) (UK)	Construction of a new pedestrian and cyclist bridge across the river towards the city center	No	5 locations to enter the city from the side of the bridge. City-wide cycling trends are presented for comparison	Direct observation; Time between measurements; 4 years Time exposed; 2 years	Bikes counted per 2 days	Difference over time, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested A: 500 bikes/2 days R: 62% Rest of the city: A: 1700 bikes/2 days R: 48%

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Merom (38) (US)	The construction of a cycle way (16.5 km) and the associated promotional campaigns	No	4 locations along the new infrastructure	Automatic counters Time between measurements; 5 months Time exposed; 3 months	Bikes per day	Difference, tested by regression models	Weather variables, day of the week and holiday season	In favor of the intervention, significant Average of the 2 locations: A: Not reported R: 31%
Nguyen (51) (Singapore)	Improvement of 20 street segments (4.8km in total) to complete a well-developed cycling network	Yes	20 intervention street segments vs 55 control street segments	Direct observation Time between measurements; 2 years Time exposed; 12 months	Bikes per hour	Difference-in-difference, but no statistical test conducted	Not reported	In favor of the intervention, significance not tested A: 18 bikes/hour R: 62%
Parker (52) (US)	Introduction of a bike lane (5.0 km) with multiple bus stops, schools, businesses, a police station and private residences located along the intervention	No	1 location on the study road	Direct observation Time between measurements; 12 months Time exposed; 6 months	Bikes per day	Difference tested by regression models	Not reported	In favor of the intervention, significant A: 53 bikes/day R: 58%

7

Table 1: (Continued) Study characteristics of infrastructural interventions to promote cycling

Reference (country)	Infrastructural intervention	Controlled comparison	Type of comparison	Data collection method; time between measurements; time exposed	Outcome studied	Analytical methodology	Confounders	Direction of the results; significance; absolute (A) and relative (R) change
Parker (53) (US)	Introduction of a bike lane (1.6 km) with multiple schools, churches and businesses located along the intervention	Yes	1 location on the study road vs 1 location at 2 control streets	Direct observation Time between measurements; 12 months Time exposed; 3 months	Bikes per day	Difference-in-difference, but no statistical test conducted	Not reported	In favor of the intervention, significant A: 196 bikes/day R: 385%
Wilmink and Hartman (42) (the Netherlands)	Improvements to an existing cycle route network, creating a comprehensive and interconnected network	Yes	Counts made along roads in the intervention neighborhoods vs counts made in the control neighborhood	Count data, methods not described Time between measurements; 3 years Time exposed; not specified but could range from 1-3 years	Bike counts	Difference-in-difference, no statistical test conducted	Not reported	In favor of the intervention, significance not tested A: Not reported R: 14%

Table 2: Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Cycling behavior								
Aittasalo (23) (Finland)	C	Participation: 49% Only limited data was available regarding the working-age population in the region. The study population was broadly representative with the general adult population in the region	No comparison group	Survey: no info on validity	45%	· Half of the workplaces went through economic problems and workforce adjustment during the study	No other comparison groups	·Published protocol ·Survey among employees ·Safety monitoring ·Count data
Aldred (24) (UK)	A	Participation: 2% There was an underrepresentation of 16 to 24 year olds, non-white individuals and unemployed individuals. They were more likely to have a car or van in the household, and to have cycled	Comparisons groups were broadly comparable. Adjusted for a wide range of variables	7-day recall instrument with acceptable validity	50%	· Dose response effects were reported · The first interventions were targeting areas perceived as more receptive to cycling and walking interventions	No other comparison groups	·Survey among residents
Brown (25) (US)	A	Participation: 29% Representativeness was not shown	Adjusted for some of the characteristics in which the groups significantly differed at baseline	GPS and accelerometer data, using validated algorithm	59%	· Multiple improvements to other nearby infrastructure · Spill-over effect occurred: control residents were exposed to the intervention	No other comparison groups	·Published protocol ·Survey among residents Health indicators: ·Energy expenditure ·BMI

7

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Brown (26) (US)	C	Same as above	No comparison group	Same as above	Same as above	No comments made	No other comparison groups	Same as above
Burbidge and Goulias (27) (US)	C	Participation was not shown. Study population was older, had less cars in the household and were more often unemployed	No comparison group	1-day activity diary, modified from a validated household activity diary	56%	No comments made	No other comparison groups	·Survey among residents and new residents ·Intercept survey Health indicators: ·Physical activity
Chowdhury (28) (New Zealand)	C	Participation was not shown. Study population was representative	No comparison group	Survey, methods not described	Not applicable	No comments made	No other comparison groups	·Survey among residents
Crane (29) (Australia)	A	Participation was not shown. Study population was highly educated and more physically active than the general population	Adjusted for some of the characteristics in which the groups significantly differed	Survey: no info on validity Travel diary: no info on validity	48%	· No dose response effects were observed · Suburbs furthest away from the cycle way were quite diverse in infrastructure · Spill-over effect occurred: users of the cycle way included participants living in control areas	No other comparison groups	·Published protocol ·Survey among residents ·Count data Health indicators: ·Physical activity ·Quality of life

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Deegan (30) (UK)	C	Participation and representativeness were not shown	No comparison group	Census data, methods not described	Not applicable	· Congestion charge and bombings on public transport resulted in sharp increases in cycling levels	· The increase in cycling in the intervention areas was larger than observed in other areas	·Safety monitoring
Dill (31) (US)	A	Participation: 3% Representativeness was not shown	Adjusted for variables that were tested to be significant	GPS and accelerometer data, shown to successfully predict 79% of the cycling trips	72%	· The city may have chosen to install bicycle boulevards in areas where residents were supportive of new cycling infrastructure · Unknown changes in the physical and social environment in specific areas may have influenced the results · Data collection by means of GPS and accelerometers may have changed behavior	No other comparison groups	·Survey among residents ·Count data Health indicators: ·Physical activity

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Evenson (32) (US)	C	Participation: 47% Study population was more highly educated	No comparison group	Non validated method of interviewing	64%	· Questions mentioning the trail were only asked at follow-up and after assessing cycling behavior · Substitution of physical activity behavior may have occurred	· Comparing users and non-users of the intervention did not change the results	·Survey among residents Health indicators: ·Physical activity
Goodman (33) (UK)	A	Participation: 16% Study population was broadly representative, except that fewer young adults were included, and they were somewhat healthier, better educated, and less likely to have children	Adjusted for a wide range of variables	7-day recall instrument with acceptable validity Survey, validated	42%	· Dose response effects were reported · The increase in cycling was only seen for users of the intervention	· Comparing users and non-users of the intervention showed that the increase in cycling was only seen for users of the intervention	·Published protocol ·Survey among residents Health indicators: ·Physical activity
Song (34) (UK)	C	Same as above	No comparison group	7-day recall instrument with acceptable validity Survey, validated	45%	· The increase in cycling may have resulted from the economic crisis, rising fuel costs, and the ageing of the sample	No other comparison groups	Same as above

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Hirsch (35) (US)	A	Participation and representativeness were not shown.	Adjusted for a wide range of variables	Census data (before) and a community survey (after): no info on validity	Not applicable	· Dose response effects were reported · Unknown if people moving into the neighborhood cycle more, or if existing residents change their behaviors · Other infrastructure changes, including a new light rail service, may have influenced the results	· Historical trends showed that the increase in cycling in the intervention period was larger than in previous years	No other methods used

7

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Krizek (36) (US)	C	Participation and representativeness were not shown.	No comparison group	Census data, methods not described	Not applicable	Many potential factors were listed, but only those with an explanation were listed here: · Minor other infrastructural improvements were made in the study areas · Small demographic differences were not the sole explanation of the results · Intervention areas had already a higher cycling level at baseline. The facilities might be the effect, rather than the cause, of high cycling levels	· The increase in cycling in the intervention area was larger than observed in the area as a whole	No other methods used

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Lanzendorf (37) (Germany)	C	Participation and representativeness were not shown.	No comparison group	National travel survey, valid for comparison over time according to the authors	Not applicable	· Hard to disentangle the effects of the infrastructure and marketing campaigns. A combination of both may have the largest impact	· The increase in cycling in the intervention cities was comparable to the change in other big cities, but larger than in the country as a whole	·Document analysis and expert interview to analyze the development of cycling policies
Merom (38) (US)	A	Participation: 48% Representativeness was not shown.	Not adjusted for characteristics in which the groups statistically differed at baseline	Telephone interviews, validated	79%	No comments made	No other comparison groups	·Survey among residents ·Bike counts ·Campaign reach

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Panter (39) (UK)	A	Participation was not shown. The sample contained a higher percentage of woman, older adults and those with a degree, and a smaller proportion of those who rented their home	Adjusted for a wide range of variables	7-day recall instrument with acceptable validity Survey, validated	41%	· Dose response effects were reported	No other comparison groups	·Published protocol ·Survey among residents Health indicators: ·Physical activity
Pedroso (40) (US)	B	Participation and representativeness were not shown	No comparison group	Census data, methods not described	Not applicable	· Several other programs and interventions were implemented during the study period	No other comparison groups	·Safety monitoring
Smith (41) (US)	C	Participation and representativeness were not shown	No comparison group	Methods not described	Not applicable	· The percentage of cyclists using bike lanes declined over time	No other comparison groups	·Safety monitoring
Wilmink and Hartman (42) (the Netherlands)	A	Participation and representativeness were not shown	No information on comparability	Postal home interview survey, no info on validity	Not shown	· There was no change observed in total mobility over time	No other comparison groups	·Survey among residents ·Bike counts

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Usage of the infrastructure								
Aittasalo (23) (Finland)	C	Not applicable	No comparison group	Automatic counters: 4 locations, continuous measurements for 2 years	Not applicable	· Half of the workplaces went through economic problems and workforce adjustment during the study	No other comparison groups	· Published protocol · Survey among employees · Safety monitoring · Cycling behavior
Barnes (43) (US)	C	Not applicable	No comparison group	54 hours of video observations: before and after at 1 location, 6 days, 4.5 hours per day	Not applicable	· No unusual weather or traffic patterns were observed · It is unclear whether cyclist simply changed their routes	No other comparison groups	· Safety monitoring

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Crane (29) (Australia)	B	Not applicable	No comparison group	Automatic counters: 2 locations, measurements in October for 3 years on weekdays, 6 hours per day	Not applicable	· Results may reflect population growth	· The increase in cyclists was only seen in the intervention area, while it decreased in the city as a whole · Historical trends in the number of cyclists were comparable between the intervention areas and the city as a whole	· Published protocol · Survey among residents · Cycling behavior Health indicators: · Physical activity · Quality of life
Dill (31) (US)	C	Not applicable	No comparison group	Not described in the paper	Not applicable	· Unknown changes in the physical and social environment in specific areas	No other comparison groups	· Survey among residents · Cycling behavior Health indicators: · Physical activity

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Fitzhugh (44) (US)	A	Not applicable	Broadly comparable	72 hours of direct observations: before and after at 3 locations, 2 days, 6 hours per day	Not applicable	· Study neighborhoods were not exposed to any marketing or awareness campaigns · Spill-over effect may have occurred: people cycling may not live in the intervention neighborhood	No other comparison groups	No other methods used
Goodno (45) (US)	C	Not applicable	No comparison group	Not described in the paper	Not applicable	· Weather conditions and seasonality may have influenced the results	· The increase in cyclists in the intervention area was larger than in the city as a whole	· Survey among residents · Survey among business owners · Safety monitoring · Intercept survey
Hans (46) (Denmark)	B	Not applicable	No comparison group	Automatic counters calibrated by visual/ manual counts: 2 locations, continuous measurements for 3 years	Not applicable	· Most of the increase in cyclists can be attributed to switching from alternative routes	No other comparison groups	· Intercept survey

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Heesch (47) (Australia) - direct observations	C	Not applicable	No comparison group	7.5 hours of direct observations: Before: 1 location, 1 day, 2.5 hours After: 2 locations, 1 day, 2.5 hours	Not applicable	· The findings suggest some shifting of cyclist	No other comparison groups	· Intercept survey · Mobile phone application to capture movements of cyclists
Heesch (47) (Australia) - mobile phone application	A	Only 10% of the population uses the app, and those were not representative of the broader cycling community	Comparison streets were all connecting the suburbs and the city center	1 year counts made by a mobile phone application: 4 locations, continuous measurement for 1 year	Not applicable	· The findings suggest some shifting of cyclist · The increase in people using the app may have influenced the results · Data on trips was analyzed, and it is unknown if the same cyclists were travelling more frequent, or if more cyclists were travelling	No other comparison groups	· Intercept survey · Direct observations

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Law (48) (UK)	C	Not applicable	No comparison group	Before: direct observations: 21 locations, 1 day, 10 hours After: automatic counters: 21 locations, 1 day,12 hours	Not applicable	· The intervention effect is likely to be over-estimated due to seasonal differences · Change in data collection methods may have influenced the results	No other comparison groups	· Safety monitoring
Marques (49) (Spain)	B	Not applicable	No comparison group	Counts data, no description of the protocol	Not applicable	· Changes in population were not meaningful · Changes in data procedures over time may have influenced the results	No other comparison groups	· Safety monitoring
McCartney (50) (UK)	B	Not applicable	No comparison group	560 hours of digital video recordings manually checked: before and after at 5 locations, 4 days, 14 hours	Not applicable	· Displacement effects were observed · Weather and seasonality may have influenced the results · Traffic conditions may have influenced the results	· The relative increase in cyclists in the intervention area was larger than in the city as a whole	No other methods used

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Merom (38) (US)	B	Not applicable	No comparison group	Automatic counters: 4 locations, continuous measurements for 5 months	Not applicable	· No comments made	No other comparison groups	· Survey among residents · Cycling behavior · Campaign reach
Nguyen (51) (Singapore)	A	Not applicable	No information on comparability	Direct observations: few weekdays during peak periods, no precise description of the protocol	Not applicable	· Shifting of routes was observed · No major change in land use · Possibly reverse causation since segments that were improved had a high demand before the intervention	· The increase in cyclists was even larger on segments that were already improved before start of the current study	· Survey among residents · Intercept survey

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Parker (52) (US)	C	Not applicable	No comparison group	216 hours of direct observations; Before: 1 location, 10 days, 9 hours After: 1 location, 14 days, 9 hours	Not applicable	· Displacement from other streets may have occurred · It is possible that more people ride a bike because of the rising costs of car ownership · The population increase may have influenced the results, but it is unlikely that this explains the total change in cycling	No other comparison groups	No other methods used
Parker (53) (US)	A	Not applicable	No information on comparability	660 hours of direct observations: before and after at 3 location, 10 days, 11 hours	Not applicable	· Some displacement of cyclists from nearby streets was observed · Change in population size is unlikely to be the reason for the increase in cycling	No other comparison groups	No other methods used

7

Table 2: (Continued) Description of the methodological quality, design elements and additional analyses

Reference (country)	Quality criteria (19)[a]					Methodological items (12)[b]		
	Study design[c]	Participation and representativeness	Comparability at baseline	Credibility of data collection methods	Retention	Attributability of effect to intervention	Multiple comparison groups	Complementing research methodologies
Wilmink and Hartman (42) (the Netherlands)	A	Not applicable	No comparison group	Count data, no description of the protocol: 250 locations	Not applicable	· Population growth may have influenced the findings	No other comparison groups	· Survey among residents · Cycling behavior

[a] None of the studies was a randomized experiment, therefore randomization was not applicable for any of the studies and was not shown.

[b] None of the studies presented data for neutral outcomes that were hypothesized to be unaffected by the new infrastructure designed to promote cycling, therefore this parameter was not shown.

[c] A =controlled before-after study; B = uncontrolled study with at least two before and two after data points; C = uncontrolled study with only 1 before and after data point.

DISCUSSION

We identified 31 studies that assessed the effect of infrastructural interventions on cycling in adult populations. All were conducted in urban areas in high-income countries. Most of the evaluations found effects in favor of the intervention, showing that the number of cyclists using the facilities increased, and to a lesser extent that cycling behavior increased. Studies that collected behavioral data more often provided insights in characteristics of people engaging in cycling as compared to studies that reported bike counts. Seven studies reported on physical activity levels, and findings were mixed. Only three studies tested for equity effects, therefore we cannot draw any conclusions as to whether some population subgroups benefitted more than others. We provided data on relative changes that indicates the magnitude of the findings. We acknowledge that in context where only few people use a bike, large relative changes may result in only small population-health benefits. However, due to the large variety in outcomes used we could not further summarize the results.

Our findings suggest that the approach and the specific methods did provide different results. Previous reviews have indicated that this might be the case, but our synthesis of studies exclusively focusing on cycling according to the method used, provides more evidence of this (15, 16). This review built on earlier findings by including studies with various study designs and published in health-related and transportation-related journals. Furthermore, we quantitatively summarized the findings to assess whether the magnitude of the change in cycling differed across study design. In the following three sections we describe the implications of the study design, data collection methods and statistical approaches for the study findings.

Study design and implications for causal inference

An important aspect of study design is the choice of outcome. In this review we categorised outcomes broadly into those that assessed either cycling behavior and infrastructure usage. We found that studies on behavioral outcomes found smaller relative changes than studies presenting usage of the infrastructure. If researchers are interested in outcomes relevant for public health, it is recommended that outcomes are framed around the duration and frequency of cycling, as these measures can be directly linked to health impacts. Assessing the proportion of cyclists in a population or the numbers using a route may be a good alternative. If researchers are interested in understanding usage, count data may be used to measure the number of cyclists on the new infrastructure. Other reviews also found that studies measuring outcomes more closely related to the intervention (e.g., cycling) were more likely to find intervention effects than studies measuring more general outcomes (e.g., physical activity or BMI) (15, 54). Bike count data may support the findings from other evaluations on cycling behavior, but it cannot directly be translated into health gains in the population.

Another important design element is whether to include a control population when evaluating built environment changes. The changes in cycling differed for controlled and uncontrolled studies that assessed usage of the infrastructure, but not for cycling behavior. Uncontrolled studies have a stronger basis for causal inference if they can provide evidence that the observed effects do not solely reflect underlying time trends in cycling in the wider area (29, 36, 37, 45, 50). For example, Crane (29) counted the number of bikes passing two locations along the new infrastructure. They also presented city-wide cycling trends during the same time period. An increase of 3.7% of cyclist was found along the intervention road, whereas a decrease of 2.0% was seen in the city as a whole. This finding suggests that the number of cyclist increased in the area with the new infrastructure, and this increase does not solely reflect underlying time trends in cycling. To strengthen causal inference, we recommend that studies use controlled designs where possible, and present different measures of cycling and physical activity. Evaluating similar interventions across different sites could give further insights in the variation in the change in these sites if controlled designs are not possible. For example, Lanzendorf (37) evaluated improvements made to the cycling infrastructure in four German cities. Cycling frequency on average increased by 27%, which differed between cities from 3% to 38%. They also reported an average increase of cycling frequency by 31% in all big German cities. This approach illustrates that the observed changes in cycling in the intervention sites were comparable to the country-wide increase in cycling. The large range in changes in cycling in the four intervention sites also gives insight into the potential range of effects which could be expected in other cities.

The duration of time that populations are exposed to the new infrastructure is another important design element, which can be difficult to control in large infrastructural projects. In studies that assessed changes in cycling behavior we found that the changes were larger when exposure time was longer than 1 year. In studies that assessed the usage of cycling infrastructure, those with shorter exposure time reported larger changes than those with longer exposure time. We noted that some count studies did not count on rainy days (44, 53), or only collected data during peak hours (23, 29, 45, 47, 51), which may have resulted in larger changes than what could be expected if data was measured throughout by means of automatic counters (46). Most studies that found changes that were not in favor of the intervention were less than 6 months exposed (23, 25, 31, 32), suggesting that longer follow-up periods may be needed to allow behavioral changes to be detected. Including questions on infrastructure usage within ongoing surveys, or nested within cohorts, may ensure that if the construction work is delayed, there is data available with sufficient exposure time to measure the impact.

Data collection methods and implications for causal inference

Studies were categorized according to whether the focus was on usage or cycling behavior, and large differences in results were found between these two types of outcome. Studies presenting count data of infrastructure found larger changes than studies that assessed behavioral change in the population. Studies counting the number of bikes that passed tracking locations are at risk of assessing the displacement of existing riders to the new infrastructure, and seven studies specifically mentioned this phenomena (43, 46, 47, 50-53). Some studies had offset some of the so-called funneling biases by selecting strategic counting locations where most cyclist pass, or used multiple counting locations to capture cycling behavior in a wider area. Some studies complemented bike count data with intercept surveys among users of the infrastructure, and asked about their previous travel behaviors. These studies showed that the proportion of users that would not have cycled, had the infrastructural improvement not taken place, was much smaller than the increase in counts of cyclists (46, 47, 51). Bike count data is useful when aiming to describe at what times of the day, and under which weather conditions, cyclists are using the facility (46).

Another important consideration is choosing between objective or self-reported measures to collect data on cycling behavior. We found that studies using GPS and other objective measures of cycling reported smaller changes than those using self-reported measures. Using GPS and objective assessments of activity could potentially be used to distinguish cycling on and off the new infrastructure (26), and yields estimates of total physical activity levels (26, 31). However, such measures are often applied to a small sample, are limited to a small period of time, and participants who wear such devices might be quite different to the general population. Therefore the findings might be subject to some selection biases. Furthermore, it is possible that the novelty of wearing such devices might lead to changes in physical activity behaviors (31). Subjective measures of cycling behaviors, such as travel diaries and surveys, provide alternatives when interested in larger groups of people, but many of these have not been validated for cycling specifically.

It is attractive to use already available data when studying so-called “natural experiments” in which researchers lack control over the intervention. Collecting new data to match the timescale of intervention delivery is challenging. A third of the studies evaluating cycling behavior used data that were already collected for a regular monitoring or as part of other studies for the evaluation of other built environment interventions (28, 30, 35-37, 40, 41). For example, four US studies used census data to estimate changes in cycling after the introduction of new cycling facilities (35, 36, 40, 41). Other evaluations of natural experiments were planned, allowing to collect specific data to evaluate the intervention of interest in detail. This resulted in powerful analyses in which the method of data collection was tailored to the research questions, but sometimes resulted in limited time being exposed to the intervention. For example, Dill (31) assessed cycling at baseline and after 2-years of follow-up. The construction work was significantly delayed, resulting in a short time period between the opening of the facilities and the second assessment of cycling. Moreover, two of the nine projects were not completed within this period. This may have influenced study outcomes. Using existing data may be useful if researchers were not aware of the new intervention, did not obtain funding in time to design a study around the natural experiment, or if large delays in the construction are expected.

Analytical approaches and implications for causal inference

Like other reviews (55), we found that many studies did not perform statistical tests (for cycling behavior: 15% (8/52) and for usage: 67% (14/21)). Smaller changes were found for studies that did not test for statistical significance than those that performed statistical tests. We recommend that studies test for statistical significance which provides more robust evidence that the results are not due to chance, as recommended by guidance for the clear reporting of observational studies (56). This review included some studies that used more complex analytical methods, such as fixed-effect models (27), interrupted time series (47), or estimated the difference in cycling over time by using a regression analyses that included group, period, and an interaction term between group and period (29, 31, 35). Fixed-effects models allow to account for observed time-varying and unobserved time-invariant characteristics. Perhaps most prominently, individual attitudes towards physical activity may both determine living at a place with opportunities to be physically active and their physical activity behavior. Fixed-effect models allow to control for such unobserved time-invariant confounding, allowing for better causal inference. One study conducted a time series analyses by using GPS tracking information from a mobile phone application, thereby correcting for time trends prior to the intervention (47). Studies that specified an interaction term between group and period are able to control for observed differences between groups, thereby reducing the risk of bias. The usage of multiple analytical strategies, and the usage of methods that are able to correct for time trends, and measured or unmeasured confounders at the individual or neighborhood level may strengthening the basis of causal inference.

Strengths and limitations

In this review, we focused on the methodological aspects in the evaluation of infrastructural interventions to promote cycling and extracted information on the magnitude of the change in cycling. This allowed us to examine differences in change in cycling according to the methods used. This study was comprehensive by searching multiple electronic databases without date or language restrictions, and we included studies published in public health journals and transportation journals. Controlled and uncontrolled studies were considered for inclusion, and the final selection of studies had a large variety in study designs and methods. We added valuable information by calculating the relative and absolute changes in cycling behavior or usage of the infrastructure, which brought together different outcomes in a simple but interpretable way.

Some limitations also have to be noted. We included only studies that reported on measures of cycling and were unable to examine unreported data on cycling that were included in composite measures of active transportation, walking and cycling, or physical activity. The detail of the information provided in the papers differed between studies, which made it difficult to synthesise and interpret study findings. A pragmatic approach was used to calculate relative changes where possible, but for some studies other approaches may have been better. The evidence presented in the review came from studies that were all conducted in high-income countries. Moreover, only a few studies evaluated the impact on physical activity behaviors and studied equity effects. We focused on structural interventions here, but future research should explore the importance of and interactions with other interventions, such as financial incentives, cycle training, or behavioral interventions, together with the introduction and maintenance of high-quality cycling infrastructure.

Recommendations

Each study design, data collection method and analytical strategy has its advantages and disadvantages. To further strengthen causal inference from observational data, studies are needed that triangulate different methodologies to evaluate the effect of built environment interventions. Studies published in public health journals often report on changes in cycling behavior, while studies published in transportation journals report on usage of cycling infrastructure. Bringing experts from both fields together could result in study designs that better capture the range of impacts of new cycling infrastructure. We are not recommending a specific method or approach, as the research questions of interest should drive the method of data collection. When existing data are used, careful consideration needs to be given to the appropriateness of that data. The reporting of evaluations should adhere to guidelines, such as STROBE which seeks to strengthen the quality of work reported (56). We suggest, where possible, to combine count data that provides information on how many people are using new infrastructure, with behavioral outcomes of duration and frequency of cycling to ensure estimates of the population health impact. Such estimates could be used in combination with modelling or scenario building tools to estimate the current or future health impacts on outcomes that cannot be observed in studies with limited follow-up. Future studies should focus on the question who are benefiting from the intervention, and identify contexts, barriers and choice constraints to better understanding why cycling changed. This review focused on interventions that changed the cycling infrastructure, but findings and recommendations are likely applicable to any built environment intervention to promote health behaviors.

CONCLUSION

Introducing cycling facilities in cities is likely to increase the number of cyclist using the facilities, and may result in increases in cycling. Evidence on total physical activity following cycling facilities was mixed. Equity effects were rarely studied. Research questions interest should drive the method of data collection and reporting of evaluations should adhere to published guidelines. Triangulation of methods is warranted to overcome potential issues that evaluators may encounter when evaluating infrastructural interventions within the built environment, and to strengthen the basis of causal inference.

REFERENCES

1. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. BMJ. 2016;354:i3857.
2. Oja P, Titze S, Bauman A, de Geus B, Krenn P, Reger-Nash B, et al. Health benefits of cycling: a systematic review. Scandinavian Journal of Medicine & Science in Sports. 2011;21(4):496-509.
3. Fishman E, Böcker L, Helbich M. Adult Active Transport in the Netherlands: An Analysis of Its Contribution to Physical Activity Requirements. PLoS ONE. 2015;10(4):e0121871.
4. Kelly P, Kahlmeier S, Götschi T, Orsini N, Richards J, Roberts N, et al. Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship. International Journal of Behavioral Nutrition and Physical Activity. 2014;11(1):132.
5. Rasmussen MG, Grøntved A, Blond K, Overvad K, Tjønneland A, Jensen MK, et al. Associations between Recreational and Commuter Cycling, Changes in Cycling, and Type 2 Diabetes Risk: A Cohort Study of Danish Men and Women. PLOS Medicine. 2016;13(7):e1002076.
6. Blond K, Jensen MK, Rasmussen MG, Overvad K, Tjønneland A, Østergaard L, et al. Prospective Study of Bicycling and Risk of Coronary Heart Disease in Danish Men and Women. Circulation. 2016;134(18):1409-11.
7. de Hartog JJ, Boogaard H, Nijland H, Hoek G. Do the Health Benefits of Cycling Outweigh the Risks? Environmental Health Perspectives. 2010;118(8):1109-16.
8. Tainio M, de Nazelle AJ, Götschi T, Kahlmeier S, Rojas-Rueda D, Nieuwenhuijsen MJ, et al. Can air pollution negate the health benefits of cycling and walking? Preventive Medicine. 2016;87:233-6.
9. Pucher J, Buehler R. Making Cycling Irresistible: Lessons from the Netherlands, Denmark and Germany. Transport Reviews. 2008;28(4):495-528.
10. Garrard J, Rose G, Lo SK. Promoting transportation cycling for women: The role of bicycle infrastructure. Preventive Medicine. 2008;46(1):55-9.
11. Dill J. Bicycling for Transportation and Health: The Role of Infrastructure. Journal of Public Health Policy. 2009;30(1):S95-S110.
12. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health. 2012;66(12):1182-6.

13. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annual Review of Public Health. 2017;38(1):39-56.

14. Barnighausen T, Tugwell P, Rottingen JA, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series - Paper 4: uses and value. J Clin Epidemiol. 2017;89:21-29.

15. Stappers NEH, Van Kann DHH, Ettema D, De Vries NK, Kremers SPJ. The effect of infrastructural changes in the built environment on physical activity, active transportation and sedentary behavior – A systematic review. Health & Place. 2018;53:135-49.

16. Panter J, Guell C, Humphreys D, Ogilvie D. Can changing the physical environment promote walking and cycling? A systematic review of what works and how. Health & Place 2019;58:102161.

17. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine. 2009;6(7):e1000097.

18. Goodman A, Panter J, Sharp SJ, Ogilvie D. Effectiveness and equity impacts of town-wide cycling initiatives in England: A longitudinal, controlled natural experimental study. Social Science & Medicine. 2013;97:228-37.

19. Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, et al. The harvest plot: A method for synthesising evidence about the differential effects of interventions. BMC Medical Research Methodology. 2008;8(1):8.

20. Briss PA, Zaza S, Pappaioanou M, Fielding J, Wright-De Agüero L, Truman BI, et al. Developing an evidence-based guide to community preventive services—methods. American Journal of Preventive Medicine. 2000;18(1, Supplement 1):35-43.

21. Thomas H. Quality assessment tool for quantitative studies. Toronto: Effective Public Health Practice Project McMaster University. 2003.

22. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. Journal of Clinical Epidemiology. 2014;67(1):56-64.

23. Aittasalo M, Tiilikainen J, Tokola K, Suni J, Sievänen H, Vähä-Ypyä H, et al. Socio-Ecological Natural Experiment with Randomized Controlled Trial to Promote Active Commuting to Work: Process Evaluation, Behavioral Impacts, and Changes in the Use and Quality of Walking and Cycling Paths. Int J Environ Res Public Health. 2019;16(9):1661.

24. Aldred R, Croft J, Goodman A. Impacts of an active travel intervention with a cycling focus in a suburban context: One-year findings from an evaluation of London's in-progress mini-Hollands programme. Transportation Research Part a-Policy and Practice. 2019;123:147-69.

25. Brown BB, Smith KR, Tharp D, Werner CM, Tribby CP, Miller HJ, et al. A Complete Street Intervention for Walking to Transit, Nontransit Walking, and Bicycling: A Quasi-Experimental Demonstration of Increased Use. J Phys Act Health. 2016;13(11):1210-9.

26. Brown BB, Tharp D, Tribby CP, Smith KR, Miller HJ, Werner CM. Changes in bicycling over time associated with a new bike lane: Relations with kilocalories energy expenditure and body mass index. Journal of Transport & Health. 2016;3(3):357-65.

27. Burbidge SK, Goulias KG. Evaluating the Impact of Neighborhood Trail Development on Active Travel Behavior and Overall Physical Activity of Suburban Residents. Transportation Research Record. 2009(2135):78-86.

28. Chowdhury S, Costello SB. An examination of cyclists' and non-cyclists' mode choice under a new cycle network. Road & Transport Research. 2016;25(4):50-61.

29. Crane M, Rissel C, Standen C, Ellison A, Ellison R, Wen LM, et al. Longitudinal evaluation of travel and health outcomes in relation to new bicycle infrastructure, Sydney, Australia. Journal of Transport & Health. 2017;6:386-95.

30. Deegan B. Cycling infrastructure in London. Proceedings of the Institution of Civil Engineers-Engineering Sustainability. 2016;169(3):92-100.

31. Dill J, McNeil N, Broach J, Ma L. Bicycle boulevards and changes in physical activity and active transportation: Findings from a natural experiment. Prev Med. 2014;69(S):S74-S8.

32. Evenson KR, Herring AH, Huston SL. Evaluating change in physical activity with the building of a multi-use trail. Am J Prev Med. 2005;28(2 Suppl 2):177-85.

33. Goodman A, Sahlqvist S, Ogilvie D, iConnect C. New walking and cycling routes and increased physical activity: one- and 2-year findings from the UK iConnect Study. Am J Public Health. 2014;104(9):e38-46.

34. Song Y, Preston J, Ogilvie D, iConnect c. New walking and cycling infrastructure and modal shift in the UK: A quasi-experimental panel study. Transp res Part A policy pract. 2017;95:320-33.

35. Hirsch JA, Meyer KA, Peterson M, Le Z, Rodriguez DA, Gordon-Larsen P. Municipal investment in off-road trails and changes in bicycle commuting in Minneapolis, Minnesota over 10 years: a longitudinal repeated cross-sectional study. International Journal of Behavioral Nutrition & Physical Activity. 2017;14:1-9.

36. Krizek KJ, Barnes G, Thompson K. Analyzing the effect of bicycle facilities on commute mode share over time. J Urban Plann Dev. 2009;135(2):66–73.

37. Lanzendorf M, Busch-Geertsema A. The cycling boom in large German cities-Empirical evidence for successful cycling campaigns. Transport Policy. 2014;36:26-33.

38. Merom D, Bauman A, Vita P, Close G. An environmental intervention to promote walking and cycling - The impact of a newly constructed Rail Trail in Western Sydney. Prev Med. 2003;36(2):235-42.

39. Panter J, Heinen E, Mackett R, Ogilvie D. Impact of New Transport Infrastructure on Walking, Cycling, and Physical Activity. Am J Prev Med. 2016;50(2):45-53.

40. Pedroso FE, Angriman F, Bellows AL, Taylor K. Bicycle Use and Cyclist Safety Following Boston's Bicycle Infrastructure Expansion, 2009-2012. Am J Public Health. 2016;106(12):2171-7.

41. Smith A, Zucker S, Lladó-Farrulla M, Friedman J, Guidry C, McGrew P, et al. Bicycle Lanes: Are We Running in Circles or Cycling in the Right Direction? J Trauma Acute Care Surg. 2019;87(1):76–81.

42. Wilmink A, Hartman JB. Evaluation of the Delft Bicycle Network Plan. Final Summary Report. the Netherlands: Ministry of Transport and Public Works, 1987.

43. Barnes E, Schlossberg M. Improving Cyclist and Pedestrian Environment While Maintaining Vehicle Throughput Before- and After-Construction Analysis. Transportation Research Record. 2013(2393):85-94.

44. Fitzhugh EC, Bassett DR, Jr., Evans MF. Urban trails and physical activity: a natural experiment. Am J Prev Med. 2010;39(3):259-62.

45. Goodno M, McNeil N, Parks J, Dock S. Evaluation of Innovative Bicycle Facilities in Washington, DC Pennsylvania Avenue Median Lanes and 15th Street Cycle Track. Transportation Research Record. 2013(2387):139-48.

46. Hans SP, Bredahl JJ, Elizabeth VS, Nielsen TAS, Simon R. Effects of upgrading to cycle highways - An analysis of demand induction, use patterns and satisfaction before and after. Journal of Transport Geography. 2017;64:203-10.

47. Heesch KC, James B, Washington TL, Zuniga K, Burke M. Evaluation of the Veloway 1: A natural experiment of new bicycle infrastructure in Brisbane, Australia. Journal of Transport & Health. 2016;3(3):366-76.

48. Law S, Sakr FL, Martinez M. Measuring the Changes in Aggregate Cycling Patterns between 2003 and 2012 from a Space Syntax Perspective. Behav Sci (Basel). 2014;4(3):278-300.

49. Marqués R, Hernández-Herrador V. On the effect of networks of cycle-tracks on the risk of cycling. The case of Seville. Accid Anal Prev. 2017;102:181-90.

50. McCartney G, Whyte B, Livingston M, Crawford F. Building a bridge, transport infrastructure and population characteristics: Explaining active travel into Glasgow. Transport Policy. 2012;21:119-25.

51. Nguyen PN, Koh PP, Wong YD. Impacts of bicycle infrastructure: a case study in Singapore. Proceedings of the Institution of Civil Engineers-Municipal Engineer. 2015;168(3):186-98.

52. Parker KM, Gustat J, Rice JC. Installation of bicycle lanes and increased ridership in an urban, mixed-income setting in New Orleans, Louisiana. J Phys Act Health. 2011;8 Suppl 1:S98-S102.

53. Parker KM, Rice J, Gustat J, Ruley J, Spriggs A, Johnson C. Effect of Bike Lane Infrastructure Improvements on Ridership in One New Orleans Neighborhood. Ann Behav Med. 2013;45:S101-S7.

54. Mayne SL, Auchincloss AH, Michael YL. Impact of policy and built environment changes on obesity-related outcomes: a systematic review of naturally occurring experiments. Obesity Reviews. 2015;16(5):362-75.

55. Yang L, Sahlqvist S, McMinn A, Griffin SJ, Ogilvie D. Interventions to promote cycling: systematic review. BMJ. 2010;341.

56. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. International Journal of Surgery. 2014;12(12):1495-9.

SUPPLEMENTAL MATERIAL

Appendix 1: Calculations of relative and absolute change

Example 1: Aittasalo – Uncontrolled study

Outcome: Bikes per day during afternoon peak hour
Results - absolute change presented in text:

Before: 646 cyclists
After: 1013 cyclists

Results - absolute change: 1013 – 646 = 367 cyclists/peak hour
Results - relative change: 1013 / 646 = 1.57 (57%)

Example 2: Brown – Controlled study

Outcome: Made a bike trip on the intervention road (yes-no)
Results - absolute change presented in text:

Among Near residents (intervention group):
- Before: 8% were detected cycling on the intervention road
- After: 10% were detected cycling on the intervention road

Among Far residents (control group):
- Before: 5% were detected cycling on the intervention road
- After: 7% were detected cycling on the intervention road

Results - absolute change: (10 – 8) – (7 – 5) = 0
Results - relative change: (10 / 8) / (7 / 5) = 0.89 (-11%)

Supplemental Table 1: Summary of the results

	Cycling behavior			Usage of infrastructure		
	Outcomes reported	In favor of the intervention	Median relative change[a]	Outcomes reported	In favor of the intervention	Median relative change[a]
	n	n	%	n	n	%
All outcomes	52	38	NA	21	21	NA
All outcomes, quantitative summary	36	29	23	20	20	62
Study design						
Controlled	18	13	19	4	4	143
Uncontrolled	18	16	30	16	16	60
Exposure time						
≥1 year	17	16	40	12	12	55
<1 year	19	13	15	8	8	93
Data collection methods						
Objective measures[b]	7	3	-1	8[d]	8	39
Subjective measures[c]	29	26	27	8	8	73
Statistical analyses						
Tested	26	22	12	6	6	163
Not tested	10	7	7	14	14	60

[a]No units were presented for the median relative change because it can refer to various metrics.
[b]Objective registration methods included GPS and accelerometer data for studies assessing cycling behavior, and GPS, accelerometer data and automatic counting stations for studies assessing usage of the infrastructure.
[c]Subjective registration methods included self-reported cycling by travel diaries, telephone interviews, and surveys for studies assessing cycling behavior, and direct observations for studies assessing usage of the infrastructure.
[d] For four outcomes it was unclear what type of data collection method was used, or methods changed over time.

8

A framework for exploring non-response patterns over time in health surveys

Famke J.M. Mölenberg, Chris de Vries, Alex Burdorf, Frank J. van Lenthe

BMC Medical Research Methodology 2021;21(1):37

ABSTRACT

Introduction

Most health surveys have experienced a decline in response rates. A structured approach to evaluate whether a decreasing - and potentially more selective - response over time biased estimated trends in health behaviours is lacking. We developed a framework to explore the role of differential non-response over time. This framework was applied to a repeated cross-sectional survey in which the response rate gradually declined.

Methods

We used data from a survey conducted biannually between 1995 and 2017 in the city of Rotterdam, the Netherlands. Information on the sociodemographic determinants of age, sex, and ethnicity was available for respondents and non-respondents. The main outcome measures of prevalence of sport participation and watching TV were only available for respondents. The framework consisted of four steps: 1) investigating the sociodemographic determinants of responding to the survey and the difference in response over time between sociodemographic groups; 2) estimating variation in health behaviour over time; 3) comparing weighted and unweighted prevalence estimates of health behaviour over time; and 4) comparing associations between sociodemographic determinants and health behaviour over time.

Results

The overall response rate per survey declined from 47% in 1995 to 15% in 2017. The probability of responding was higher among older people, females, and those with a Western background. The response rate declined in all subgroups, and a faster decline was observed among younger persons and those with a non-Western ethnicity as compared to older persons and those with a Western ethnicity. Variation in health behaviours remained constant. Prevalence estimates and associations did not follow the changes in response over time. On the contrary, the difference in probability of participating in sport gradually decreased between males and females, while no differential change in the response rate was observed.

Conclusion

Providing insights on non-response patterns over time is essential to understand whether declines in response rates may have influenced estimated trends in health behaviours. The framework outlined in this study can be used for this purpose. In our example, in spite of a major decline in response rate, there was no evidence that the risk of non-response bias increased over time.

INTRODUCTION

Health surveys provide valuable information on the distribution of disease in a population, health behaviours associated with diseases, and trends over time. Most surveys have experienced major decreases in response rates over the past decades (1-3). This increases the possibility of selective participation that may bias estimates of prevalence of health behaviours and also associations between determinants and health behaviours. Especially for studying trends, it is important to disentangle an actual change in disease and underlying health behaviours in the entire population from an observed change due to systematic changes in the composition of respondents of the survey over time.

A low response rate in repeated cross-sectional surveys may influence the prevalence estimates of health behaviours in two ways. First, selective participation may occur by sociodemographic subgroups (4, 5). If a response rate becomes lower among younger compared to older individuals, and younger individuals participate more frequently in sport activities, the observed prevalence estimate of sport activities in the study sample may underestimate sport activities in the entire population. Selective participation between sociodemographic subgroups is often taken into account by calculating weighting factors such that the composition of the study sample mirrors the entire population for known sociodemographic characteristics (6, 7). Second, selective participation may take place by the outcome of interest, either a health behaviour or a disease (4, 8). If a response rate is higher among individuals who participate more frequently in sport activities, the observed prevalence estimate of sport activities in the study sample may overestimate sport activities in the entire population. Correcting for the selective participation for outcomes is often not possible given that, in most surveys, the outcome among those who do not respond is unknown. To make it more complex, the combination of both mechanisms of selective participation may also occur, whereby unhealthy younger persons respond less than healthy older persons, resulting in a gross underestimation of the prevalence of disease and overestimation of health behaviour.

A low response rate also plays a role in estimating the associations between sociodemographic determinants and outcomes, albeit in a different way. Selective participation by sociodemographic subgroups will not influence associations between age and sport participation as long as the persons who respond are a random sample. If younger individuals respond less, but those who do respond participate in sport as frequently as individuals who do not respond, the association between age and sport participation is not influenced by the low response rate. However, if selective participation takes place by the outcome of interest, and only the young individuals who participate in sport respond, the observed estimate for the association between age and sport participation may be overestimated. In addition, at this point a combination of the two selection processes may occur; however, only selective participation by the outcome will bias associations. Correcting for the latter is not possible, although it has been suggested that a low response rate will not bias associations in studies with large samples sizes and sufficient variation in the outcome of interest (9, 10). This suggestion was supported by a modelling study showing that estimates of associations between determinants and outcomes were relatively unbiased in scenarios where variation was maintained by including some individuals with outcomes at extreme values of the distribution (11).

Bias in estimated trends in health behaviours due to a decreasing - and potentially more selective - number of responses over time can be investigated by comparing the sociodemographic characteristics and outcomes of those who respond to those who do not. Some studies will have complete information of the total sample, for example using information collected in previous studies (4), through linkage with registries on sociodemographic characteristics (5) or databases for specific outcomes such as membership of a sports club (8). Nevertheless, the majority of outcomes will only be known for those who participate in the survey. Although most health surveys are suffering from a decline in response rates, a step-by-step approach to describe how a decline in responses over time can be explored in repeated cross-sectional surveys is currently lacking.

This study aimed to outline a framework combining multiple analytical strategies to explore the risk of non-response bias over time and to evaluate its usefulness in a population survey where the response rate has considerably decreased over time.

METHODS

Study population

We used the Rotterdam leisure-time survey *(in Dutch: Vrijetijdsonderzoek)* to obtain data on health behaviours of Rotterdam citizens, the 2nd largest city in the Netherlands with ~650,000 inhabitants. This biannual cross-sectional survey was conducted by researchers from the City of Rotterdam from 1995 until 2017. Participants were randomly selected from the municipal registration database and sent a letter of invitation and a questionnaire. The registration database provided information on age, sex, ethnicity, and district of residence, and was linked with questionnaire information on demographics, leisure-time related topics, and a section on sport participation and sedentary behaviours. Respondents were informed that completing the questionnaire implied consent to use the data anonymously for research purposes. The approach to inviting participants and the mode of delivery of data collection changed over time. An overview of the changes is presented in Supplemental Table 1. From 1995 until 2003, a random selection of individuals was invited to participate. Since 2005, specific population subgroups were oversampled to increase the precision of estimates of subgroups with few respondents.

The total study sample for analyses included 33,934 respondents aged 15–75 years for which complete data on age, sex, ethnicity, income, and city district were available. The total study sample for analyses included 25,897 respondents with data on sport participation and 29,615 on watching TV.

Sport participation and watching TV

The main outcomes of the study were sport participation and watching TV, since these were repeatedly measured with similar questions in at least ten subsequent biannual surveys. A standardised questionnaire *(in Dutch: Richtlijn Sportdeelname Onderzoek)* was used to ask about sport participation frequency over the past year with the open question: “How many times did you participate in sport over the past 12 months?”, and answers were categorised into weekly (≥46 times/year) or less than weekly (<46 times/year). Responses on sport participation were available from 1999 to 2017 as the question was asked neither in 1995 nor 1997. Time spent watching TV was asked by the open question: “How many hours, on an average day, do you spend watching TV?”, and responses were categorised into <3 hours/day or ≥3 hours/

day. The cut-off value was based on the results of an earlier study suggesting that watching TV for ≥3 hours/day was associated with increased mortality, independent of physical activity (12). Responses on watching TV were available from 1995 to 2015, but not 2017 as the question was not asked in that year.

Sociodemographic variables

Information on age, sex, ethnicity, and city district was derived from the municipal registration database. Age was categorised into 16-24, 25-44, 45-64, and 65-75 years. Ethnicity was based on the country of birth and classified as Western and non-Western, following the definition by Statistics Netherlands (13). Income was obtained from the survey. For each survey, the social minimum income and modal income of that specific year was used as the reference to classify self-reported household income into low (below social minimum), mid-low (social minimum to modal), mid-high (modal to 2x modal), and high (>2x modal) income.

Statistical analyses

The framework consists of four steps that enable the estimation of whether non-response over time has biased observed trends in prevalence estimates of sport participation and TV watching in the general population and their associations with sociodemographic characteristics. In the first step, we investigated the sociodemographic determinants of responding to each biannual survey, using data from all persons invited to participate in each survey. Logistic regression analysis was conducted with response as the binary outcome variable, and time (survey year), age (four categories), sex (male vs female) and ethnicity (Western vs non-Western) as determinants, using pooled data of all surveys. The odds ratio (OR) estimates the trend in response per two-year interval as the survey was repeated biannually. We tested whether the response over time differed between age groups, sex, and ethnic groups by adding interaction terms with time to the model. Significant interaction terms indicate that the probability of responding for particular age groups, sex, or ethnic groups changed over time.

For the second step, we estimated the variation in outcomes by calculating the mean and standard deviation (SD) in sport participation (times/year) and watching TV (hours/day) in each cross-sectional sample by sociodemographic subgroup. We visualised the patterns in SD over time by age group, sex, and ethnic group.

For the third step, we calculated weighted and unweighted prevalence estimates of sport participation and watching TV in each cross-sectional sample by sociodemographic subgroup. Prevalence estimates were weighted by age, sex, and ethnicity distributions per city district to reflect the Rotterdam population of that year. All weighting variables were obtained from the municipal registration database, and thus available for respondents and non-respondents. The difference between weighted and unweighted prevalence estimates illustrates how the differential response rates between population subgroups affected the prevalence estimates of sport participation and watching TV.

For the fourth step, we estimated associations between sociodemographic determinants and health behaviours. For each cross-sectional sample, two separate logistic regression analyses were conducted with sport participation and watching TV as the binary dependent variable, respectively. Independent variables were age (four age categories), sex (male vs female), and ethnicity (Western vs non-Western). Studies have shown that higher socioeconomic subgroups more often participate in sport activities and spend less time watching TV (14-16), therefore models were additionally adjusted for household income (four income categories). The OR estimates per wave were analysed for changes over time, indicating whether particular age groups, sex, or ethnic groups reported increasing or decreasing sport participation or TV watching over time.

We presented a checklist to summarise whether differential non-response over time may have biased estimated trends in health behaviours. All analyses were conducted in SPSS version 24. Two-sided P-values of <0.05 were considered statistically significant. Interactions were explored for P-for-interactions of <0.10.

RESULTS

Step 1: Sociodemographic determinants of responding to the survey

During 12 waves, a total of 136,696 subjects were invited to participate, and 35,587 subjects returned the questionnaire. The overall response rate declined from 47% in 1995 to 15% in 2017. The characteristics of the study sample are presented in Supplemental Table 2. Over the whole period, the probability of responding was higher among older people, females, and those with a Western background (Table 1).

The average decline in response was 8% every two years. Over a 22-year period, this accumulated to a relative decrease of 62% in survey responses. In the first survey conducted in 1995, the overall response rate was 47%, and the absolute difference between the most and least responsive subgroup was 9%-points (Supplemental Table 3). For the most recent survey conducted in 2017, the overall response rate was 15%, and the absolute difference between the most and least responsive subgroup was 15%-points (Supplemental Table 3).

Between 1995 and 2017, the response rate declined for all groups. The changes in response over time differed for age groups and ethnic groups (P-for-interaction<0.10) but not for sex. The probability of responding decreased by 10% every two years for individuals aged 16-24 years, as compared to 4% for individuals aged 65-75 years (Table 2). Over a 22-year period, this accumulated to a 70% decrease in response for individuals aged 16-24 years and a 36% decrease in response for individuals aged 65-75 years. The probability of responding decreased by 7% per 2 years (57% over 22 years) for individuals with a Western background as compared to 10% (69% over 22 years) for individuals with a non-Western background.

Table 1: Sociodemographic determinants of response to a health survey

	Response OR (95% CI)
Time (per 2 years)	0.92 (0.92; 0.92)
Age (years)	
16-24	1.01 (0.97; 1.05)
25-44 (ref)	1.0
45-64	1.29 (1.25; 1.33)
65-75	1.54 (1.48; 1.61)
Sex	
Male (ref)	1.0
Female	1.29 (1.26; 1.32)
Ethnicity	
Western (ref)	1.0
Non-Western	0.51 (0.49; 0.52)

Results were obtained from a logistic regression model using response to the health survey as a binary outcome, including time as a continuous variable (per 2 years), and sociodemographic determinants as listed. All determinants were derived from the municipal registration database, and thus available for respondents and non-respondents.

Step 2: Variation in outcomes over time

Between 1995 and 2017, the variation in sport participation remained fairly stable across sociodemographic groups (Figure 1). In the total population, the variation changed from a SD of 76 times/year in 1999 to 70 times/year in 2017. The variation in TV watching increased in the last three rounds of data collection, from a SD of 1.9 hrs/day in 2011 to 3.1 hrs/day in 2015. However, there were no large differences between sociodemographic subgroups. No differences in variance were seen in the period that less-responsive subgroups were oversampled as compared to the situation before in which a random selection was invited to participate.

Table 2: Sociodemographic determinants of the change in response to a health survey between 1995 and 2017

	Change in response per 2 years OR (95% CI)
Age (years)	
16-24	0.90 (0.90; 0.91)
25-44	0.91 (0.91; 0.91)
45-64	0.93 (0.93; 0.94)
65-75	0.96 (0.95; 0.97)
Sex	
Male	0.92 (0.92; 0.92)
Female	0.92 (0.92; 0.92)
Ethnicity	
Western	0.93 (0.93; 0.93)
Non-Western	0.90 (0.90; 0.91)

Results were obtained from a stratified logistic regression model using response to the health survey as a binary outcome, including time as a continuous variable (per 2 years), and sociodemographic determinants as listed. All determinants were derived from the municipal registration database, and thus available for respondents and non-respondents.

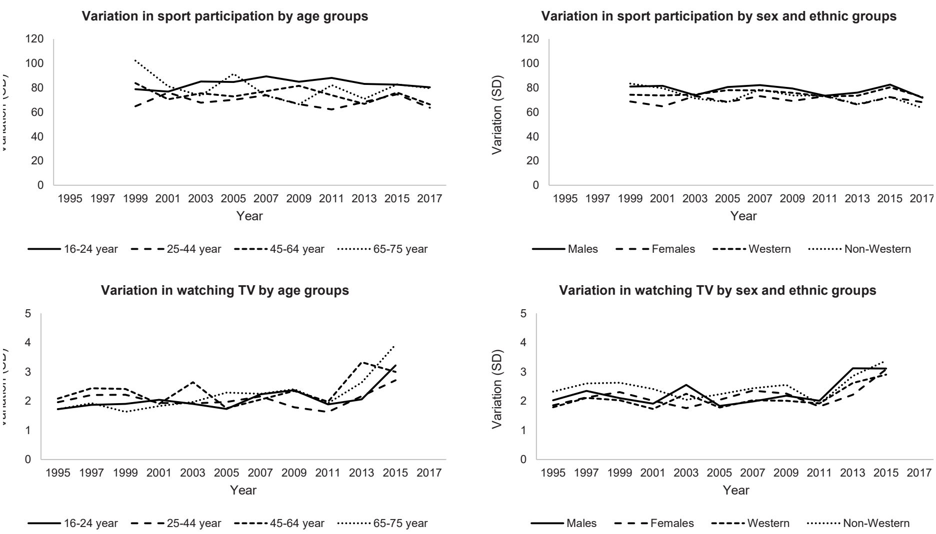


Figure 1: Variation (SD) in sport participation and watching TV by sociodemographic determinants in each cross-sectional sample

Abbreviation: standard deviation (SD).

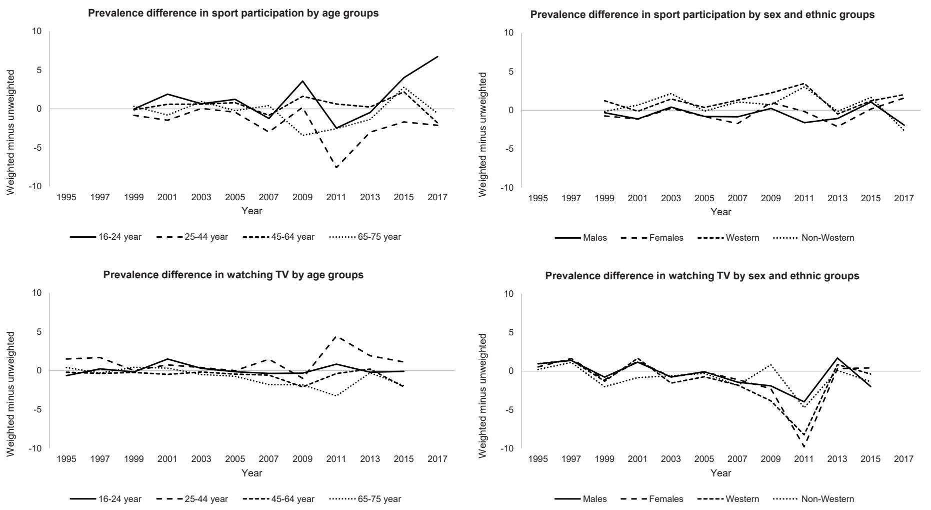


Figure 2: Weighted minus unweighted prevalence estimates (%-points) in sport participation and watching TV by sociodemographic determinants in each cross-sectional sample

Prevalence estimates were weighted for age, sex, and ethnicity distributions per city district to reflect the population of that year.

Step 3: Comparison of weighted and unweighted prevalence estimates

The difference between weighted and unweighted prevalence estimates for sport participation was small for most comparisons in the cross-sectional samples [mean difference: 0.0%-points (interquartile range (IQR): -1.0–1.1)]. Likewise, weighted and unweighted prevalence estimates for watching TV did not differ largely [mean difference: -0.5%-points (IQR: -1.1–0.5)]. There was no clear trend across sociodemographic subgroups (Figure 2).

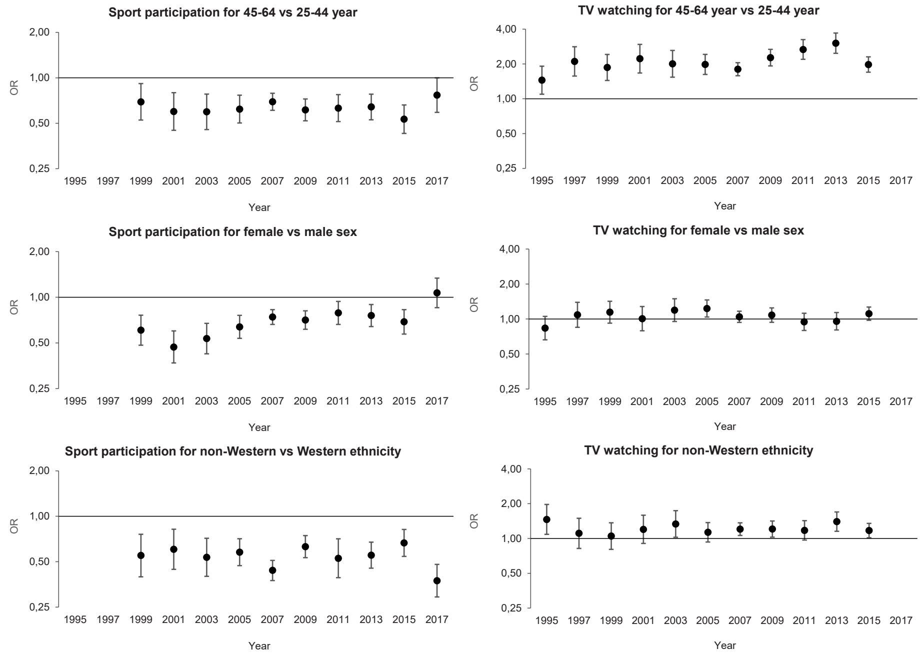


Figure 3: Associations between health behaviours and sociodemographic determinants in each cross-sectional sample

Results were obtained from a logistic regression model using sport participation (left panel) and watching TV (right panel) as a binary outcome, and adjusted for age, sex, ethnicity and household income. Abbreviation: odds ratio (OR).

Step 4: Comparisons of associations between health behaviours and sociodemographic groups

In all years, the probability of participating in sport was lower for individuals aged 45-64 years and with a non-Western background as compared to those aged 25-44 years and with a Western background (Figure 3). The probability of participating in sport was lower for females in earlier years, but over time the difference in sport participation between males and females gradually decreased. The probability of watching TV was higher for individuals aged 45-64 years and those with a non-Western background as compared to those aged 25-44 years and with a Western background. No differences were seen between males and females.

Table 3: Checklist for the evaluation of non-response over time in repeated cross-sectional studies

Health behaviour as outcome		Sport participation			Watching TV		
Sociodemographic subgroup		Age	Sex	Ethnicity	Age	Sex	Ethnicity
Step 1	Did the change in response over time differ between subgroups?	Y	N	Y	Y	N	Y
Step 2	Did the variation in outcome change over time?	N	N	N	N	N	N
Step 3	Did the difference between weighted and unweighted prevalence estimates of the outcome change over time?	N	N	N	N	N	N
Step 4	Did the association between sociodemographic determinants and outcome change over time?	N	Y	N	N	N	N
Summary	Is it likely that increasing non-response over time biased estimated time trends in outcomes?	N	N	N	N	N	N

Summary of the findings

A summary of the findings is presented in Table 3. In spite of a faster decline in response among persons aged 25-44 years and those with a non-Western ethnicity as compared to persons aged 45-64 years and those with a Western ethnicity, the variation in sport participation and watching TV in the study sample remained constant over time. Weighted and unweighted prevalence estimates and associations did not follow the changes in response over time. On the contrary, we observed that the difference in probability of participating in sport gradually decreased between males and females, while the change in response and variation in sport participation in the study sample did not differ over time between males and females. We did not find evidence that the risk of non-response bias increased over time.

DISCUSSION

We outlined a four-step approach to obtain insights in differential non-response patterns over time, and the influence on estimations of prevalence in health behaviours and associations between sociodemographic characteristics and health behaviours. We applied this framework to the Rotterdam leisure-time survey, a repeated cross-sectional study in which the overall response rate declined from 47% in 1995 to 15% in 2017. We observed that the response rate declined more rapidly among younger participants and those with a non-Western ethnicity. The results did not suggest that the differential non-response biased estimated trends in prevalence of sport participation and watching TV and associations between sociodemographic determinants and sport participation and watching TV.

Health surveys are frequently used by national and local institutions to gain insights on the health of the population at a specific point in time and to evaluate trends over time. It is important to thoroughly examine potential non-response bias in questionnaire surveys. We suggest to report the change in response for each subgroup to provide insights on the risk of bias in some subgroups as compared to others. Whereas most health surveys experienced major decreases in response rates over time, only few have described the presence of differential non-response and its impact on reported findings.

Two studies have previously evaluated the decline in response across subgroups. Responses to a Finnish health survey between 1978 and 2002 declined faster among younger participants and males, and authors suggested that this may have resulted in biased trend estimates (5). In a Lithuanian study, the decrease in response between 1994 and 2010 was similar in all age groups and between males and females, and authors concluded that this could not seriously bias trends in smoking behaviour (17). One study evaluated trends in smoking behaviour in 17 populations, using information from persons who immediately returned the questionnaire, and compared this with information collected for part of the population who initially did not respond (18). They showed that the decreasing trend in smoking prevalence among males was overestimated in most populations when calculating trends only based on data from respondents who immediately handed in the survey. For females, however, the decreasing trend was overestimated in half of the population and underestimated in the other half. Some studies evaluating trends in health behaviours have stated that their findings may be biased due to a larger decline in response among some groups compared to others (5, 8, 19-23). Most studies, however, did not give further detail on the differences in non-response over time (8, 19-23), therefore judging whether this poses a serious risk is not possible. The framework proposed in this study may be useful for this purpose.

The percentage of people who respond to a survey is often used as the first indicator of potential non-response bias. Tools for the evaluation of methodological quality often classify studies that have a response below 60% as studies with a high risk of bias. However, there is no scientific evidence that supports this cut-off (24). This classification can be too simplistic, especially when sampling strategies are used whereby people from less-responsive subgroups are more frequently invited. Inviting more people from less-responsive subgroups will directly result in a lower overall response, however, this does not necessarily introduce bias. On the contrary, a modelling study showed that by including at least some persons in upper and lower percentiles of the distributions of outcomes reduced the risk of bias in associations between determinants and outcomes (11). Recruiting persons from the extremes of the distribution likely increases the variation in the study sample. Following this rationale, we would expect a larger variation in the outcomes from 2005 onwards, but the variation remained fairly constant. However, we do not know what the variation would have been in the years after 2005 without the oversampling strategy. It is likely that the sampling methodology contributed to the relatively high variation in the study sample throughout the study period. When using a sampling strategy, it remains important to invite persons within oversampled subgroups randomly and to capture information that allows for the correct weighting of the study sample back to the population that they represent. Oversampling may come at a cost of small losses in precision for other subgroups and the total population when the total number of questionnaires is fixed (25). Recruiting and retaining individuals from less-responsive subgroups at the end of the distributions of relevant outcomes is warranted to obtain precise prevalence estimates and lower the risk of bias for associations between determinants and outcomes.

Strengths and limitations

A strength of this study was the long time series of repeated surveys, which enabled the application of the developed framework to an existing survey that experienced large reductions in response over a 22-year period. The possibility to link both respondents and non-respondents to the municipal registration database provided the unique opportunity to analyse changes in the characteristics of non-respondents over time. An identical set of questions was used to assess sport participation and watching TV in all years. Datasets were harmonised and a weighting variable per survey year was computed using similar methodology. The framework presented in this paper can be used retrospectively to evaluate whether differential non-response has biased trends for various types of outcomes. No additional data needs to be collected.

Our study also has some limitations. The framework depends on the availability of sociodemographic characteristics for respondents and non-respondents and outcomes obtained from the survey. Differential non-response may have occurred for sociodemographic characteristics that remain unknown. Recent developments in linkage with registries, including taxation office and educational achievements, may allow to further detail non-response between sociodemographic subgroups and by outcomes in future studies. Another limitation is that the framework can only be used to evaluate whether the risk of non-response bias has changed over time. Especially for studies with low response rates in early years, the framework does not enable the evaluation of whether estimates in each cross-sectional sample are unbiased. The framework was applied to a population survey in which several aspects around the data collection have changed, possibly influencing the number of responses. In 2011, the respondents could choose to complete the survey digitally or take the pen and paper version. In the same year, door-to-door interviews to recruit non-Dutch participants were discontinued, and the survey was divided into two shorter surveys. We were unable to attribute these multiple changes at the same time to differences in response and whether response rates in some subgroups was more affected than in others.

Future recommendations

It is important to thoroughly examine potential non-response bias when studying trends in health behaviours. Therefore, we suggest to report response by sociodemographic subgroups. The framework outlined in this paper starts with presenting the differential change in non-response, and from there insights are obtained in bias in estimated trends of health behaviours over time. Recruiting persons from less-responsive subgroups is needed to ensure substantial variation in outcomes in the study sample. The framework can be applied to surveys that collect data on at least two demographic subgroups and two outcome variables at multiple time-points. This allows for the exploration of whether trends in the two outcomes follow the patterns in non-response across these subgroups. A larger number of demographic subgroups and outcome variables will strengthen the conclusion on likelihood of bias in the trends in outcomes due to differential non-response.

CONCLUSION

Surveys are an important source of information about the health of a population. A decreasing - and potentially more selective - response over time could increase the risk of bias in studies investigating trends. We provided a framework to assess whether non-response patterns over time resulted in biased trends in health behaviours. The framework was applied to a survey in which response rates gradually decreased over time. In spite of a major decline in overall response rates, and a faster decline in response in some subgroups than others, we did not find evidence that the risk of bias due to non-response has increased over time. Our results suggests that ensuring substantial variation in outcomes rather than ensuring a high response rate is important to reduce the risk of bias.

REFERENCES

1. De Heer W, De Leeuw E. Trends in household survey nonresponse: A longitudinal and international comparison. Survey nonresponse. 2002;41:41-54.

2. Galea S, Tracy M. Participation Rates in Epidemiologic Studies. Annals of Epidemiology. 2007;17(9):643-53.

3. Mindell JS, Giampaoli S, Goesswald A, Kamtsiuris P, Mann C, Männistö S, et al. Sample selection, recruitment and participation rates in health examination surveys in Europe--experience from seven national surveys. BMC Med Res Methodol. 2015;15:78.

4. Van Loon AJM, Tijhuis M, Picavet HSJ, Surtees PG, Ormel J. Survey Non-response in the Netherlands: Effects on Prevalence Estimates and Associations. Annals of Epidemiology. 2003;13(2):105-10.

5. Tolonen H, Helakorpi S, Talala K, Helasoja V, Martelin T, Prättälä R. 25-year trends and socio-demographic differences in response rates: Finnish adult health behaviour survey. Eur J Epidemiol. 2006;21(6):409-15.

6. Lynn P. Weighting for non-response. Survey and statistical computing. 1996:205-14.

7. Little RJ, Vartivarian S. On weighting the rates in non-response weights. Statistics in Medicine. 2003;22(9):1589-99.

8. Harvey JT, Charity MJ, Sawyer NA, Eime RM. Non-response bias in estimates of prevalence of club-based sport participation from an Australian national physical activity, recreation and sport survey. BMC Public Health. 2018;18(1):895.

9. Collins R. What makes UK Biobank special? The Lancet. 2012;379(9822):1173-4.

10. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. BMJ. 2020;368:m131.

11. Gustavson K, Røysamb E, Borren I. Preventing bias from selective non-response in population-based survey studies: findings from a Monte Carlo simulation study. BMC Med Res Methodol. 2019;19(1):120.

12. Ekelund U, Steene-Johannessen J, Brown WJ, Fagerland MW, Owen N, Powell KE, et al. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. Lancet. 2016;388(10051):1302-10.

13. Statistics Netherlands, 2010. *Annual Report on Integration 2010 (in Dutch: Jaarrapport Integratie 2010)*. Den Haag/Heerlen, the Netherlands

14. Gidlow C, Johnston LH, Crone D, Ellis N, James D. A systematic review of the relationship between socio-economic position and physical activity. Health Education Journal. 2006;65(4):338-67.

15. Beenackers MA, Kamphuis CBM, Giskes K, Brug J, Kunst AE, Burdorf A, et al. Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: a systematic review. Int J Behav Nutr Phys Act. 2012;9:116.

16. Rhodes RE, Mark RS, Temmel CP. Adult Sedentary Behavior: A Systematic Review. American Journal of Preventive Medicine. 2012;42(3):e3-e28.

17. Klumbiene J, Sakyte E, Petkeviciene J, Prattala R, Kunst AE. The effect of tobacco control policy on smoking cessation in relation to gender, age and education in Lithuania, 1994–2010. BMC Public Health. 2015;15(1):181.

18. Tolonen H, Dobson A, Kulathinal S, Project WM. Effect on trend estimates of the difference between survey respondents and non-respondents: results from 27 populations in the WHO MONICA Project. Eur J Epidemiol. 2005;20(11):887-98.

19. Hotchkiss JW, Davies C, Gray L, Bromley C, Capewell S, Leyland AH. Trends in adult cardiovascular disease risk factors and their socio-economic patterning in the Scottish population 1995-2008: cross-sectional surveys. BMJ Open. 2011;1(1):e000176.

20. Ernstsen L, Strand BH, Nilsen SM, Espnes GA, Krokstad S. Trends in absolute and relative educational inequalities in four modifiable ischaemic heart disease risk factors: repeated cross-sectional surveys from the Nord-Trøndelag Health Study (HUNT) 1984-2008. BMC Public Health. 2012;12:266.

21. Vikum E, Bjørngaard JH, Westin S, Krokstad S. Socio-economic inequalities in Norwegian health care utilization over 3 decades: the HUNT Study. European Journal of Public Health. 2013;23(6):1003-10.

22. Abouzeid M, Wikström K, Peltonen M, Lindström J, Borodulin K, Rahkonen O, et al. Secular trends and educational differences in the incidence of type 2 diabetes in Finland, 1972–2007. European Journal of Epidemiology. 2015;30(8):649-59.

23. Graff-Iversen S, Ariansen I, Naess O, Selmer RM, Strand BH. Educational inequalities in midlife risk factors for non-communicable diseases in two Norwegian counties 1974-2002. Scand J Public Health. 2019;47(7):705-12.

24. Johnson TP, Wislar JS. Response Rates and Nonresponse Errors in Surveys. JAMA. 2012;307(17):1805-6.

25. Mohadjer L, West J. Effectiveness of Oversampling Blacks and Hispanics in the NHES Field Test. National Household Education Survey Technical Report. 1992.

SUPPLEMENTAL MATERIAL

Appendix 1	Questionnaire items
Supplemental Table 1	Description of the approach to invite participants and the data collection procedures from 1995 to 2017
Supplemental Table 2	Population characteristics from 1995 to 2017
Supplemental Table 3	Response rate in the total population and by age group, sex and ethnic group

Appendix 1: Questionnaire items

The following questions were used to collect information on the main outcomes sport participation and watching TV. These outcomes were repeatedly measured with similar questions in at least 10 subsequent biannual surveys. The translated English language version and the original Dutch language version are given.

Sport participation:

How many times did you participate in sport over the past 12 months *<u>in total</u>*?
If you frequently participated in sports, please make an estimation of the number of times.

[ ] times

Hoeveel keer heeft u in de afgelopen twaalf maanden *<u>in totaal</u>* gesport?
Als u veel aan sport heeft gedaan, maak dan een schatting van het aantal keer.

[ ] keer

Watching TV:

How many hours, on an average day, do you spend <u>watching TV</u>?

[ ] hours per day

Hoeveel uur kijkt u gemiddeld per dag <u>televisie</u>?

[ ] uur per dag

Supplemental Table 1: Description of the approach to invite participants and the data collection procedures from 1995 to 2017

	1995	1997	1999	2001	2003	2005	2007	2009	2011	2013	2015	2017
Pen and paper survey	X	X	X	X	X	X	X	X				
Digital survey								X	X	X	X	X
Additional door-to-door interviews to recruit non-Dutch participants	X	X	X	X	X	X	X	X				
No additional door-to-door interviews									X	X	X	X
Random selection of participants	X	X	X	X	X							
Stratified by city district and oversampling of less-responsive subgroups						X	X	X	X	X	X	X
Oversampling of elderly aged 55-75 years							X					
Single survey with all items	X	X	X	X	X	X	X	X		X		X
Two surveys with half of the items									X		X	

8

Supplemental Table 2: Population characteristics from 1995 to 2017

	1995 (n=1404)	1997 (n=1285)	1999 (n=1610)	2001 (n=1270)	2003 (n=1473)	2005 (n=2603)	2007 (n=5966)	2009 (n=4091)	2011 (n=5207)	2013 (n=2743)	2015 (n=4705)	2017 (n=1577)
Age, years (%)												
16-24	16,7	15,7	15,7	15,9	16,6	16,6	16,6	16,7	16,5	16,5	16,4	16,3
25-44	43,8	44,2	44,4	44,2	43,3	42,9	41,7	41,0	40,7	40,3	39,9	39,9
45-64	27,5	28,1	28,5	28,8	29,4	30,0	31,3	32,0	32,7	32,2	32,2	32,3
65-75	12,0	12,0	11,4	11,0	10,6	10,5	10,5	10,3	10,1	11,0	11,5	11,5
Sex (%)												
Male	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0
Female	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0	50,0
Ethnicity (%)												
Western	77,3	75,7	73,5	70,6	68,8	67,1	66,0	64,7	63,7	62,7	61,8	61,0
Non-Western	22,7	24,3	26,5	29,4	31,2	32,9	34,0	35,3	36,3	37,3	38,2	39,0
Household income (%)												
Low	23,8	20,5	19,9	21,8	22,4	23,6	22,4	23,6	25,8	34,5	31,5	26,3
Mid-low	24,6	27,0	25,7	27,8	29,3	28,3	28,1	22,5	21,0	20,5	21,5	24,4
Mid-high	36,1	37,3	36,9	34,0	27,2	28,6	28,0	31,3	28,7	26,4	25,1	25,7
High	15,5	15,3	17,5	16,4	21,1	19,5	21,5	22,5	24,4	18,6	22,0	23,6

Supplemental Table 2: (Continued) Population characteristics from 1995 to 2017

	1995 (n=1404)	1997 (n=1285)	1999 (n=1610)	2001 (n=1270)	2003 (n=1473)	2005 (n=2603)	2007 (n=5966)	2009 (n=4091)	2011 (n=5207)	2013 (n=2743)	2015 (n=4705)	2017 (n=1577)
Sport participation (%)												
Less than weekly	NA[a]	NA[a]	66,0	61,4	61,3	62,9	62,2	58,9	56,8	59,6	51,0	58,0
Weekly	NA[a]	NA[a]	34,0	38,6	38,7	37,1	37,8	41,1	43,2	40,4	49,0	42,0
Watching TV (%)												
<3 hours/ day	44,8	50,6	51,9	48,7	52,9	54,0	56,5	56,1	57,5	60,0	61,7	NA[b]
≥3 hours/ day	55,2	49,4	48,1	51,3	47,1	46,0	43,5	43,9	42,5	40,0	38,3	NA[b]

Prevalence estimates were weighted for age, sex, and ethnicity distributions per city district to reflect the population of that year.

[a] Responses on sport participation were available from 1999 to 2017.

[b] Responses on watching TV were available from 1995 to 2015.

Supplemental Table 3: Response rate in the total population and by age group, sex and ethnic group

	1995	1997	1999	2001	2003	2005	2007	2009	2011	2013	2015	2017
Overall response (%)	46,8	43,5	54,0	48,6	53,8	42,8	36,7	20,1	33,1	23,0	20,5	14,8
Age, years (%)												
16-24	44,8	47,0	52,8	46,2	52,1	38,4	29,8	18,3	25,0	19,0	17,5	12,3
25-44	48,2	43,4	53,1	47,3	50,2	40,7	33,7	17,7	26,3	20,9	17,7	12,5
45-64	48,1	42,9	53,4	51,5	57,6	47,9	41,0	23,0	36,5	26,0	24,0	17,2
65-75	41,7	40,9	59,9	49,2	59,3	53,3	44,3	24,5	46,6	34,6	30,9	25,1
Sex (%)												
Male	42,3	42,1	50,3	44,8	50,5	37,9	34,0	18,5	29,8	21,2	18,4	13,8
Female	50,5	44,9	57,5	52,2	57,3	47,6	39,4	21,7	36,2	24,8	22,6	15,8
Ethnicity (%)												
Western	47,0	45,3	57,8	52,4	56,5	49,9	40,8	24,8	36,9	30,3	27,3	20,1
Non-Western	46,3	38,1	43,3	40,1	48,0	32,7	27,6	14,0	20,2	15,1	14,5	10,3

9

The paradox of getting control over natural experiments

Famke J.M. Mölenberg, Francisca Vargas Lopes

Submitted

The opportunity of natural experiments in public health have been discussed for some years (1-5). Natural experiments allow the retrospective and prospective evaluation of policies, interventions or programs in real-world settings (1). Importantly, they present a valuable alternative to evaluate changes to a system for which it would be unethical, unfeasible or simply impossible to conduct randomised controlled trials (RCTs). Although there is not a widely accepted definition, the key element of natural experiments is that the change in exposure is caused by external shocks or factors outside researchers control, and that manipulation of exposure by researchers is not possible (1). This allows the identification of intervention and control groups. While under ideal circumstances there is an “as-if” random allocation to the intervention, it is not uncommon that potential confounding remains in the effect of exposures on outcomes of interest. In combining good knowledge of the allocation process, careful choice of methods, and transparent reporting and assumption testing, studies based on natural experiments can approximate causal evidence (1). At a time where numerous opportunities to study natural experiments will arise from the sudden and disruptive changes linked to COVID-19 (6-8), it is important to understand the barriers to conduct these studies. In this essay we build on learnings gained during our PhD trajectories focused on natural experiments. We discuss three key aspects hindering the potential of this type of research. We argue that, paradoxically, some level of control is needed to shape conditions in which evaluations of natural experiments is possible.

I. The convenience of unpredictability

There is considerable unpredictability in researching natural experiments, which may pose serious challenges for its evaluation. Unpredictability can be related to the implementation of the intervention (e.g., timing, intensity and reach), but also to aspects related to the study conduction (e.g., suitability of datasets and power). Studies using natural experiments in prospective evaluations may face difficulties aligning implementation, evaluation and funding timelines. For example, infrastructural interventions can be substantially delayed, while legislation is sometimes sooner implemented than anticipated; both impact heavily on timelines. Studies evaluating policies or interventions that have already been implemented will rely on previously collected data. This may sound like a secure route to minimise unpredictability, but exploratory data analysis is needed to assess whether assumptions and other statistical requirements of the study design are met. Not rarely, evaluations are altered or discontinued if evaluation in a meaningful way is not possible as expected when the natural experiment was identified. To overcome these challenges, researchers should be involved in early phases of intervention and policy planning, ensuring that key requirements to conduct evaluations through natural experiments are not missed. Based on our experience, early career researchers with relatively short contracts may benefit from joining existing collaborations with the fundaments for evaluation already present. Furthermore, research environments need to accommodate the intrinsic uncertainty of these studies. Providing the incentives to swiftly react on societal changes that suddenly occur are key: additional data can often be collected now, or never (8). For example, quick and flexible sources of funding have become available over the past months to study the COVID-19 pandemic. Similar initiatives are needed to combat big public health challenges that have been around for a while, including the "obesity epidemic", the persistence of social inequalities, and the climate crisis.

II. The challenge of data linkage

Even when data to evaluate the natural experiment are available, a complex factor is creating the database that includes all information needed. Linking datasets has been emphasised as an important aspect to foster the evaluation of natural experiments (9). Fortunately, databanks with linked administrative datasets are increasingly becoming available for entire countries or regions. An example is the microdata on all Dutch residents made available by Statistics Netherlands that brings together individual-level data on various domains while protecting privacy (10). Such databases provide excellent opportunities to evaluate natural experiments based on already collected data, but researchers are not always aware of their potential. Training on secondary data as formal part of research education might increase the opportunities for natural experiments evaluation. In absence of databanks, datasets need to be linked on a one-by-one basis. Informed consents – especially those that have been signed years ago – are often not designed to accommodate the linkage of datasets for the retrospective evaluation of natural experiments. Existing ethical and regulatory frameworks for sharing and processing personal data are sometimes subject for debate, making the alignment of different stakeholders a main barrier to proceed. Researchers need support by their own institutions to create multidisciplinary teams in which persons from various backgrounds (e.g., legal officers, policymakers, practitioners, researchers) jointly facilitate the timely linkage of databases within ethical and regulatory frameworks.

III. Embracing natural experiments in public health

Natural experiments provide unique opportunities to strengthen the evidence base. To increase their adoption in public health, it is essential to improve the understanding of studies based on natural experiments. Over the years, we have received disappointing reviewer comments when submitting evaluations of natural experiments to public health journals. Some of the misunderstanding may result from different ways of conceptualising natural experiments (5). Given the tendency in some journals to consider scientific rigor and associated uncertainties as more important than implications for professional practice, as well as the strict criteria for the use of causal language solely allowed for RCTs (11), means that the continuum of evidence from associations to causal conclusions is being ignored. Interdisciplinary research can be one way to learn from methodologies used by researchers in other fields (12). Training of public health professionals, non-academic stakeholders, funders, and policymakers to understand the value and specificities of natural experiments is likely needed to increase the use of these evaluation strategies.

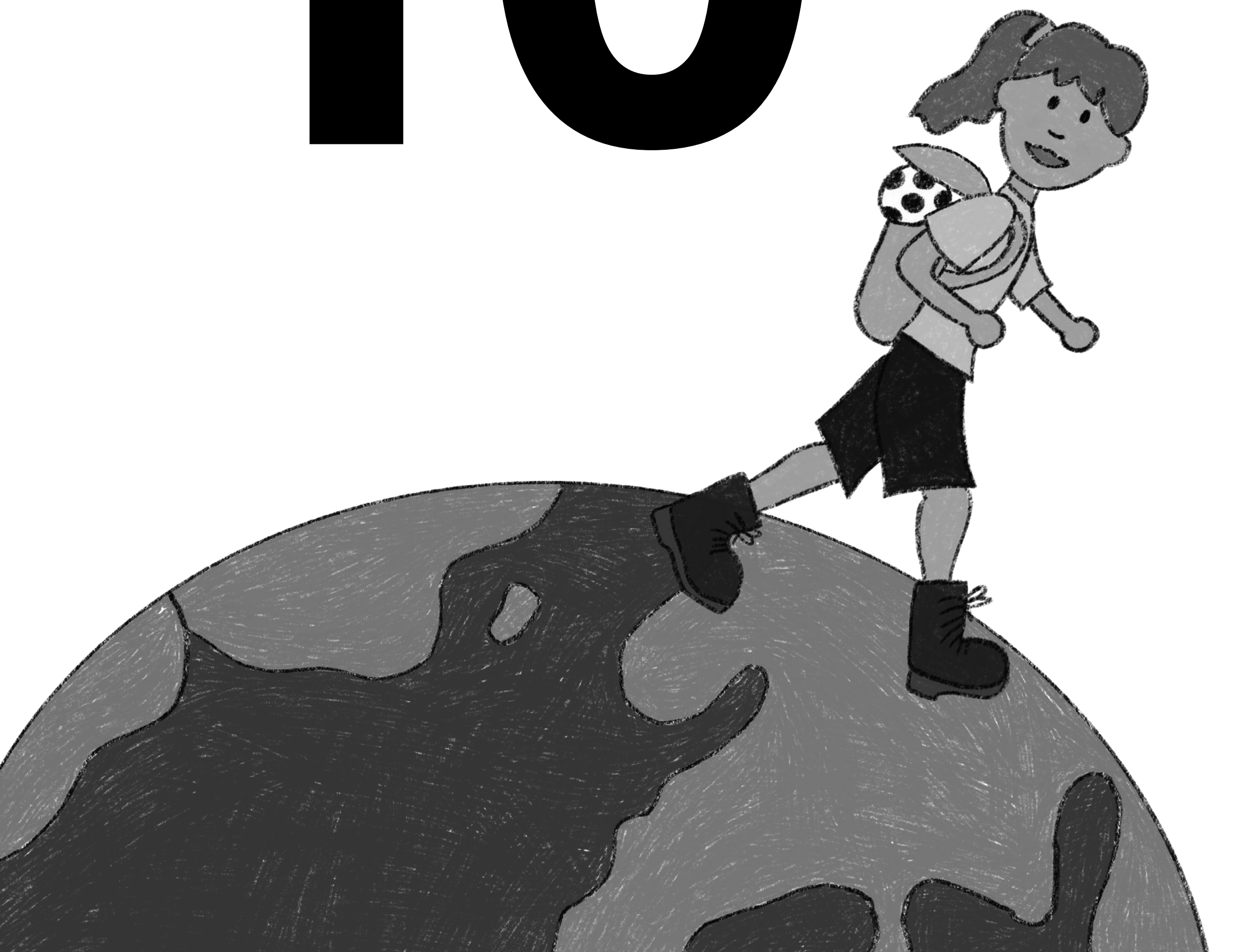
At last, we would like to draw attention to the career perspectives of researchers working on projects that capitalise on natural experiments. In a system where publications are still key to obtain new research funding and academic positions, evaluating small-scale interventions or conducting descriptive research might provide a more secure route to progress in academia (4). This may pose a serious risk that the existing "evaluative bias" will increase, whereby most evidence is available for interventions that were easiest to study and to publish (13). We need institutional changes where researchers are acknowledged for the societal relevance of their studies, not primarily on the quantity of publications. These risks are even larger for PhD candidates with strict timings and output objectives, possibly demotivating early career researchers from devoting their projects to evaluate natural experiments. Ultimately this may lead to less senior researchers being experts on natural experiments, and moving it forward in public health. Better understanding of the use and value is needed to ensure that brave researchers more often evaluate the interventions that hold large promise to inform decisions about population health.

In conclusion, natural experiments provide opportunities to inform policymaking on exposures that are impossible to randomise. While successful evaluations are available in literature, much can be learned from the barriers ultimately leading to unexplored opportunities, ceased projects, and unpublished manuscripts. As long as barriers are not addressed jointly by the research and policy environments, opportunities to provide evidence with extensive societal impact will continue to be missed.

REFERENCES

1. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health. 2012;66(12):1182-6.
2. Petticrew M, Cummins S, Ferrell C, Findlay A, Higgins C, Hoy C, et al. Natural experiments: an underused tool for public health? Public Health. 2005;119(9):751-7.
3. Barnighausen T, Tugwell P, Rottingen JA, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series - Paper 4: uses and value. J Clin Epidemiol. 2017;89:21-29.
4. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. Journal of Epidemiology and Community Health. 2020;74(2):203-8.
5. de Vocht F, Katikireddi SV, McQuire C, Tilling K, Hickman M, Craig P. Conceptualising natural and quasi experiments in public health. BMC Medical Research Methodology. 2021;21(1):32.
6. Been JV, Sheikh A. COVID-19 must catalyse key global natural experiments. J Glob Health. 2020;10(1):010104.
7. Thomson B. The COVID-19 Pandemic: A Global Natural Experiment. Circulation. 2020;142(1):14-6.
8. Erren TC, Lewis P, Shaw DM. The COVID-19 Pandemic: Ethical and Scientific Imperatives for “Natural” Experiments. Circulation. 2020;142(4):309-11.
9. Academy of Medical Sciences London. Improving the health of the public by 2040. 2016.
10. Bakker BFM, van Rooijen J, van Toor L. The System of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. Statistical Journal of the IAOS. 2014;30(4):411-24.
11. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. American Journal of Public Health. 2018;108(5):616-9.
12. Matthay EC, Hagan E, Gottlieb LM, Tan ML, Vlahov D, Adler NE, et al. Alternative causal inference methods in population health research: Evaluating tradeoffs and triangulating evidence. SSM - Population Health. 2020;10:100526.
13. Ogilvie D, Egan M, Hamilton V, Petticrew M. Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go? J Epidemiol Community Health. 2005;59(10):886-92.

10



General discussion

The main objectives of this thesis were to evaluate public health interventions introduced by local governments, and to appraise study designs and analytical methods used to evaluate interventions by means of natural experiments. This final chapter presents the main findings in light of these objectives, discusses methodological challenges, presents a guidance for evaluating natural experiments, and makes recommendations for future directions in research and policymaking.

MAIN FINDINGS OF THIS THESIS

Findings part I – Evaluation of local public health interventions

Chapter 2 examined the introduction of 18 physical activity spaces in deprived neighbourhoods. Findings indicated that outdoor play and sedentary behaviours were unaffected by the opening of a new physical activity space near home. Although not statistically significant, subgroup analyses suggested that reducing the distance to the nearest physical activity space increased outdoor play for children from socially disadvantaged families. For these children, sedentary behaviours including television viewing and computer gaming also increased. These findings were not observed for children from socially advantaged families. The results suggest that improving a neighbourhood environment that allows for outdoor play is more important for children from lower than higher socioeconomic subgroups. It also suggests that other interventions are needed to decrease sedentary behaviours among children.

Chapter 3 studied inequalities in exposure to the food environment near home, and their contribution to childhood obesity. Children from lower educated mothers were exposed to more fast food outlets and an unhealthier food environment as compared to children from higher educated mothers. Furthermore, existing inequalities in the food environment widened over time. In contemporary societies that could best be described as "fast food paradises" given the abundancy of fast food outlets, subsequent changes in food outlets were not associated with children's body composition. The introduction of fast food outlets was associated with small increases in body-mass index for children from lower educated mothers initially not living in close proximity of fast food outlets. This pattern was not observed for children from higher educated mothers. Policymakers should aim at improving the food environment around home, and prioritise the poor environment of children from lower socioeconomic groups.

Chapter 4 evaluated the effect of hosting elite sport events on recreational sport participation. From 2010 onwards, events specifically aimed at increasing sport participation in the population; we evaluated if this shift in focus resulted in higher sport participation levels. We showed that three out of the ten sport events were followed by an increase in sport participation, while one was followed by a decrease. The sport events organised in the period with the specific aim to target sport participation, were followed by an increase in sport participation in the general population. No differences in sport participation were found between events with lower or higher number of visitors, and between events located in the city centre or in stadiums in outskirts. Not all recent elite sports events with an aim to increase recreational sport participation were able to do so. Therefore, we need a better understanding of the type of events and associated promotional activities that will lead to an increase in sport participation in the general population.

Chapter 5 outlined a study protocol for a systematic review and meta-analysis to evaluate smoke-free policies on tobacco smoke exposure and respiratory health in children. Chapter 6 presented the findings of this comprehensive assessment. 11 studies were included, of which six evaluated smoke-free car policies. Pooled estimates showed that the introduction of smoke-free car policies was followed by substantial reductions in tobacco smoke exposure in cars among children, potentially reducing the health burden associated with tobacco smoke exposure. We found some evidence suggesting that the implementation of smoke-free school grounds and city-wide smoke-free policies protects children from the deleterious effects of tobacco smoke exposure, but more studies are needed to support these findings.

Findings part II – Methods and approaches to evaluate natural experiments

Chapter 7 presented the findings of a systematic review on the methodological approaches undertaken to evaluate interventions aimed at improving the cycling infrastructure in cities using natural experiments. We included 31 studies and most evaluations found that the number of cyclists using the facilities increased, and to a lesser extent that persons changed their behaviour towards more cycling.

Larger changes were found for studies that tested for statistical significance and for studies that used subjective measurement methods, as compared to those that did not perform statistical tests, and those that used objective measurement methods. Triangulation of methods is warranted to strengthen the basis of causal inference, and to overcome practical limitations that one may encounter when evaluating built environment interventions.

Natural experiments often rely on already collected data, and health surveys may provide data before and after an intervention was introduced. However, most surveys have experienced major declines in response rates, which could have biased findings of studies that evaluate trends in outcomes. Chapter 8 outlined a framework to explore non-response patterns over time in repeated cross-sectional surveys. The framework presented in this chapter consists of four steps that allow to estimate whether patterns in non-response over time differ across sociodemographic subgroups, and to assess the potential influence on trends of prevalence in outcomes and associations between sociodemographic characteristics and outcomes. We applied this framework to a population survey in which the response rate gradually declined. Despite a major decline in response, we did not find evidence that the risk of non-response bias increased over time in this specific survey.

This thesis describes successful evaluations of natural experiments, but it is important to understand the barriers to do so. In **chapter 9**, we reflected on the challenges encountered during PhD trajectories. Paradoxically, some level of control is likely to be needed to allow for meaningful evaluations of natural experiments. We have identified three main aspects hindering the adoption of natural experiments. First, timelines of natural experiments can be highly unpredictable. Collaborations between researchers and policymakers is warranted to ensure that key elements for evaluation are secured. Furthermore, flexible sources of funding are needed to be able to respond to sudden opportunities for evaluation that may arise at unforeseen times. Second, data linkage can be a real challenge. Setting up data platforms and multidisciplinary teams is important to facilitate data linkage in a timely manner within ethical and regulatory frameworks. Third, we need to embrace evidence derived from natural experiments. Acknowledging the value of studies that capitalise on natural experiments is essential to increase the use of these evaluation strategies. It is important to create a scientific and policy environment in which researchers are able to discover and evaluate natural experiments in all its complexities.

Findings through an equity lens

Population-level interventions that do not heavily rely on individual capacities to engage in health-related behaviours are suggested to reduce socioeconomic inequalities in health (1, 2). In each of the chapters, we conducted subgroup analyses to study differential effects if data allowed and if deemed appropriate. In **chapters 2 and 3**, findings suggested that built environment changes around home (i.e., physical activity spaces, and fast food outlets) were impacting on behaviours of children from lower socioeconomic groups, but not on children from higher socioeconomic groups. In the systematic reviews on smoke-free policies and cycling infrastructure, we explored equity effects by extracting relevant information as presented by the included studies. **Chapters 6 and 7** reported on the findings of the systematic reviews, but only few studies investigated differential effects. One of the 11 included studies in the review on smoke-free policies reported if some subgroups benefited more than others. Three of the 31 studies included in the review on cycling infrastructure tested for equity effects. Given the limited evidence available, we cannot draw conclusions on equity aspects following smoke-free policies and cycling infrastructure.

METHODOLOGICAL CONSIDERATIONS

Natural experiments and the C-Word

The disadvantage of lacking control over the intervention includes the uncertainty around causal inference, the "C-Word" that is often avoided in studies without randomisation (3). There are three key assumptions that needs to hold to ensure that one factor *caused* another factor, namely exchangeability, positivity and consistency (4). The so-called "fundamental problem of causal inference" indicates that it is impossible to observe the effect of the intervention on the outcome, and at the same time the effect of the control condition on the outcome within the same individual (5). Indeed, this problem applies to randomised and non-randomised studies. To overcome this fundamental problem, it is essential to identify intervention and control groups that are comparable, and to select an appropriate study design and analytical methodology. Ideally, the estimated intervention effect would be similar to a hypothetical scenario where the control group would be exposed to the intervention instead of the intervention group, and *vice versa*. However, exposure to public health interventions is seldom a random act, thus identifying intervention and control groups that are truly exchangeable will not often occur. All evaluations of interventions that capitalise on natural experiments are at risk of bias, but some studies are more likely to estimate causal effects than others (6). A common procedure to control for determinants related to both the intervention and outcome is to add them as confounders in multivariable models. Unfortunately, confounding variables are often unmeasured or unknown, and therefore residual confounding is likely to remain present. Alternative study designs and analytical methodologies are needed under these circumstances to avoid risk of bias due to selective exposure to the intervention and residual confounding (7). In the primary research articles presented in chapters 2 to 4, we minimised the risk of bias using approaches that do not primarily rely on confounder control. In the following paragraphs, we will describe the used methods and analytical techniques used to minimise the risk of bias in each of the studies.

First, we aimed to avoid selective exposure to the intervention and therefore evaluated changes in exposure to interventions over time that were outside control of individuals. In **chapter 2**, the majority of physical activity spaces were initiated by the local government. The government does not systematically investigate the preferences for new physical activity spaces in specific neighbourhoods, thus new facilities are likely to be randomly introduced in deprived neighbourhoods that met the criteria set by the charities that offer financial support. In **chapter 3**, we studied the change in exposure to fast food outlets. On the long term, it is plausible that fast food outlets arise in areas with a high demand for this type of food. However, changes to fast food outlets within 2 to 4 years between measurements are unlikely to be strongly influenced by individual preferences. In **chapter 4**, we studied the changes in sport participation following the hosting of international elite sport events. The selection of international sport events is partly guided by the interest of the population, but the year of hosting an event and changes in policy priority can be considered a random act.

Second, we aimed to identify intervention and control groups that were likely to have similar characteristics, except for their exposure status. In **chapters 2 and 3**, we used the home address to calculate exposure to physical activity spaces and fast food outlets around home at multiple time-points. As a result of this individual assignment, children living in one neighbourhood may have had different exposure levels despite living in the same neighbourhood and likely sharing similar characteristics. In **chapter 4**, intervention and control group were based on the year in which the elite sport events were organised, and the timing of the shift in policy focus towards targeting recreational sport participation. The characteristics of the population before and after the change in policy priority are likely to be comparable. The rapid decline in response over time may have introduced some bias, especially if the reduction in response is faster in some subgroups than in other subgroups. Additional analyses presented in **chapter 8** suggested that the differential change in response did not introduce bias in trends in sport participation in the population.

Third, we applied state-of-the-art methodologies to reduce the risk of bias that resulted from unmeasured confounding variables. In **chapters 2 and 3**, we used fixed-effect models to take into account unmeasured time-invariant confounding, and extended models by additionally controlling for measured time-varying confounders such as income. **In chapter 4**, we used interrupted time series analysis to control for the pre-intervention trend in outcomes. We also conducted a falsification test where the time of interruption was placed at random, to ensure that the change in outcomes was not reflecting the time trend in outcome.

Fourth, for studies that evaluated built environment interventions, we reduced the risk of bias that may result from selective relocation. In **chapter 2**, we studied the relation between new physical activity spaces and time spent playing outdoors and sedentary behaviours. In **chapter 3**, we evaluated the relation between exposure to fast food around home and body composition measures. In both studies, we restricted the population to those who did not move houses during follow-up, to avoid that residential moves influenced the findings.

Fifth, we outlined study designs that closely mimic a hypothetical target trial (8). In each of the studies we asked: how would a trial look like? Who would be invited to participate in a trial? What would define the intervention and comparison groups? What analytical approaches would be needed? Causal inference in observational studies may be hindered if the approach does not align with what would have been done in an experimental setting (8). To approximate an experimental setting, we excluded some observations. In **chapter 2**, we excluded children who were already exposed to physical activity spaces at baseline. In **chapter 3**, we conducted an additional analysis restricting the population to children not exposed to fast food outlets at start of the study. Excluding observations from the analysis may result in substantial power losses. However, it is important that the study population and analytical sample is correctly defined and will allow to answer the question of interest.

Towards further improvement of causal evidence

The lack of randomisation and the related uncertainty about who was exposed to the intervention require careful examination of the robustness of the study findings. Combining different types of evidence, for example by using triangulation and mixed-methods studies, may be useful for this purpose (7, 9-12). Triangulation is the practice of obtaining multiple estimates of the intervention effect using combinations of study designs, datasets, and analytical methods, whereby each approach addresses different sources of potential bias (10). Mixed-methods evaluation of natural experiments are characterised by using multiple sources of evidence on a number of outcomes to study the wider effect of the intervention, and to demonstrate causal pathways. We were not able to support the findings presented in chapter 2 to 4 with additional insights from other databases on the same outcome, or on other outcomes to demonstrate causal mechanisms. In the following paragraphs we will introduce two studies to illustrate how different sources of data can be used to strengthen causal inference. Due to data linkage difficulties and the COVID-19 pandemic that disturbed data collection, we were not able to present the findings of these evaluations in this thesis.

First, we illustrate how triangulation might be used to evaluate the long-term effectiveness of the school-based intervention Lekker Fit! (in Dutch: "*enjoy being fit*") aimed at reducing overweight among school-aged children. The main component of the intervention is the delivery of 3 physical education (PE) sessions a week provided by trained PE teachers. The results from a cluster RCT performed in 2006-2007 gave some first indications of the effectiveness of the intervention after 1-year of follow-up (13). The findings indicate that obesity prevalences were lower, waist circumference reduced, and fitness increased for children aged 6 to 9 years. The results of the RCT contributed to the implementation of the program to cover nearly half of the primary schools in the city of Rotterdam, the Netherlands. In 2018, 18,000 children aged 6 to 12 years were reached. There are no insights on the long-term effects of multiple years of the intervention, and on the sustained behaviour change after leaving elementary school. Therefore, we aimed at evaluating the long-term effectiveness of the school-based intervention Lekker Fit!.

In absence of randomisation, it is important to consider the differences between children attending intervention and control schools. Most important here is that the intervention is introduced at schools in lower income neighbourhoods, where children are at highest risk of developing overweight and obesity. Therefore, we cannot simply compare overweight and obesity rates between intervention and control schools. The phased introduction of the intervention since 2006 can be used to identify intervention and control groups that are more comparable. The intended evaluation aimed to use three different databases that repeatedly measured body composition at different time-points during childhood. Each of the databases selected participants in a different way, collected information on a different set of outcomes and confounding variables, and varied in timings of measurements and years of follow-up. Another strength of this study would be that the short-term findings derived from the three evaluation studies could be compared with the findings of the cluster RCT (13). If the short-term effects of the cluster RCT are comparable with the short-term effects found in the three databases, one can be more certain that the long-term effects of the intervention – beyond the follow-up time of the experiment – are also attributable to the intervention. Panel 1 describes the three datasets that were considered. Unfortunately, privacy issues have delayed the data access to such an extent that the study could not be conducted as part of this thesis

Triangulation has the major benefit that multiple estimated intervention effects can be compared. If all estimates point in the same direction, or if differences can be sufficiently explained by elements in which the evaluations differ (e.g., controlled confounders, outcome assessment, selection of the study population), one can be more confident about the estimated intervention effect. Natural experiments that closely follows the design of previously conducted experiments can yield important insights on the risk of bias. Designing RCTs that follow persons for years is often unwarranted, impractical of even unethical. Natural experiments can overcome this limitation by evaluating random variation in the population for longer periods than RCTs allows.

Panel 1 – Triangulation to evaluate a school-based intervention

Data from the intervention schools: This database includes yearly measurements for all children joining the intervention schools. Measurements include body weight, body-mass index (BMI), and the 20 meter Shuttle Run test as an indicator of physical fitness. All measurements were collected during physical education lessons. Data would allow to answer the question if body weight, BMI and physical fitness changed for children that were 1 to 6 years exposed to the intervention. The benefit of this database is that it contains a large number of observations of children from more socially disadvantaged families, a group that is underrepresented in most health surveys and cohorts. A limitation is the lack of data on changes in outcomes for children attending control schools with a regular curriculum.

Data from the Child & Family Centres: This database consists of children's body weight, BMI, and psychosocial wellbeing, collected during regular monitoring activities at the Child & Family Centres during childhood. All children living in the city of Rotterdam were invited. We therefore could compare the changes in outcomes between children attending intervention and control schools, adjusting for a limited number of confounders. An advantage of the dataset is that all children are invited, and the majority are attending the regular monitoring activities. Therefore, the risk of selection bias is expected to be low. The evaluation is at high risk of bias due to confounding by unmeasured factors that may have influenced whether children were attending intervention or control schools.

Data from a birth-cohort study: The database from the Generation R study, an ongoing birth-cohort study in the city of Rotterdam, includes information on children's body weight, BMI, psychosocial wellbeing, and fat-mass and fat-free mass derived from DXA-scans. This data would be used to evaluate the change in outcomes for children that were attending intervention schools as compared to control schools, adjusting for a wide range of confounders. A strength of this database is that information on a variety of demographic and socioeconomic factors was collected, and that fat-mass was measured. A disadvantage is that children from more socially disadvantaged families less often participate, therefore the generalisability of the study findings may be limited.

As a second example, we introduce a study that uses multiple data sources to evaluate the introduction of taps at public squares around Dutch train stations. In November 2018, the Dutch railway operator announced that they will install 200 taps at train stations by the end of 2019 so that more than 90% of the passengers have access to facilities to refill bottles with water at train stations. We hypothesised that the introduction of taps would increase water consumption, limit the intake of sugar-sweetened beverages, and reduce the environmental impact of single-use plastic bottles. The evaluation study was based on this natural experiment. The introduction of taps was staggered per region, and the order was dependent on the cooperation with water companies and the ease of building the facilities in the specific areas. Most importantly, the order was independent from travellers requests or needs, and therefore was unlikely to be related to the behavioural outcomes under study. We contacted the Dutch railway operator to identify possibilities for evaluation, and agreed on the outline of the study. Three rounds of data collection took place in 2019. We planned other rounds of data collection after completion of the intervention in 2020, but this was not possible due to the COVID-19 pandemic. Restrictions to travel by public transport were in place, and travellers who used public transport during the pandemic might be hesitant using public taps to minimise the risk of becoming infected. The lack of data in the same season the year after implementation did not allow to disentangle changes in behaviours that were the result of the introduction of taps, from changes in behaviours due to weather conditions. An evaluation was not possible and we had to abort the study.

The data sources for the planned evaluation are presented in Panel 2. The evaluation builds on data collected by means of surveys, purchasing data from shops at train stations, and the amount of plastic littered at the train stations. Each of the databases allows to study a specific outcome related to the introduction of the taps. Combining the evidence from three data sources would allow to study in detail the multiple effects of the intervention at various domains, and may identify causal mechanisms underlying the intervention effect. If all results point in the same direction, one would be more confident that the intervention was effective. For example, if the evaluation would indicate that the number of tap users increased and the number of water bottles sold decreased, one could be more certain that the introduction of taps has influenced persons behaviours when travelling by train.

Panel 2 – Multiple data sources to evaluate taps at train stations

Surveys among train travellers: A survey was designed to collect information on tap use, beverage consumption during traveling by train, purchasing behaviour at train stations, and the usage of single-use plastic bottles. The survey was sent to a panel of frequent train travellers at several time-points. Data would allow to answer the question if the taps were used by train travellers and if the usage increased over time, and how the taps were perceived. The data was also used to evaluate the effect of taps on the consumption of water while travelling by train, the amount of beverages sold at train stations, and the usage of single-use plastic bottles.

Sales data from train stations: This database includes monthly beverage sales from shops at a selection of train stations. Data were available for shops owned by the Dutch railway operator, and we only included smaller stations that do not have other shops that sell hot beverages. This database would have been used to evaluate if the introduction of taps changed the number of water bottles and sugar sweetened beverages sold in shops at train stations.

Plastic littered at trains and train stations: This database contained the monthly amount of plastic littered at trains and train stations. Data would allow to investigate if the amount of plastic littered changed following the introduction of taps at train stations.

Environmental interventions may trigger a wide range of responses, and important information on the effects of an intervention may be missed if focussing on a limited set of outcomes. Also the context in which the intervention was introduced may have played a role, and research is needed that understands how intervention works in varying contexts (14). Reviews that capitalise on natural experiments should allow for summarising the evidence derived from multiple data sources and reporting on multiple outcomes. The two reviews conducted as part of this thesis evaluated smoke-free policies to reduce the burden associated with tobacco smoke exposure in children, and new infrastructure to promote cycling. These reviews are presented in **chapters 6 and 7**, and we extracted all information from quantitative and qualitative sources presented in the included studies. We did not explore to which extend the context enabled or constrained behaviour change. Some studies reported on additional elements, such as multiple comparison groups to test the robustness of findings, or evidence from related sources that allowed for exploration of causal mechanisms (e.g., surveys, policy documents or qualitative studies). Most studies relied on a single database and provided only limited insights on the mechanisms how the intervention may have caused an effect.

Embracing the evidence derived from natural experiments

In this thesis, we used observational data to estimate causal effects of interventions. Building on natural experiments allows to achieve strong causal evidence between interventions and outcomes in a real-world setting. To date, the valorisation of evidence derived from natural experiments is limited, which may hinder its usage and potential to inform policymaking. The scientific community tend to have strict criteria for causal language solely allowed for RCTs (3), and as a consequence the continuum of evidence derived from natural experiments – from associations to causal conclusions – is being ignored. RCTs can be considered as the simplest settings to identify causal effects, however, observational data can be used to approximate the evidence derived of trials (5). The expected lack of confounding makes researchers and policymakers *feel* more confident of a causal interpretation of the estimated intervention effect (3). Despite randomisation and causal language, findings from RCTs needs to be interpreted with care. The strong selection of participants that are willing to participate in RCTs may hinder the generalisability of the results to the target population (15). Effect estimates derived from a highly selected population are not necessarily a good estimate of the effect in the general population. The benefit of evidence derived for natural experiments is that it provides real-world evidence of how populations are affected by the intervention under normal circumstances (12). Another important feature of natural experiments is that it allows to evaluate change in exposures that are impossible to be studied through RCTs (9). A better understanding on the value and specificities of natural experiments, and their usage for identifying causal effects, is urgently needed. Relying on RCTs as the sole form of estimating causal effects may hinder the adoption of natural experiments.

GUIDANCE FOR EVALUATING NATURAL EXPERIMENTS

To date, most natural experiments in the field of obesity prevention rely on data collected as part of health monitoring surveys, ongoing or pre-existing cohort studies, hospital administrative data, or other secondary data (16, 17). The interventions evaluated as part of this thesis were using data from existing cohort studies and monitoring surveys. Based on our experiences, we identified 7 steps that are of importance to identify and evaluate natural experiments. A summary of the key steps for the evaluation of natural experiments is presented in the Panel 3.

Panel 3 – Key steps for evaluating natural experiments

1. Identify natural experiments seeking evaluation
2. Create common interest for the new research
3. Identify data sources
4. Write a study proposal
5. Acquire funds
6. Collect data or achieve data access
7. Analyse data

Step 1 compromises of the identification of experiments from which a substantive population health impact can be expected. Ideally, researchers are involved in early phases of intervention and policy planning to ensure that all opportunities for evaluation are explored. Changes in the built environment or legislative measures may be announced years before actual implementation, and as such offer the opportunity to design an evaluation study. Some projects that we identified could not be evaluated, but with some small adaptions in the past would have allowed for an impact evaluation. Therefore, it is crucial that researchers and policymakers are jointly identifying opportunities for evaluation. Research environments that accommodate long-term collaborations between various stakeholders may allow to react on sudden opportunities for evaluation. We searched for promising environmental interventions introduced at the local level, and that were possible to evaluate within a 4-year time period. Opportunities outside public health are recognised to be of importance to promote health-related behaviours and reduce health inequalities (18, 19), but in this thesis we were not able to study interventions related to other domains such as education, work, and housing.

Step 2 describes the collaboration with key stakeholders (e.g., policymakers, local partners, other research institutes) to create common interest for the evaluation of the identified experiment, and to build capacity to conduct the study. Collaboration is needed to identify data sources that might be used for evaluation, and in a later stage for accessing data, acquiring funds, and disseminating the study findings. Furthermore, the insights from stakeholders are crucial for understanding the allocation procedures. Step 2 is a major factor for the successful use of natural experiments.

Step 3 includes the exploration of potential data sources. For some evaluations it may be possible to collect additional data, and these possibilities should be explored. We had to rely on data that was already collected for other research purposes, and it is important to search for data collected by various stakeholders. The main reason that some identified interventions were not further considered was that insufficient data were available, or that outcome data were not closely linked to the intervention. The latter is important, since reviews have shown that studies measuring more distal outcomes were more likely to conclude that the intervention had no impact, as compared to studies that reported on more proximal behaviours (20, 21).

Step 4 is the writing of a study proposal. This requires epidemiological knowledge on study designs and analytical methods. The best evaluation strategy needs to be outlined, depending on the data availability, which often requires methodological creativity. Intervention and outcome data is seldom presented on a silver platter in settings where researchers lack control over the intervention. Understanding the specificities and requirements of studies that capitalise on natural experiments is needed to design an evaluation study that allows to answer the research questions of interest. Training of public health professionals, non-academic stakeholders, funders, and policymakers to understand the value and specificities of natural experiments is essential to increase the use of these evaluation strategies.

Once a study proposal is agreed upon, step 5 may be needed to acquire additional funding. During the COVID-19 pandemic, flexible sources of funding were available to study experiments related to the enormous disruptions in our society. Similar funding opportunities are needed to ensure that unexpected variations in exposure can be evaluated that have the potential to influence population health.

Step 6 and 7 consist of data access and data analyses. The process of collecting data or accessing data can be time consuming, especially if the duration of the project is long (e.g., when aiming to collect data before and after large infrastructural interventions) or if multiple data sources need to be linked (e.g., when legal advice is needed to link datasets). Only at the stage when the dataset is available for the researcher, one can explore whether the specific requirements related to the study design are met. Some studies will be discontinued after months of exploration, because the datasets or analytical approaches do not allow to evaluate the intervention in a meaningful way. The two studies presented in Panel 1 and 2 were discontinued after investing months of time, because the data did not allow for an evaluation that would make a contribution to the existing evidence base, or simply because the data could not be shared in time.

FUTURE DIRECTIONS OF NATURAL EXPERIMENTS

Call for consequential research

The evidence base in social epidemiology is dominated by studies that focus on causes and distributions of diseases in populations (22). Although such studies will remain important to inform policymaking, we also need epidemiology that is more "consequential" in nature, guiding policymakers how to address public health issues by means of interventions or policies (22, 23). For example, insights on the role of the neighbourhood environment on health-related behaviours is dominated by cross-sectional studies (e.g., (24-26)), but what the impact is of specific interventions to alter the neighbourhood environment remains often unknown. Natural experiments provide excellent opportunities to derive more consequential evidence whereby the relationship between changes in exposures and outcomes is studied, guiding policymaking nationally and locally.

Towards a healthy, equitable and sustainable future

In 2015, all United Nations (UN) Member States have adopted a set of 17 Sustainable Development Goals (SDGs) to achieve a better, equitable, and more sustainable future for all (27, 28). To meet these goals, we need to create the conditions that foster health and well-being in a more sustainable way. For example, by promoting healthy diets rich in plant-based sources to reduce the environmental impact of current diets (29, 30), or by creating neighbourhoods that promote high levels of walking and active transportation in expense of car use (31). The SDGs make clear that most measures towards sustainability require multiple actions that create synergy within a particular SDG, but also create synergies among different SDGs. National and local interventions are needed that focus on these synergies and to accommodate shifts in behaviours. To allow for the evaluation of natural experiments to promote health and sustainable behaviours, the linkage of databases from various sources will become increasingly important. To ensure a future for natural experiments, we should invest today in data collection platforms that allow evaluation of the most promising interventions in the near future across domains.

What type of experiments at the local level can we expect in the near future? It is forecasted that the health effects of climate change will considerably compromise population health (32). Future local experiments are likely to be centred around the direct consequences of heavy rainfalls, windstorms, and heat waves, but will also address indirect consequences of climate change such as migration (33). It is important to create a policy and research environment in which sudden disruptions in our society can be evaluated, initiated within and outside the health domain, using various types of data that were not primarily collected for research purposes. A flipped socioecological model where the environment instead of the individual is placed at the centre may facilitate a change in thinking from interventions targeting individuals towards interventions targeting the environments as causes of obesity, undernutrition and climate change (32).

FUTURE DIRECTIONS IN POLICYMAKING

To curb the high levels of smoking, physical inactivity and poor diets, and the increasing prevalence of overweight and obesity in the population, bold policy decisions are needed. Many public health interventions still rely on conveying information on individual's disease risk, but it has been argued that we should move towards approaches that target the environments to facilitate this behavioural change (34). In this thesis we aimed to contribute to the expansion of natural experiments to evaluate local interventions targeting populations. Our findings showed that environmental interventions may induce behaviour change, but that the impact may differ between population subgroups. It is important to understand these differences so that future interventions are designed more equitable. Local governments must commit to creating environments in which health-related behaviour are self-evident. Creating safe opportunities for outdoor play within neighbourhoods, ensuring access to affordable and nutritious foods, creating smoke-free environments, and building high-quality cycling infrastructure are some interventions that local governments can implement to support health-related behaviours of its citizens. Most important when implementing such interventions is to plan for an evaluation so that we better understand how changes impact the population and specific population subgroups. We encourage policymakers and researchers to timely collaborate to ensure that possibilities for meaningful impact evaluations are explored. Collaboration between governments, companies, researchers and the public is needed to create cities in which health-related behaviours are promoted.

Industry has a large influence on the public and policymakers understanding of health issues, which may hinder the implementation of effective public health policies (35). Also at the local level, we need more insights on the so-called "commercial determinants of health" and the role of the private sector through actions directed to individuals or by shaping the political environment (36). One may think that only few people support taxation or regulatory interventions when following debates in newspapers, television and on social media, but studies have shown that public support is remarkably high. For example, 40% of the Dutch adult population is supportive of a sugar-sweetened beverage tax, and this percentage substantially increased to 66% if the earnings are invested on health initiatives (37). Presenting quantitative evidence of the health effects of interventions has been shown to increase public support (38), emphasising the need to measure the impact of interventions and to effectively communicate the evidence to the public. For the latter, scientific products and activities (e.g., manuscripts, research projects, or programmes) are unlikely to reach the public and alternative strategies are needed to create societal impact of research activities.

Achieving a healthier Netherlands, make it happen!

In 2018, the Dutch government in collaboration with 70 public and private organisations has presented the National Prevention Agreement (39). The agreement describes the ambitions and strategies to improve the health of the Dutch population by 2040. It addresses three behaviours that are known to largely contribute to the disease burden: smoking, overweight, and problematic alcohol consumption. It is promising that prevention is acknowledged as an important strategy to address public health issues. Encouraging is that a modelling study forecasted that the preventive measures related to smoking will be sufficient to meet the smoking-related goals by 2040 (40). However, the same study showed that only limited impacts can be expected from the strategies taken to reduce overweight and problematic alcohol consumption in the population (40). More comprehensive measures are needed to ensure that the goals are met. Evidence derived from natural experiments is essential to select the best strategies to counteract the rise in overweight and obesity in the population, and to decrease the prevalence of problematic alcohol consumption.

In conclusion, the implementation of effective public health policies can considerably improve the health of populations. Natural experiments can be helpfull to study the impact of such policies. The introduction of policies should be based on the best evidence available, and we should strive for collecting the best evidence possible. Policymakers, local partners and researchers should jointly identify possibilities for evaluation to ensure that future interventions are designed more effective and equitable. We should look for opportunities beyond domains, disciplines, and conventional methods to ensure that we know more tomorrow than we know today. Every day is a natural experiment!

REFERENCES

1. McLaren L, McIntyre L, Kirkpatrick S. Rose's population strategy of prevention need not increase social inequalities in health. International Journal of Epidemiology. 2009;39(2):372-7.
2. Beauchamp A, Backholer K, Magliano D, Peeters A. The effect of obesity prevention interventions according to socioeconomic position: a systematic review. Obesity Reviews. 2014;15(7):541-54.
3. Hernán MA, The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. American Journal of Public Health. 2018;108(5):616-9.
4. Rehkopf DH, Glymour MM, Osypuk TL. The Consistency Assumption for Causal Inference in Social Epidemiology: When a Rose is Not a Rose. Curr Epidemiol Rep. 2016;3(1):63-71.
5. Holland PW. Statistics and Causal Inference. Journal of the American Statistical Association. 1986;81(396):945-60.
6. Humphreys DK, Panter J, Ogilvie D. Questioning the application of risk of bias tools in appraising evidence from natural experimental studies: critical reflections on Benton et al., IJBNPA 2016. International Journal of Behavioral Nutrition and Physical Activity. 2017;14(1):49.
7. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Health. 2017;38:39-56.
8. Hernán MA. Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.
9. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health. 2012;66(12):1182-6.
10. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. International Journal of Epidemiology. 2017;45(6):1866-86.
11. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. International Journal of Epidemiology. 2016;45(6):1776-86.
12. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. Journal of Epidemiology and Community Health. 2020;74(2):203-8.
13. Jansen W, Borsboom G, Meima A, Zwanenburg EJ-V, Mackenbach JP, Raat H, et al. Effectiveness of a primary school-based intervention to reduce overweight. International Journal of Pediatric Obesity. 2011;6(2Part2):e70-e7.

14. Panter J, Guell C, Humphreys D, Ogilvie D. Can changing the physical environment promote walking and cycling? A systematic review of what works and how. Health Place. 2019;58:102161.
15. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. Prev Sci. 2015;16(3):475-85.
16. Bennett WL, Wilson RF, Zhang A, Tseng E, Knapp EA, Kharrazi H, et al. Methods for Evaluating Natural Experiments in Obesity: A Systematic Review. Ann Intern Med. 2018;168(11):791-800.
17. Crane M, Bohn-Goldbaum E, Grunseit A, Bauman A. Using natural experiments to improve public health evidence: a review of context and utility for obesity prevention. Health Res Policy Syst. 2020;18(1):48.
18. Whitehead M, Dahlgren G. What can be done about inequalities in health? Lancet. 1991;338(8774):1059-63.
19. Marmot M, Bell R. Fair society, healthy lives. Public Health. 2012;126:S4-S10.
20. Mayne SL, Auchincloss AH, Michael YL. Impact of policy and built environment changes on obesity-related outcomes: a systematic review of naturally occurring experiments. Obes Rev. 2015;16(5):362-75.
21. Stappers NEH, Van Kann DHH, Ettema D, De Vries NK, Kremers SPJ. The effect of infrastructural changes in the built environment on physical activity, active transportation and sedentary behavior – A systematic review. Health & Place. 2018;53:135-49.
22. Nandi A, Harper S. How Consequential Is Social Epidemiology? A Review of Recent Evidence. Curr Epidemiol Rep. 2015;2(1):61-70.
23. Galea S. An Argument for a Consequentialist Epidemiology. American Journal of Epidemiology. 2013;178(8):1185-91.
24. Cerin E, Nathan A, van Cauwenberg J, Barnett DW, Barnett A, et al. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. International Journal of Behavioral Nutrition and Physical Activity. 2017;14(1):15.
25. Timperio A, Reid J, Veitch J. Playability: Built and Social Environment Features That Promote Physical Activity Within Children. Current Obesity Reports. 2015;4(4):460-76.
26. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CAM. The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results. Obesity. 2015;23(7):1331-44.
27. Costanza R, Fioramonti L, Kubiszewski I. The UN sustainable development goals and the dynamics of well-being. Frontiers in Ecology and the Environment. 2016;14(2):59-59.

28. United Nations General Assembly. Transforming our world: the 2030 Agenda for Sustainable Development. 2015.

29. Willett W, Rockström J, Loken B, Springmann M, Lang T, Vermeulen S, et al. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. The Lancet. 2019;393(10170):447-92.

30. Mertens E, Kuijsten A, Kanellopoulos A, Dofková M, Mistura L, D'Addezio L, et al. Improving health and carbon footprints of European diets using a benchmarking approach. Public Health Nutr. 2020:1-11.

31. den Braver NR, Kok JG, Mackenbach JD, Rutter H, Oppert J-M, Compernolle S, et al. Neighbourhood drivability: environmental and individual characteristics associated with car use across Europe. International Journal of Behavioral Nutrition and Physical Activity. 2020;17(1):8.

32. Swinburn BA, Kraak VI, Allender S, Atkins VJ, Baker PI, Bogard JR, et al. The Global Syndemic of Obesity, Undernutrition, and Climate Change: Lancet Commission report. The Lancet. 2019;393(10173):791-846.

33. Costello A, Abbas M, Allen A, Ball S, Bell S, Bellamy R, et al. Managing the health effects of climate change: Lancet and University College London Institute for Global Health Commission. The Lancet. 2009;373(9676):1693-733.

34. Marteau TM, White M, Rutter H, Petticrew M, Mytton OT, McGowan JG, et al. Increasing healthy life expectancy equitably in England by 5 years by 2035: could it be achieved? Lancet. 2019;393(10191):2571-3.

35. Petticrew M, Katikireddi SV, Knai C, Cassidy R, Maani Hessari N, Thomas J, et al. 'Nothing can be done until everything is done': the use of complexity arguments by food, beverage, alcohol and gambling industries. Journal of epidemiology and community health. 2017;71(11):1078-83.

36. Kickbusch I, Allen L, Franz C. The commercial determinants of health. The Lancet Global Health. 2016;4(12):e895-e6.

37. Eykelenboom M, van Stralen MM, Olthof MR, Renders CM, Steenhuis IH, Consortium PEN. Public acceptability of a sugar-sweetened beverage tax and its associated factors in the Netherlands. Public Health Nutr. 2020:1-11.

38. Reynolds JP, Pilling M, Marteau TM. Communicating quantitative evidence of policy effectiveness and support for the policy: Three experimental studies. Soc Sci Med. 2018;218:1-12.

39. Nationaal Preventieakkoord. Naar een gezonder Nederland. 2018.

40. RIVM. Quickscan mogelijke impact Nationaal Preventieakkoord. Bilthoven: 2018.

SUMMARY

Non-communicable diseases, including cardiovascular diseases, cancers, respiratory diseases, and diabetes, are the leading contributor to mortality worldwide. The majority of deaths and incidence of chronic diseases are related to tobacco use, unhealthy diets, alcohol consumption and physical inactivity. Creating supportive environments that facilitate health-related behaviours may substantially improve the health of the population, and may also be of importance to reduce inequalities in health. Most environmental interventions are not under control of researchers, and therefore the evidence about their effectiveness is unlikely to be derived solely from randomised controlled trials (RCTs). Evaluation of interventions by means of so-called "natural experiments" that capitalise on changes in real-world settings is important to strengthen the evidence base. This thesis aimed to identify and evaluate local interventions to improve health-related behaviours in the population (part I, chapter 2 to 6), and to elaborate on the methodological approaches when studying natural experiments (part II, chapter 7 to 9).

In **chapter 2**, we studied the introduction of dedicated physical activity spaces on outdoor play and sedentary behaviour in school-aged children. The physical activity spaces were built in deprived neighbourhoods which host relatively more persons from lower socioeconomic and ethnic minority groups; we also evaluated if effects were similar for children from various backgrounds. Children participating in the Generation R study at the age of 6 and 10 years were included. As if it was a trial, we only included children that did not live close to a physical activity space at baseline. Over time, 18 new physical activity spaces were built across the city. We estimated changes in outdoor play time and sedentary behaviour time for children who were living closer to the nearest dedicated physical activity space, as compared to children without changes in distance, using a fixed-effect approach. The introduction of a dedicated physical activity space around home did not increase time spent playing outdoors or decrease watching TV or gaming. Although not statistically significant, stratified analyses suggested that reducing the distance to the nearest physical activity space increased outdoor play for children from more socioeconomically disadvantaged families. Compensation may have taken place, since we also found a non-significant increase in sedentary behaviours in this group of children. Findings suggest that modifying the neighbourhood environment to facilitate outdoor play

is important for children from socially disadvantaged families, without decreasing sedentary behaviours.

In **chapter 3**, we evaluated socioeconomic inequalities in exposure to the food environment and subsequent changes in body composition measures in children. We included children participating in the Generation R study with at least two measures of body composition between the age of 4 and 14 years. At each time-point, we estimated the relative and absolute exposure of fast food outlets, and the healthiness of the food environment around home. Fixed-effect models were used to study changes in the food environment and related changes in body composition measures within children. At all time-points, children from lower educated mothers were exposed to more fast food outlets, and an unhealthier food environment, as compared to children from higher educated mothers. Children from lower educated mothers were also at higher risk that the food environment unfavourably changed over time by offering more fast food outlets and other unhealthy food outlets. Changes in the food environment were not associated with changes in body composition. For children from lower educated mothers without fast food exposure during the first measurement round, the introduction of fast food was associated with small increases in body-mass index. Our findings showed that inequalities in exposure to the food environment have widened over time. Within an already poor food environment, a further increase in fast food outlets or an unhealthier food environment had no additional impact on children's body composition.

In **chapter 4**, we assessed whether hosting international elite sport events has increased recreational sport participation in the population. In 2010, the city of Rotterdam introduced policies aimed at targeting the general population, and we evaluated if this shift in focus contributed to larger increases in sport participation for events organised in this period. Data on sport-specific participation in the general population was obtained from a biannual cross-sectional survey conducted between 1995 and 2017. We evaluated if elite sport events significantly interrupted the underlying time trend in the percentage of people participating in that specific sport. Ten events were included. Findings suggested that the Tour de France Grand Départ, the World Championship Table tennis and the World Championship Gymnastics were followed by an increase in sport-specific participation the year after the event was organised, whereas the World Championship Korfball was followed by a decrease.

Only the increase in the number of recreational cyclists in the population sustained during the 5 years after the event was hosted. The pooled effect of the ten events, however, did not show that sport-specific participation changed. Significant group differences by period that the events took place were found. Subgroup analyses suggested that only for events organised in the period with the more explicit aim to target recreational sport participation, an increase in sport-specific participation in the host population was observed. No group differences were found between elite events with lower or higher number of visitors, of between events organised in the city centre or in stadiums in outskirts. This study provided some evidence that a city with a clear ambition to stimulate recreational sports participation by organising elite sport events may be able to do so. Not all events were followed by an increase in sports participation, therefore a careful selection of the type of sports activities being promoted seems important.

In **chapter 5**, we described a study protocol for a systematic review and meta-analysis on smoke-free policies covering outdoor areas and (semi-)private spaces and the impact on tobacco smoke exposure and respiratory health in children. Findings were presented in **chapter 6**. Through a systematic and comprehensive search, we included 11 studies of which six evaluated smoke-free car policies. Ten effect estimates from four studies were pooled, demonstrating that smoke-free policies substantially reduced tobacco smoke exposure in cars among children. This reduction could potentially decrease the health burden associated with environmental tobacco smoke exposure in children. Other studies evaluated policies in outside places frequented by children, and also reported that tobacco smoke exposure in locations covered by policies declined. There were only limited insights about the differences in effects between socioeconomic groups, therefore we cannot draw conclusions on equity effects.

In **chapter 7**, we conducted a systematic literature review to evaluate the impact of new cycling infrastructure to promote cycling. We included studies that used a wide range of study designs and methods, to assess whether differences in the observed effects were found by methodological approach taken. A broad and systematic search resulted in the inclusion of 31 studies. All studies that reported on infrastructure usage found that the number of cyclists increased. The majority of studies that evaluated changes in cycling behaviour as an outcome found that cycling increased, but effect sizes were considerably smaller as compared to studies that evaluated usage of the infrastructure. Larger changes were found for studies that tested for statistical significance and studies that used subjective measurement methods, compared to studies that did not perform statistical tests and used objective measurement methods. Seven studies also reported on changes in physical activity following the infrastructural intervention, and findings were mixed. We cannot draw conclusion on equity effects, given that only three studies tested for differential effects.

In **chapter 8**, we outlined a framework to explore non-response patterns in repeated cross-sectional surveys, and to evaluate if differential non-response may have increased the risk of non-response bias over time. The framework consists of four steps: 1) investigating differential non-response over time between sociodemographic groups; 2) estimating variation in outcomes over time; 3) comparing weighted and unweighted prevalence estimates of outcomes over time; and 4) comparing associations between sociodemographic determinants and outcomes over time. This framework was applied to a biannual population survey conducted between 1995 and 2017 in which the response rate gradually declined. Information on sociodemographic characteristics of respondents and non-respondents was obtained from the municipal registration database. The health-related behaviours sport participation and watching TV were the outcomes of this study. These outcomes were only available for respondents. Findings suggested that response decreased more rapidly among younger persons and those with a non-Western ethnicity. We did not find evidence that the differential decline in response biased trends in health-related behaviours, or associations between sociodemographic characteristics and health-related behaviours. This framework can be used to understand whether declines in response rates, and larger declines in some subgroups than in others, may have biased trends in health behaviours.

In **chapter 9**, we reflected on the barriers that we encountered in evaluating interventions through natural experiments during PhD trajectories. We have identified three aspects hindering these studies. First, natural experiments itself, but also aspects related to the evaluation, are often unpredictable. We need a scientific and policy environment in which researchers and policymakers jointly identify possibilities for evaluation in early phases of intervention planning. Furthermore, flexible sources of funding are needed to respond to opportunities for evaluation that may suddenly occur. Second, the majority of evaluations rely on already collected data. Their linkage can be challenging. Platforms linking databases from various domains in a safe but accessible way may foster the usage of natural experiments. Without such platforms, multidisciplinary teams are needed ensuring that datasets are linked in a timely manner within ethical and regulatory frameworks. Third, the understanding and acknowledging the value of evidence derived from natural experiments should be improved. Researchers need to be supported to discover and evaluate natural experiments in all its complexities.

In conclusion, the studies in this thesis showed that environmental interventions can considerably improve health-related behaviours in the population. We found some indications that built environment interventions around home may affect children from lower and higher socioeconomic groups in a different way. Ensuring that the home environment makes it easier for children to play outside, and in which children are less exposed to fast food locations, seem promising strategies to reduce socioeconomic inequalities. Natural experiments make it possible to evaluate changes beyond what would be possible in RCTs. Researchers should be involved in early stages of intervention and policy planning to ensure that all opportunities for evaluation are explored. To find the best evidence possible, we should look for opportunities for evaluation of interventions beyond domains, disciplines, and conventional methods. We should strive for knowing more tomorrow than we do today.

SAMENVATTING

Niet-overdraagbare ziekten, zoals cardiovasculaire aandoeningen, kanker, chronische ademhalingsstoornissen en diabetes, zijn de belangrijkste oorzaken van vroegtijdig overlijden wereldwijd. De meeste van deze sterftegevallen en de incidentie van ziekten kan worden toegeschreven aan roken, een ongezond voedingspatroon, alcoholconsumptie en onvoldoende lichaamsbeweging. Het creëren van een omgeving die gezond gedrag faciliteert, kan de gezondheid van de bevolking aanzienlijk verbeteren, en is mogelijk van belang bij het verkleinen van gezondheidsverschillen. Vaak hebben onderzoekers geen controle over omgevingsinterventies, daarom is het onwaarschijnlijk dat bewijs over effectiviteit uitsluitend zal worden verkregen uit gerandomiseerde gecontroleerde onderzoeken. Onderzoekers beroepen zich op het evalueren van zogenaamde “natuurlijke experimenten” waarbij wordt ingespeeld op veranderingen in de maatschappij. Dit proefschrift heeft als doel om lokale volksgezondheidsinterventies, die als doel hebben om gezondheidsgedragingen in de populatie te verbeteren, te identificeren en te evalueren (deel I, hoofdstuk 2 tot en met 6), en om de methodologische vraagstukken die komen kijken bij het evalueren van natuurlijke experimenten te bespreken (deel II, hoofdstuk 7 tot en met 9).

In **hoofdstuk 2** bestudeerden we de introductie van speelvelden op buitenspelen en sedentair gedrag bij schoolgaande kinderen. De speelvelden werden voornamelijk geopend in minder welvarende buurten, waar relatief meer mensen uit lagere sociaaleconomische groepen en etnische minderheden wonen, daarom evalueerden we ook of de effecten vergelijkbaar waren voor kinderen met diverse achtergronden. Kinderen die deelnamen aan de Generation R studie op de leeftijd van 6 en 10 jaar werden geïncludeerd. Net als bij een experiment werden enkel kinderen die bij aanvang van de studie niet in de buurt van een speelveld woonden meegenomen in de analyse. Over tijd werden er, verspreid over de stad, 18 nieuwe speelvelden geopend. Met behulp van een fixed-effect methode evalueerden we de veranderingen in buitenspelen en sedentair gedrag voor kinderen waarbij de afstand tot het dichtstbijzijnde speelveld kleiner werd, in vergelijking met kinderen zonder veranderingen in afstand. Uit de resultaten blijkt dat kinderen niet meer buitenspelen of minder TV kijken en gamen als er een speelveld wordt geopend rondom huis. Hoewel niet statistisch significant, vonden we aanwijzingen dat het verkleinen van de afstand tot het dichtstbijzijnde speelveld ervoor zorgt dat kinderen uit minder welvarende gezinnen meer buitenspelen. Mogelijk heeft er compensatie van gedrag plaatsgevonden, aangezien de resultaten ook suggereren dat sedentair gedrag is toegenomen in deze groep. De bevindingen brengen naar voren dat het creëren van een buurt waarin buitenspelen wordt gefaciliteerd van belang is voor kinderen uit minder welvarende gezinnen, zonder dat het sedentaire gedrag vermindert.

In **hoofdstuk 3** evalueerden we sociaaleconomische ongelijkheid in blootstelling aan de voedselomgeving, en de daaruit voorkomende veranderingen in lichaamssamenstelling bij kinderen. We includeerden kinderen die deelnemen aan de Generation R studie met ten minste twee metingen van lichaamssamenstelling tussen de leeftijd 4 en 14 jaar. Op elk meetmoment bepaalden we de relatieve en absolute blootstelling aan fastfood locaties, en de gezondheidsscore van alle voedselaanbieders rondom huis. Een fixed-effect methode werd gebruikt om veranderingen in de voedselomgeving en gerelateerde veranderingen in lichaamssamenstelling van kinderen te schatten. Resultaten laten zien dat kinderen van lager opgeleide moeders vaker worden blootgesteld aan fastfood locaties en ongezonde voedselaanbieders dan kinderen van hoger opgeleide moeders. Kinderen van lager opgeleide moeders lopen ook een groter risico dat de voedselomgeving in de loop van de tijd ongunstig verandert, en dat het aanbod van fastfood en andere ongezonde eetgelegenheden toeneemt. Veranderingen in de voedselomgeving waren niet geassocieerd met veranderingen in de lichaamssamenstelling. Voor kinderen van lager opgeleide moeders die tijdens de eerste meetronde niet in de nabijheid van fastfood locaties woonden, ging de introductie van fastfood locaties gepaard met een kleine stijging van de body-mass index. Onze bevindingen laten zien dat ongelijkheid in blootstelling van de voedselomgeving in de loop der tijd is toegenomen. In een voedselomgeving die zich kenmerkt door een groot aanbod van ongezonde eetgelegenheden, heeft een verdere toename van fastfood locaties of een ongezonde voedselaanbieders geen extra gevolgen op de lichaamssamenstelling van kinderen.

In **hoofdstuk 4** hebben we onderzocht of het organiseren van internationale topsportevenementen zorgt voor een toename in sportdeelname in de bevolking. In 2010 heeft de Gemeente Rotterdam zich specifiek als doel gesteld om sportdeelname te stimuleren door middel van sportevenementen, dus evalueerden we of deze verschuiving in focus bijdroeg aan een grotere toename van de sportdeelname voor evenementen die in deze periode werden georganiseerd. Gegevens over sportdeelname in de bevolking werd verkregen uit een tweejaarlijkse monitor, uitgevoerd tussen 1995 en 2017. We evalueerden of evenementen de onderliggende tijdstrend, in het percentage van mensen dat aan die specifieke sport deelneemt, onderbreken. Tien evenementen werden geïncludeerd. Onze bevindingen laten zien dat de Grand Départ van de Tour de France, het wereldkampioenschap tafeltennis en het wereldkampioenschap turnen werden gevolgd door een toename in sportdeelname, één jaar nadat het evenement was georganiseerd. Het wereldkampioenschap korfbal werd gevolgd door een afname. Alleen het aantal recreatieve fietsers bleef hoger gedurende de vijf jaar na het evenement. Het gemiddelde effect van de tien evenementen tezamen suggereerde dat de sport-specifieke deelname niet veranderde. Significante groepsverschillen werden gevonden voor de periode waarin het evenement plaatsvond. Subgroep analyse suggereerde dat alleen voor evenementen die in de periode werden georganiseerd met het expliciete doel om recreatieve sportdeelname te stimuleren werden gevolgd door een toename in sportdeelname. Er werden geen groepsverschillen gevonden voor het aantal bezoekers van het evenement en de locatie van het evenement. In deze studie vonden we enig bewijs dat een stad met een duidelijke ambitie om sportdeelname te stimuleren via het organiseren van topsportevenementen, de sportdeelname kan verhogen. Een zorgvuldige selectie van het type sportactiviteit dat wordt gepromoot lijkt belangrijk, aangezien niet alle evenementen werden gevolgd door een toename in sportdeelname.

In **hoofdstuk 5** hebben we een studieprotocol beschreven voor een systematische literatuurstudie en een meta-analyse van rookvrije wetgevingen die betrekking hebben op buitenruimten en semi-private ruimten, en de impact ervan op de blootstelling aan tabaksrook en luchtwegaandoeningen bij kinderen. De bevindingen hiervan worden gepresenteerd in **hoofdstuk 6**. Door middel van een systematische en uitgebreide zoekactie hebben we 11 onderzoeken geïncludeerd, waarvan zes onderzoeken de gevolgen van rookvrije wetgeving specifiek gericht op roken in auto's in de aanwezigheid van kinderen evalueerden. Tien effectschattingen van vier onderzoeken werden samengevoegd, waaruit blijkt dat de invoering van rookvrije wetgeving de blootstelling aan tabaksrook in auto's bij kinderen aanzienlijk vermindert. Zulke wetgeving kan een belangrijke rol spelen in het verminderen van de negatieve gezondheidseffecten als gevolg van blootstelling aan tabaksrook.

Andere onderzoeken evalueerden rookvrije wetgeving gericht op buitenruimtes waar kinderen frequent komen. Zij rapporteerden ook dat de blootstelling aan tabaksrook afnam op plaatsen die door het nieuwe beleid rookvrij werden gemaakt. Er zijn vooralsnog te weinig inzichten verkregen over de verschillen in effecten tussen sociaaleconomische groepen.

In **hoofdstuk 7** hebben we een systematische literatuurstudie uitgevoerd om de impact van infrastructurele interventies, gericht op fietsen, te evalueren. We zochten naar studies met een grote diversiteit aan onderzoeksopzetten en methodes, om vervolgens te beoordelen of er verschillen in de waargenomen effecten werden gevonden naar methodologische aanpak. De systematische en uitgebreide zoekactie resulteerde in 31 onderzoeken. Alle onderzoeken die als uitkomstmaat rapporteerden over het gebruik van de nieuwe fiets infrastructuur lieten zien dat het aantal fietsers toenam. De meeste onderzoeken die als uitkomst veranderingen in fietsgedrag rapporteerden vonden dat het fietsen toenam, maar de effectgroottes waren aanzienlijk kleiner in vergelijking met onderzoeken die het gebruik van de infrastructuur evalueerden. Er werden grotere veranderingen gevonden voor onderzoeken die testten op statistische significantie en onderzoeken die subjectieve meetmethoden gebruikten, vergeleken met onderzoeken die geen statistische tests uitvoerden en objectieve meetmethoden gebruikten. Zeven onderzoeken rapporteerden over veranderingen in lichaamsbeweging als gevolg van de nieuwe fietspaden, en de bevindingen waren gemengd. We kunnen geen conclusies trekken over verschillen in effecten tussen sociaaleconomische groepen, aangezien slechts drie onderzoeken rapporteerden over verschillen in effecten naar sociaaleconomische status.

In **hoofdstuk 8** hebben we een raamwerk geschetst om verschillen in respons in cross-sectionele enquêtes te onderzoeken, en om te evalueren of de verandering in respons het risico op non-respons bias in de loop van de tijd kan hebben vergroot. Het raamwerk bestaat uit vier stappen: 1) het evalueren van differentiële non-respons over de tijd tussen sociaaldemografische groepen; 2) het schatten van variatie in uitkomsten over de tijd; 3) het vergelijken van gewogen en ongewogen prevalentieschattingen van uitkomsten over de tijd; en 4) het vergelijken van associaties tussen sociaaldemografische determinanten en uitkomsten over de tijd. Het raamwerk is toegepast op een tweejaarlijkse monitor uitgevoerd tussen 1995 en 2017 waarbij de respons geleidelijk afnam. Informatie over sociaaldemografische kenmerken van respondenten en niet-respondenten is verkregen uit de gemeentelijke basisadministratie. De gezondheidsgedragingen *sportdeelname* en *TV kijken* waren de uitkomstmaten van dit onderzoek. Deze uitkomstmaten waren enkel beschikbaar voor respondenten. Bevindingen suggereerden dat de respons sneller afnam onder jongeren en mensen met een niet-Westerse achtergrond. We vonden geen aanwijzingen dat het verschil in afname in respons bias resulteerde in trends in gezondheidsgedragingen of associaties tussen sociaaldemografische kenmerken en gezondheidsgedragingen. Het raamwerk kan worden gebruikt om inzicht te verkrijgen of de daling in respons, en een snellere daling in bepaalde groepen, bias kan introduceren in de inzichten die zijn verkregen uit cross-sectionele enquêtes.

In **hoofdstuk 9** reflecteren we op de barrières die we tegenkwamen bij het evalueren van interventies door middel van natuurlijke experimenten gedurende promotietrajecten. We identificeerden drie aspecten die deze studies bemoeilijkten. Allereest zijn natuurlijke experimenten vaak onvoorspelbaar, alsook aspecten die verband houden met de evaluatie. We hebben een wetenschappelijke en beleidsomgeving nodig waarin onderzoekers en beleidsmakers gezamenlijk optrekken en tijdig mogelijkheden identificeren voor evaluatie. Daarnaast zijn flexibele bronnen van financiering nodig om in te spelen op kansen die zich plotseling kunnen voordoen. Ten tweede zijn de meeste evaluaties gebaseerd op reeds verzamelde gegevens, maar het koppelen van verschillende datasets is vaak een uitdaging. Dataplatforms die op een veilige maar toegankelijke manier databases uit verschillende domeinen koppelen, kunnen ertoe bijdragen dat natuurlijke experimenten vaker worden geëvalueerd. Zonder dergelijke platforms zijn multidisciplinaire teams nodig, zodat datasets tijdig kunnen worden gekoppeld binnen de geldende wetgeving. Als laatste is het belangrijk om de waarde te erkennen van studies die inspelen op natuurlijke experimenten. Onderzoekers moeten worden ondersteund om natuurlijke experimenten - in al hun complexiteit - te ontdekken en te evalueren.

Samenvattend laten de onderzoeken in dit proefschrift zien dat omgevingsinterventies de gezondheidsgedragingen in de populatie aanzienlijk kunnen verbeteren. We vonden enkele aanwijzingen dat de gevolgen van interventies die ingrijpen op de leefomgeving mogelijk verschillen voor kinderen uit lagere en hogere sociaaleconomische groepen. Het creëren van een omgeving die het voor kinderen aantrekkelijk maakt om buiten te spelen, en waarin kinderen minder worden blootgesteld aan fastfood locaties, lijken veelbelovende strategieën om gezondheidsverschillen in de bevolking te reduceren. Natuurlijke experimenten maken het mogelijk om interventies te evalueren, die niet door middel van gerandomiseerde gecontroleerde onderzoeken kunnen worden onderzocht. Het is raadzaam om onderzoekers al in de beginfasen te betrekken bij interventie en beleidsplanning zodat alle mogelijkheden voor evaluatie worden verkend. Het is van groot belang om te zoeken naar mogelijkheden buiten de bekende domeinen, disciplines en conventionele methoden. We moeten er naar streven dat we morgen meer weten dan vandaag.

ABOUT THE AUTHOR

Famke Mölenberg was born in Zutphen, the Netherlands, on the 8th of May in 1992. After finishing her bachelor's degree in Biomedical Sciences in 2013, she obtained her master's degree in Nutrition and Health at Wageningen University in 2016. Her graduation internship took place at the Nestlé Research Center in Switzerland. After graduating, she worked for 10 months at Wageningen University as a junior researcher. From January 2017 to December 2020, she worked as a PhD candidate at the Department of Public Health at the Erasmus MC in Rotterdam. She conducted her research within the academic workplace CEPHIR, and was linked to the Research Department of the City of Rotterdam through a hospitality agreement. In 2019, Famke also obtained a post-initial master's degree in Health Sciences at the Netherlands Institute for Health Sciences (NIHES). During her PhD, she worked as a visiting scholar at the MRC Epidemiology Unit at the University of Cambridge, United Kingdom. In January 2021, Famke has started working as a senior researcher at the department of Public Health at the Erasmus MC, and has a hospitality agreement with the Department of Sports, Nature and Recreation of the City of Rotterdam.

Famke Mölenberg is op 8 mei 1992 geboren te Zutphen. Na het afronden van de bachelor Biomedische Wetenschappen in 2013 behaalde zij haar mastergraad Voeding en Gezondheid aan de Wageningen Universiteit in 2016. Haar afstudeerstage vond plaats bij het onderzoekscentrum van Nestlé in Zwitserland. Na haar afstuderen werkte ze voor 10 maanden bij de Wageningen Universiteit als junior onderzoeker. Van januari 2017 tot en met december 2020 was zij als promovendus werkzaam bij de afdeling Maatschappelijke Gezondheidszorg van het Erasmus MC in Rotterdam. Ze verrichte haar onderzoek binnen de academische werkplaats Publieke Gezondheid CEPHIR, en was via een gastvrijheidsovereenkomst verbonden aan de afdeling Onderzoek van de Gemeente Rotterdam. In 2019 behaalde Famke ook een postinitiële mastergraad in Health Sciences bij the Netherlands Institute for Health Sciences (NIHES). Als onderdeel van haar promotie deed ze onderzoek bij de afdeling MRC Epidemiology aan de Universiteit van Cambridge, Verenigd Koninkrijk. Sinds januari 2021 is Famke werkzaam als senior onderzoeker bij de afdeling Maatschappelijke Gezondheidszorg van het Erasmus MC, en heeft ze een gastvrijheidsovereenkomst bij de afdeling Sport, Natuur en Recreatie van de Gemeente Rotterdam.

LIST OF PUBLICATIONS

Publications part of this thesis

Radó MK*, **Mölenberg FJM***, Westenberg LEH, Sheikh A, Millett C, Burdorf A, van Lenthe FJ, Been JV. Impact of smoke-free policies in outdoor areas and (semi-)private places on children's tobacco smoke exposure and respiratory health: a systematic review and meta-analysis. The Lancet Public Health 2021. *In press.*

* Both authors contributed equally

Mölenberg FJM, de Vries C, Burdorf A, van Lenthe FJ. A framework for exploring non-response patterns over time in health surveys. BMC Medical Research Methodology 2021; 21(1):37.

Mölenberg FJM, de Waart F, Burdorf A, van Lenthe FJ. Hosting elite sport events to target recreational sport participation: an interrupted time series analysis. International Journal of Sport Policy and Politics 2020;12(4):531-543.

Radó MK, **Mölenberg FJM**, Sheikh A, Millett C, Bramer WM, Burdorf A, van Lenthe FJ, Been JV. Impact of expanding smoke-free policies beyond enclosed public places and workplaces on children's tobacco smoke exposure and respiratory health: protocol for a systematic review and meta-analysis. BMJ Open 2020;10(10):e038234.

Mölenberg FJM, Panter J, Burdorf A, van Lenthe FJ. A systematic review of the effect of infrastructural interventions to promote cycling: strengthening causal inference from observational data. International Journal of Behavioral Nutrition and Physical Activity 2019;16(1):93.

Mölenberg FJM, Noordzij JM, Burdorf A, van Lenthe FJ. New physical activity spaces in deprived neighborhoods: Does it change outdoor play and sedentary behavior? A natural experiment. Health & Place 2019;58:102151.

Mölenberg FJM, Mackenbach JD, Poelman MP, Santos S, Burdorf A, van Lenthe FJ. Socioeconomic inequalities in the food environment and body composition among school-aged children: a fixed-effects analysis. *Submitted.*

Mölenberg FJM, Vargas Lopes F. The paradox of getting control over natural experiments. *Submitted*

Other publications

Boderie NW, **Mölenberg FJM,** Sheikh A, Bramer WM, Burdorf A, van Lenthe FJ, Been JV. Assessing public support for extending smoke-free policies beyond enclosed public places and workplaces: protocol for a systematic review and meta-analysis. BMJ Open 2021; 11(2):e040167.

Mölenberg FJM, Beenackers MA, Mackenbach JD, Burdorf A, van Lenthe FJ. Is Rotterdam een fastfoodparadijs? De voedselomgeving van 2004 tot 2018. Rotterdam: Erasmus MC, 2019.

Mölenberg FJM, de Goede J, Wanders A, Zock PL, Kromhout D, Geleijnse JM. Dietary fatty acid intake after myocardial infarction: a theoretical substitution analysis of the Alpha Omega Cohort. American Journal of Clinical Nutrition 2017;106:895-901.

Van Dongen LH, **Mölenberg FJM**, Soedamah-Muthu SS, Kromhout D, Geleijnse JM. Coffee consumption after myocardial infarction and risk of cardiovascular mortality: a prospective analysis in the Alpha Omega Cohort. American Journal of Clinical Nutrition 2017;106:1113-1120.

Combet E, Vlassopoulos A, **Mölenberg FJM**, et al. Testing the Capacity of a Multi-Nutrient Profiling System to Guide Food and Beverage Reformulation: Results from Five National Food Composition Databases. Nutrients 2017;9(4):406.

Masset G, Mathias KC, Vlassopoulos A, **Mölenberg FJM**, Lehmann U, Gibney M, Drewnowski A. Modeled dietary impact of pizza reformulations in US children and adolescents. PLoS ONE 2016;11(10):e0164197.

Gijsbers L, **Mölenberg FJM**, Bakker SJL, Geleijnse JM. Potassium supplementation and heart rate: A meta-analysis of randomized controlled trials. Nutrition, Metabolism and Cardiovascular Diseases 2016;26:674-682.

Smit MS, Raat H, **Mölenberg FJM**, Wolfers MEG, Bannink R, Jansen W. Study protocol for the evaluation of long-term effects of the school-based obesity prevention program Lekker Fit! ('Enjoy being Fit'): a retrospective, controlled design. *Submitted.*

Enthoven CA, **Mölenberg FJM**, Tideman WL, Polling JR, Labrecque JA, Raat H, van Lenthe FJ, Klaver CW. Physical activity spaces not effective against socioeconomic inequalities in myopia incidence. The Generation R study. *Submitted.*

PHD PORTFOLIO

PhD Student: Famke J.M. Mölenberg
University: Erasmus University Rotterdam
Faculty: Erasmus MC
Department: Public Health
PhD period: 2017 – 2020
Promotors: Prof.dr. F.J. van Lenthe & Prof.dr.ir. A. Burdorf

	Year	Workload (ECTs)
1. PhD Training		
General academic training		
Master in Health Sciences, NIHES, Erasmus MC	2017-2019	70
Workshop Scientific Integrity, Erasmus MC	2019	0.7
Seminars, workshops and masterclasses		
Research seminars Department of Public Health, Erasmus MC	2017-2020	3.0
Section meetings Social Epidemiology, Erasmus MC	2017-2020	1.0
CEPHIR seminars, Erasmus MC/City of Rotterdam	2017-2020	1.0
NAV Workshop: Presenting in an online society	2020	0.7
KNAW Masterclass: Interdisciplinary perspectives on the food environment	2019	0.7
STeLA Workshop: Science and Technology Leadership	2019	0.7
Rotterdam Science Festival, City of Rotterdam	2018, 2019	1.4
KNAW Workshop: Intergenerational transmission of health inequalities	2018	1.4
KNAW Symposium: Health inequalities: an interdisciplinary discussion of socioeconomic status, health and causality	2018	0.7
Congress of City Makers, Architecture Institute Rotterdam	2018	0.7

	Year	Workload (ECTs)
(Poster) Presentations and Conferences		
Annual meeting of the Dutch Epidemiology Society (WEON)	2017, 2019	1.4
Annual meeting of the International Society of Behavioural Nutrition and Physical Activity (ISBNPA)	2018, 2019	2.8
Annual meeting of the Erasmus MC Health theme: Healthy Science Days	2019	0.7
Annual meeting of Dutch and Flemish Sociologists (DvdS)	2018	0.7
Keynote lecture and panel member Science Hotel, EUR	2018	0.7
2. Teaching activities		
Workshops for researchers and policymakers		
Rotterdam Science Festival, City of Rotterdam: Food environments	2019	1.0
Dutch Statistics and Research conference (VSO): Natural experiments	2018	1.0
Rotterdam Science Festival, City of Rotterdam: Natural experiments	2018	1.0
Workshop for the public		
Science Open, EUR: Healthy Rotterdam	2018	1.0
Lecturing		
Summer Campus 010, 14-16 year old children	2020	1.0
Mortality statistics, 3rd year medical students, Erasmus MC	2018-2020	3.0
Data expedition, minor Public Health, Erasmus MC/EUR	2018, 2019	2.0
Kids College, 10-12 year old children	2018	1.0

	Year	Workload (ECTs)
Supervision		
Community projects, 3rd year medical students Erasmus MC	2018-2020	6.0
Community projects, minor Public Health, Erasmus MC/EUR	2019, 2020	4.0
Thesis Anouk Mesch, bachelor University College, EUR	2018	3.0
Thesis Wendel Peters, master Nutrition and Health, Wageningen University	2018	3.0
Examination		
Bachelor Essays, 3rd year medical students, Erasmus MC	2020	1.0
Final examination, minor Public Health, Erasmus MC/EUR	2018	1.0
3. Other activities		
Research collaborations		
Collaboration with the City of Rotterdam (hospitality agreement)	2018-2020	
10-day research visit at MRC Epidemiology Unit, Cambridge University (UK)	2018	
Committees		
Chair of the Taskforce Junior Epidemiologist of the Dutch Epidemiology Society	2019-2020	
Member of the organising committee of an international conference as part of the 50-year anniversary of the department (postponed)	2019-2020	
Ergo coach of the Department of Public Health	2018-2020	
Member of the social committee of the Department of Public Health	2018-2019	
Member of the student panel Master in Health Sciences, NIHES	2017-2018	

	Year	Workload (ECTs)
Societal impact		
Artikel Foodlog: *De stad op gezond gewicht: pak de omgeving aan*	2020	
Interview Amazing Erasmus MC: *Gezond leven in een fastfoodparadijs*	2020	
Interview Gezond010: *CEPHIR: voedselaanbod Rotterdam steeds ongezonder*	2020	
Interview NRC Rotterdam bijlage: *Speelveldjes voor sommige kinderen goed*	2020	
Interview EUR news bulletin: *Hoe maak je een buurt gezond?*	2018	
Awards		
Best thesis of the year 2019, NIHES	2019	
Societal impact award 2018 of the Department of Public Health	2018	
Acquisition		
Erasmus Trustfonds call for corona related research projects (granted)	2020	
Collaborative agreement with the Department of Sport, Nature and Recreation, City of Rotterdam (granted)	2020	
Joint Programming Initiative a Healthy Diet for a Healthy Life call (not granted)	2020	
ZonMW call Gender en Preventie (not granted)	2020	
Evaluation municipal health insurance, City of Rotterdam (granted)	2019	
Evaluation long-term effects Lekker Fit!, City of Rotterdam (granted)	2018	
Klein maar fijn: Evaluation food environment, City of Rotterdam (granted)	2018	

DANKWOORD

Een marathon zoals een PhD vraagt om een gedegen begeleiding. Beste Frank en Lex, ik had mij geen beter team kunnen wensen. Het project begon met een jaar lang warmlopen, trainen en routes verkennen. Tijdens deze aanloopfase is de basis gelegd voor wat uiteindelijk dit mooie proefschrift zou worden. Vaak kwam ik terug van een van mijn verkenningstochten met nieuwe ideeën. Als free-runner heb ik altijd de ruimte gekregen om deze ideeën verder uit te werken. Niet alle paden gaven het uitzicht waar we op gehoopt hadden; een aantal paden had onverwacht veel hoogtemeters of liep dood, maar we hebben ook hele mooie routes afgelegd. Daarnaast heb ik de ruimte gekregen om mij binnen het Erasmus MC, maar ook daarbuiten, te ontwikkelen als een veelzijdig atleet. Dank voor het vertrouwen en alle tips and tricks.

Ik wil graag de oplettende lezer waarschuwen dat zowel mijn vriend als leidinggevende dezelfde voornaam hebben. Hoewel ik gedurende mijn proefschrift heb geprobeerd selectie zoveel mogelijk te voorkomen, heeft hier mogelijk toch enige vorm van selectie plaats gevonden.

Naast mijn begeleiders wil ik ook graag de commissieleden danken voor de genomen tijd om mijn proefschrift te beoordelen. In het bijzonder wil ik Marianne Geleijnse bedanken. Na het eerste examen in Wageningen werd ik gescout als talent, en mocht ik in jouw onderzoeksteam mijn thesis schrijven. Tijdens deze periode, en later ook als junior onderzoeker, leerde jij mij de vaardigheden waarmee ik kon uitvliegen. Elly en Nicole, ons epi clubje en trainingsmaatjes van het eerste uur. Het is fijn om samen stappen te maken en te groeien. Ik hoop dat we nog vaak *van gedachten kunnen wisselen* over alles wat ons bezighoudt.

Je beseft pas hoe belangrijk collega's zijn als je maanden achtereen thuis werkt. Alle collega's van MGZ, en in het bijzonder leden van de sectie sociale epidemiologie, dank voor de fijne tijd. Ook wil ik graag het secretariaat bedanken voor de altijd aanwezige ondersteuning. Een speciaal bedankje voor Mariëlle, Karen en Astrid. Het is fijn om te weten dat modderpaden, soms wat gebrekkige communicatie en vermoeide benen erbij horen. Met jullie tips kwam ik heelhuids uit het Kralingse Bos, dank daarvoor.

Graag bedank ik ook deelnemers van de Generation R studie en het vrijetijdsonderzoek van de Gemeente Rotterdam, en alle medewerkers en onderzoekers die hebben bijgedragen aan de datacollectie. Ook wil ik graag iedereen bij de Gemeente bedanken voor het warme ontvangst en de fijne samenwerkingen. Frouwkje, fijn dat ik mocht rondneuzen in jouw keuken. Twee prachtige onderzoeken zijn hieruit voortgekomen: leuk dat ook de komende tijd onze paden zullen blijven kruisen.

It is often said that running is an individual sport, though sparring partners are essential to grow and reach beyond your own limits. Co-authors, thank you for the inspiring discussions. Marti, we co-created the systematic review. Our different backgrounds and experiences worked as a catalyzer. Lauren, you made our dream team complete. Jasper, you are a great mentor and team-builder. Dear Jenna, you introduced me to the field of natural experiments within the public health domain in a very inspiring way. Thank you for the warm welcome in Cambridge. Esther, I admire the environment that you created within your team which made me feel immediately at home. I hope you will walk on socks in the office for the rest of your life.

Lieve Eline, Hanna, Mara, Francisca en Nienke, jullie hebben gezorgd voor de confetti en slingers gedurende mijn promotie. Elk 5 kilometer punt werd uitbundig gevierd, bij het 30 kilometer punt werd groots uitgepakt en gingen we naar Lissabon. Dank hiervoor. Het heeft ervoor gezorgd dat ik met een grote glimlach terugkijk op mijn promotietijd.

Olá Francisca, we found each other (among many other things) in the urge to create societal impact. We explored multiple projects together within the City of Rotterdam. The adventure was worth the (long) run. Dear Nienke, I admire your perseverance and positivity. You never say no to bulky and challenging projects. You are a great event organizer, I'm counting on you for the *bitterballen*! I'm very happy that such lovely persons and colleagues will stand beside me as my paranymphs.

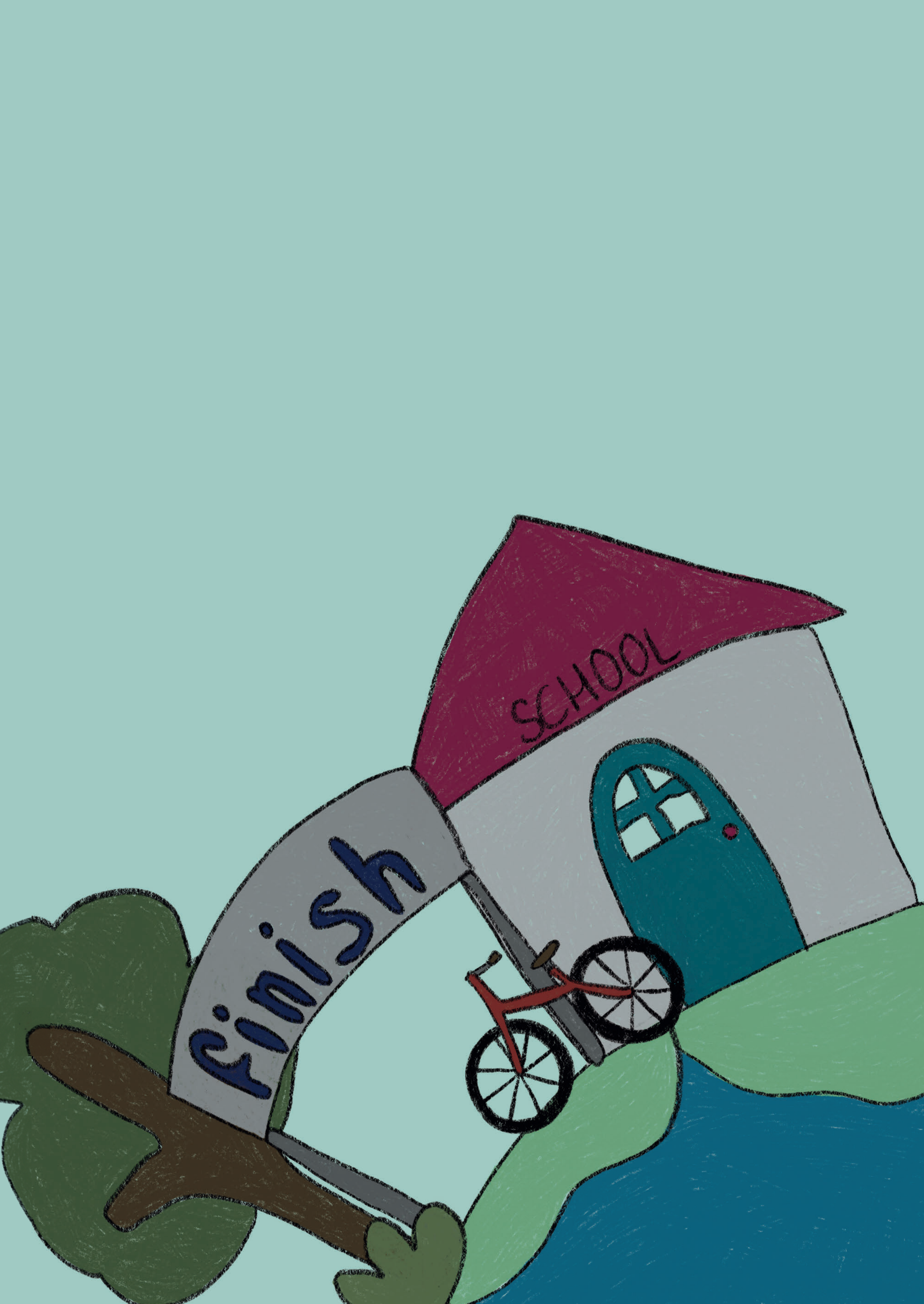
Goudsnikkels, de avonturen met elkaar zijn ontelbaar. Lieve Anouk, mijn steeple maatje, vele hordes hebben we samen genomen. Met jou zou ik in elke trein stappen, en zelfs waarschuwingsbordjes voorbij lopen waarop misschien wel stond dat het beren-broedseizoen is begonnen. Lieve Lieke, jouw promotie en feest waren memorabel. Jij liet zien dat je tijdens een promotie zelf de slingers moet ophangen.

Dank voor alle wijze lessen. Lieve Marjo, tijdens strandwandelingen en bierproeverijen heb jij altijd een spiegel bij je waar ik even in mag reflecteren. Fijn dat we zo vaak konden uitwaaien. Lieve Anne, wat is het heerlijk om af en toe samen te kunnen zagen aan stoelpoten. Ik kijk uit naar jouw promotie, en dan vooral het feest. Zilversnikkels, de Wouters en RJ, fijn dat jullie ons bijstaan in alles wat we doen.

Lieve Jan, wat fijn dat jij al sinds de middelbare school mijn vriendin bent! Myrte en Carlijn, ondanks de afstand zijn jullie altijd erg betrokken bij mijn promotie. Eva en Bjørn, jullie staan altijd klaar om bij te springen als het nodig is. Geen hooibaal te zwaar, geen dahlia knol te zanderig en geen berg te hoog. Ook alle vrienden die niet in het eerste start vak pasten, van loopmaatjes tot studiegenootjes, vriendinnen van vrienden tot vrienden van vriendinnen: lief dat jullie zo vaak hebben gevraagd naar mijn promotie en altijd een fris drankje met bubbels aanboden als het nodig was. Familie Zut-Fun, heerlijk om eens in de zoveel tijd bij elkaar te borrelen. Schoonfamilie, dank voor jullie interesse en in het laatste jaar voor de gezamenlijke koffie pauzes. Het thuiswerken werd zo een stuk draaglijker.

Lieve papa en mama, op mijn geboortekaartje staat een wereldbol afgebeeld met daarbij de tekst *mijn avontuur*. Het avontuur bracht me van Maastricht naar Wageningen, van Lissabon naar Lausanne, van een treinreis door Mongolië naar vulkanen in Zuid-Amerika, naar uiteindelijk een fijn thuis in Zoetermeer. Dank dat jullie mij hierin altijd ondersteunen. Hoe ver weg ook, fijn dat jullie er altijd zijn. Lieve Jissie, jij vertrouwde mij je afstudeerscriptie toe, met als resultaat een onvoldoende. Een smet op mijn onderzoek carrière, waarvoor nogmaals mijn excuses. Je bent de leukste juf op aarde, ik wens alle kinderen een juf Jiske toe. Lieve Rens, wat fijn dat Jissie op jou kan bouwen!

Het startvak, de Erasmusbrug, het rondje op Zuid, de Maassilo, weer de Erasmusbrug, het Kralingse Bos en uiteraard de finish op de Coolsingel. Lieve Lex, overal sta je klaar met aanmoedigende woorden en veel liefde. Jouw fanatisme, ook als ik dodelijke blikken werp, is aanstekelijk. Ik hoop met jou nog vele avonturen te mogen beleven. *We'll never walk alone.*


SCHOOL
finish