Methods for Evaluating Environmental Health Impacts at Different Stages of the Policy Process in Cities

Jaime Benavides[1] · Sebastian T. Rowland[1] · Jenni A. Shearston[1] · Yanelli Nunez[1] · Darby W. Jack[1] ·
Marianthi-Anna Kioumourtzoglou[1]

Abstract

Purpose of Review Evaluating the environmental health impacts of urban policies is critical for developing and implementing policies that lead to more healthy and equitable cities. This article aims to (1) identify research questions commonly used when evaluating the health impacts of urban policies at different stages of the policy process, (2) describe commonly used methods, and (3) discuss challenges, opportunities, and future directions.

Recent Findings In the diagnosis and design stages of the policy process, research questions aim to characterize environmental problems affecting human health and to estimate the potential impacts of new policies. Simulation methods using existing exposure–response information to estimate health impacts predominate at these stages of the policy process. In subsequent stages, e.g., during implementation, research questions aim to understand the actual policy impacts. Simulation methods or observational methods, which rely on experimental data gathered in the study area to assess the effectiveness of the policy, can be applied at these stages. Increasingly, novel techniques fuse both simulation and observational methods to enhance the robustness of impact evaluations assessing implemented policies.

Summary The policy process consists of interdependent stages, from inception to end, but most reviewed studies focus on single stages, neglecting the continuity of the policy life cycle. Studies assessing the health impacts of policies using a multi-stage approach are lacking. Most studies investigate intended impacts of policies; focusing also on unintended impacts may provide a more comprehensive evaluation of policies.

Keywords Healthy cities · Urban policy · Environmental health · Impact evaluation · Policy process

Introduction

A compelling body of evidence has demonstrated the influence of environmental exposures—such as air pollution, noise, heat, and green space—on human health in urban areas. Urban areas can be defined as built environments where large concentrations of the human population live, communicate, and exchange services. Policies are necessary to regulate the level of exposures to these environmental pathways affecting health [1••]. Policies can be defined as government-driven processes happening over time or events occurring at specific time points, which cause changes in urban infrastructure and/or human behavior and, as a consequence, impact environmental pathways and health outcomes at street, street network, neighborhood, city, or metropolitan region levels. These impacts can be direct—when policy and impact take place on the same urban environmental feature (e.g., a congestion pricing scheme that charges a fee to motor vehicles driven in an urban area acts on traffic to reduce congestion [2])—or indirect—when the policy acts on an environmental feature but impacts a different one (e.g., a congestion pricing scheme impacting air pollution [3]). Additionally, impacts can be intended (e.g., low-emission zone impacts on reducing air pollution [4]) or unintended (e.g., gentrification due to greenery enhancement [5]). At a

temporal level, impacts can be short term (e.g., air quality warnings on respiratory health admissions [6]) or long term (e.g., reduced indoor fine particulate matter ($PM_{2.5}$) concentration levels on premature mortality [7]).

Assessing a policy's impacts across different stages of the policy process, from inception to end, is important for achieving progress toward more healthy and equitable cities. The results can be used to raise awareness about the policy's potential or actual efficacy and to inform the decision-making process to modify their design, if and when necessary, toward improving urban health [8–12]. In this article, we reviewed the scientific literature on methods to quantify the urban environmental health impacts of policies at different stages of the policy process given research questions. We note that government entities and other organizations conduct impact evaluation studies that may not be published in scientific journals and may be worth discussing, but we focused on peer-reviewed scientific works in this review. We prioritized articles published in the last 5 years and included some relevant older and classic papers to illustrate specific knowledge that complements the reviewed works. While we focused on impacts in urban settings, we also included some studies at larger scales that illustrate novel approaches not yet applied at the urban scale. Most of the studies we reviewed quantitatively assessed impacts of policies instead of using qualitative information in their evaluations. Given the diversity of health impact evaluation studies, we separate our discussion by policy stage, following the policy process: diagnosis, design, pilot, implementation, operation, and dismantling. In the diagnosis stage, the need for a policy is first identified. Then, the policy is designed and may be piloted, before being implemented. After full implementation, the policy may be in effect for a period of time, defined as the operation stage, and finally, it may be dismantled. During each stage, different goals and objectives, and thus different research questions, may be relevant for health impact evaluation.

Frameworks for Evaluating Environmental Health Impacts of Policies

Several frameworks have been proposed to provide a structure for conceptualizing environmental health impacts of policies [9, 13]. Currently, two of the most prominent frameworks are (a) the framework on the relationship between urban planning, environment, and health proposed by Nieuwenhuijsen [1••] and (b) the framework for accountability research on air pollution and health by the Health Effects Institute (HEI) in collaboration with other organizations and scientists [12, 14]. The first framework links policy, urban environment, and health by taking into account multiple environmental exposures, their determinants (e.g., built environment, personal behavior), and how these exposures affect health. The second framework connects policies aiming to reduce air pollution with health impacts derived by their implementation, adding a feedback loop between the results of the impact evaluation and the policy to inform decision-making about the policy's actual efficacy. Figure 1 shows the first framework, incorporating an impact evaluation layer, representing the focus of this review and providing an analytical perspective of the impacts of the policy process at different stages. The impact evaluation layer is linked to the policy by the feedback (i.e., learning) phase from the second framework. In the learning phase, intended and unintended impacts resulting from the evaluation are communicated to policymakers involved in the policymaking process, who integrate this information and adapt the policy to be more effective in attaining its objectives. We note that policy processes result from a nonlinear complex interplay between politics, regulatory mandates, and bureaucratic processes, among other factors. We provide this framework, along with the different policy process stages, as a conceptual guide to contextualize the scientific works dealing with environmental health impacts of urban policies discussed in this review. Toward that end, we simplify the presentation of the policy process following a linear representation.

Commonly Used Methods for Policy Impact Evaluation

In general, quantitative policy impact evaluation involves predicting or estimating the impact of a policy on one or more outcomes, ranging from urban design features to public health (Fig. 1), e.g., the difference in air pollution exposure(s) under a policy versus a reality without policy. However, for studies examining implemented policies, both scenarios cannot occur at the same time, and for studies examining unimplemented policies, the policies and their impacts are unobservable. Therefore, researchers face a missing data problem—what would the outcome be under the unobserved policy scenario(s)? [15]. To overcome this challenge, researchers can estimate the potential outcomes of unobserved (i.e., counterfactual) scenarios [16]. Broadly, researchers have developed/used methods to facilitate the estimation of potential outcomes using either simulations of policy scenarios or observational evidence of policies that were at least partially implemented; studies may employ one [2, 17] or both strategies [18]. Several studies have reviewed the scientific literature on environmental health impacts of policies in recent years and provided general and methodological recommendations as summarized in Table 1.

Broadly, simulations can be characterized by how the relationship between the policy and the outcome of interest

Fig. 1 Impact evaluation of policies on urban environment and health at different stages of the policy process
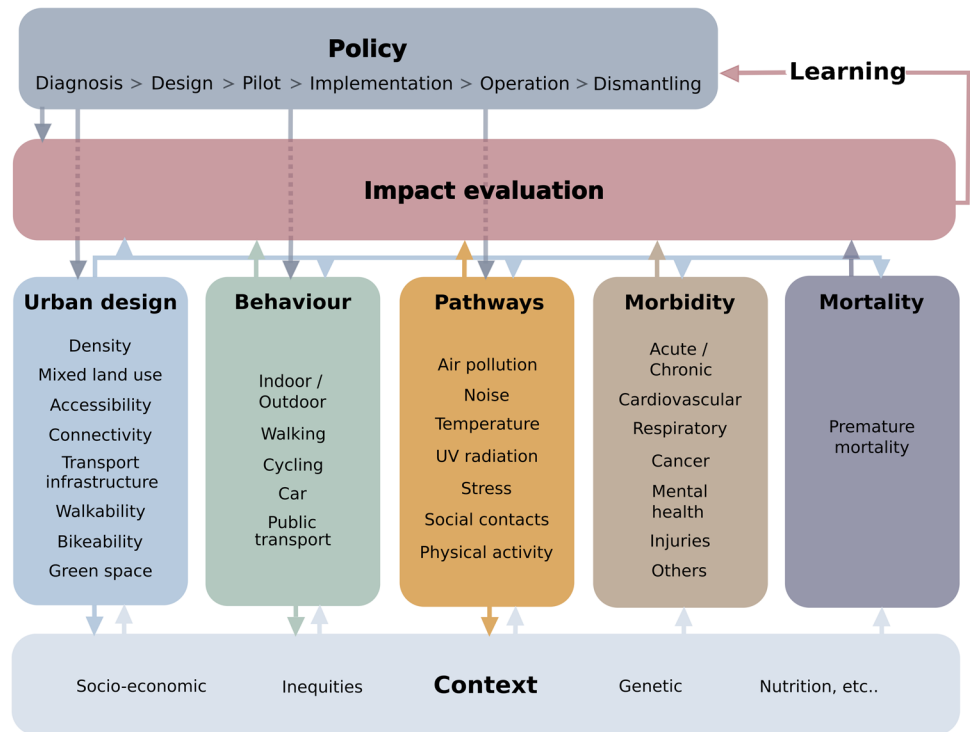

Table 1 Summary of general and methodological recommendations for future studies from recent examples of previously published reviews of policy-related health impact evaluations

Type	Category	Recommendation(s)
General	Climate change	To integrate climate change mitigation and health co-benefits and disbenefits [12, 19]
	Health equity	To disaggregate population into vulnerable groups to characterize inequities [20••]
	Multi-exposure	To move from single exposure models to integrate several environmental exposures [19, 21, 22]
	Policy monitoring	To measure impact over time to inform policy modifications [9, 11]
Methods	Confounding	In observational studies, to use control un-intervened populations and before/after policy comparisons [12, 23]
	Uncertainty	To quantify uncertainty to provide a confidence estimate [21]
	Robustness	To use multiple methods in each study and conduct sensitivity analyses [23, 24]

is modeled; process-based models use externally derived mathematical relationships (e.g., physical laws [25]), whereas data-driven approaches employ statistical and/or machine learning models (e.g., gradient boosting machine [26]) trained on researcher-provided data. Simulation studies start by estimating the policy impact on a single or multiple environmental exposures, which are assumed to be related to and impacted by the policy. For example, Baghestani and colleagues [2] estimated the impact of different congestion pricing scenarios on vehicle volume and speed over New York City, United States (US), with a transportation model and assessed emissions of air pollutants using a vehicle emission model. To isolate the impact of the COVID-19 lockdown on NO_2 concentrations from coincident meteorological changes, Petetin and colleagues

[26] simulated the hypothetical NO_2 concentrations under the same weather conditions, assuming no lockdown, via a machine learning algorithm (gradient boosting machine) trained on historical data from Spanish cities. Researchers can subsequently use exposure–response relationships (ERRs) already published in the literature to translate behavioral or exposure impacts (i.e., difference between scenarios) into anticipated health impacts [27, 28]. A strength of simulations is that researchers can reproduce the state of the urban environment to isolate the effect of the policy [26, 29]. Simulations, thus, are often used during the design stage (Table 2). A major challenge, however, is building models that accurately describe the often-complex relationship between the policy and outcome of interest [30].

Table 2 Main approaches to evaluating policies in cities at different stages. Please see Supplemental Table 1 for a complete view of the reviewed studies

Policy stage	Category	Method(s)	Health	Pathways	Policy[a]	Urban areas	Citations
Diagnosis	Simulation	Land use regression (LUR) model, ERR	Premature mortality	Air pollution		969 European cities	Khomenko et al. (2021) [33••]
		LUR model, noise propagation model, surveys, satellite obs., ERRs		Air pollution, road traffic noise, physical activity, temperature		Vienna, Austria	Khomenko et al. (2020) [34]
		Vehicle emission model, dispersion model, ERR	Premature mortality	Air pollution		Beijing, China	Tong et al. (2020) [35]
Design	Simulation	Transportation model, vehicle emission model		Air pollution	Congestion charge	New York, US	Baghestani et al. (2020) [2]
		Transportation model, surveys, dispersion model, noise propagation model, local temperature model, ERRs	Premature mortality	Air pollution, road traffic noise, physical activity, temperature	Superblocks	Barcelona, Spain	Mueller et al. (2020) [36••]
		Noise propagation model, ERRs	Premature mortality, cardiovascular diseases, and others	Road traffic noise, traffic crashes	Speed limit reduction to 30 km/h	Lausanne, Switzerland	Rossi et al. (2020) [37]
Pilot	Simulation/Observational	Traffic obs., air pollution obs., emission model, dispersion model, ERR	Premature mortality	Air pollution	Congestion charge	Stockholm, Sweden	Johansson et al. (2009) [38]
Implementation	Observational	Longitudinal survey design, regression model		Physical activity	Improve active travel infrastructure	London, United Kingdom (UK)	Aldred et al. (2019) [39]
		Liveability indicators based on spatial and street network analysis		Liveability	Improve walkability, transit access and public open space	4 Australian cities	Lowe et al. (2020) [40••]
		Bayesian point change detection for implementation time lag		Travel patterns	COVID-19 policies on transportation	Seattle, New York, US	Bian et al. (2021) [41]

Table 2 (continued)

Policy stage	Category	Method(s)	Pathways	Health	Policy[a]	Urban areas	Citations
Operation	Simulation	Industrial emission model, dispersion model, ERR	Air pollution	Premature mortality	SO_2 control strategies	Detroit, US	Martenies et al.(2018) [42]
	Observational	Propensity score, difference-in-differences	Temperature	Heat-related hospital admissions	Heat emergency plan	New York, US	Benmarhnia et al.(2019) [17]
	Observational	Interrupted time series	Air pollution	Respiratory diseases	Air quality emergency	Hong Kong, China	Mason et al. (2019) [6]
	Observational	Regression discontinuity	Air pollution		G20 Hangzhou summit plan	Zhejiang, China	Zeng et al. (2020) [43]
	Simulation/Observational	Machine learning, augmented synthetic control	Air pollution	Premature mortality	COVID-19 lockdown	Wuhan, China	Cole et al. (2020) [44••]
	Simulation/Observational	Machine learning	Air pollution		COVID-19 lockdown	Spanish cities	Petetin et al. (2020) [26]
Dismantling	Observational	Propensity score, regression model	Air pollution	Birth outcomes	Historical redlining	California, US	Nardone et al. (2020) [45]

[a]Policy is empty in the diagnosis stage because it is a stage prior to policy development which aims to investigate if an environmental health issue exists.

Observational studies evaluate the impact of already-implemented policies by comparing observed and counterfactual scenarios, which are based on historical data, from a specific spatio-temporal context in which a policy took effect [12]. A commonly used approach is the difference-in-differences design (DiD), which compares the change in an outcome in a group affected by a policy ("treated") to the change among an unaffected group ("untreated"), assuming there are no other group-specific outcome trends [31]. For example, He and colleagues [32] evaluated the impact of COVID-19 lockdowns on air quality by comparing the change in air pollution in locked-down Chinese cities to the change in non-locked-down Chinese cities, which represents the hypothetical change that the locked-down cities would have experienced without lockdown. Interrupted time series (ITS) is another widely used approach in observational studies. In ITS, the trend in outcome in the post-policy period is compared to the trend in outcome in the pre-policy period, assuming that outcome trends pre-policy would have continued in absence of the policy implementation [31]. ITS requires sequential data with a clear trend in the outcome and a well-defined change point that corresponds to the introduction of a policy. For instance, Mason and colleagues [6] used ITS to estimate the impact of implementing an air quality alert warnings program on respiratory-related hospitalizations in Hong Kong, China, by leveraging hospitalization trends from 3 years prior and post implementation. In both DiD and ITS cases, recent methods have been developed to address residual confounding if the assumptions are not met (please see the "Main Challenges and Future Directions" section for examples). Several other methods have also been applied in observational studies (Table 2). Overall, a major advantage of observational studies is the use of historical data from the observed reality. However, disentangling policy-related changes from other time-varying factors remains a challenge and hinders a reliable quantification of the impacts attributed to the policy implementation [12].
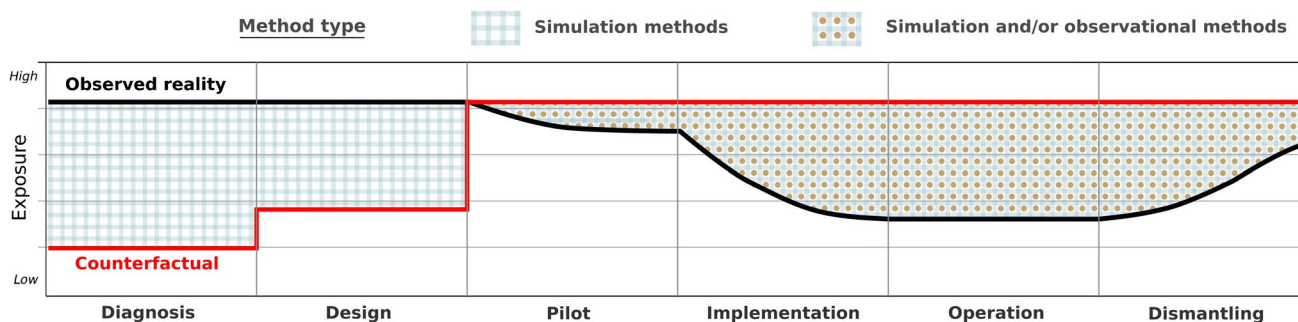
Research Questions by Policy Stage

We have reviewed common methods that researchers use to evaluate health impacts of urban policies. Next, we discuss common research questions that employ these methods to evaluate policies. Figure 2 shows an example of a policy impact evaluation from inception to end. As can be seen in the figure (black and red lines), a shift in typical counterfactuals occurs throughout the policy process: in the diagnosis and design stages, counterfactual exposure levels represent pollution standards and simulated exposure reductions from a hypothetical policy. In contrast, in the pilot, implementation, operation, and dismantling stages, counterfactual exposure levels represent a scenario without policy and may

Fig. 2 Example of observed and counterfactual exposures along with methods to assess health impacts at different stages of the policy process. Simulation methods, represented by light blue-colored squares, are used across all stages of the policy process, while observational methods (brown-colored dots) are used once the policy takes effect. Observed reality (black line) represents exposure measurements that can be observed directly, and are typically higher, as they may prompt the development of a policy. The red line is the counterfactual expo-sure level that is being compared; in the diagnosis phase, it could be the exposure level recommended by government guidelines; in the design phase, it could be the lower level achieved by a hypothetical intervention; after implementation begins, it represents the business-as-usual scenario without the policy. The modeled counterfactual in the design stage might be higher than recommended values because the designed policy may only partially address the environmental health issue

come from actual experimental data or complex simulations. The research question determines which methods and counterfactuals are chosen to evaluate the policy. Below, we discuss the research questions typically used for health impact evaluation in each policy stage and give key examples from the literature.

Diagnosis

Researchers usually identify the need for a policy during the diagnosis stage. However, it is rare that environmental health directly drives urban policies. The environmental health-relevant research questions of this stage are characterized by health impact evaluations that investigate if environmental health issues exist, their extent, and their main causes. Policymakers can apply the results of these analyses to understand if a policy is required to potentially ameliorate environmental health [9]. Often, these studies compared the current observed exposure levels with a counterfactual scenario representing recommended exposure levels (e.g., the WHO guideline for the annual mean $PM_{2.5}$ [33••]). Studies commonly used existing ERRs from the scientific literature to compare health burdens between the observed exposure and the counterfactual recommended scenarios [33••, 34, 46–49]. Common research questions for this policy stage include:

(1a) What is the health burden for exceeding international/national/local exposure guidelines for an environmental pollutant?
(1b) Are the health burdens from exceeding international/national/local exposure guidelines for an environmental pollutant distributed equitably by social factors (e.g., race/ethnicity, economic income, etc.)?

(1c) What are the predominant environmental causes of a health burden or health disparity?

As an example of research questions 1a and 1b, Khomenko and colleagues [34] assessed the impact of various environmental pathways (air pollution, road traffic noise, heat, physical activity) and an urban design feature (green space) on premature mortality in Vienna, Austria. The authors compared the existing premature mortality burden from the observed exposures with the counterfactual mortality burden that would have occurred had air pollution, noise, heat, green space, and physical activity recommendations been met. This study used existing ERRs from the literature and considered health disparities by investigating the distribution of premature mortality by socioeconomic status (SES). For research question 1c, Zhao and colleagues [46] determined the source-specific contribution of various chemicals (e.g., NH_3, SO_2, NO_x, and others) to mortality in the Beijing-Tianjin-Hebei Chinese region. They first estimated the impact of hypothesized emission reductions on population exposure using the extended response surface model, which combines scenarios simulated using chemical transport models with statistical techniques, and then estimated mortality impacts with ERRs.

Design

In the design stage, policymakers plan and potentially test the elements and details of a proposed policy. Health impact evaluations in this stage may estimate the potential health impacts of new policies or compare the potential health impacts of various iterations of a proposed policy (e.g., decreases in asthma hospitalizations after setting a new road toll to \$10 vs. \$25). In this stage, the unobserved

counterfactual represents the theoretical condition(s) that would occur after the policy is implemented (Fig. 2, second panel) and is often compared to the observed exposure levels (i.e., without policy). Similar to the diagnosis stage, most studies use ERRs from the literature to compare health impacts between different exposure scenarios [2, 7, 36••, 37, 50, 51]. Common research questions for this policy stage include:

(2a) What is the potential impact of the proposed policy on exposure levels and health outcomes?
(2b) What are the potential health impacts and tradeoffs of different policy scenarios or objectives?
(2c) What are the costs and benefits of a proposed policy?
(2d) What are the health and climate mitigation co-benefits of a proposed policy?

Research question 2a was the most common among the articles we selected to review; Mueller and colleagues provide a useful example [36••]. They estimated the potential health impact of implementing the Superblock Model—an innovative policy aiming to reclaim public space for people, reduce car dependency, and promote sustainable mobility and active lifestyle—across Barcelona, Spain, in two main steps. First, the authors applied simulation methods to estimate potential impacts on NO_2 concentrations, road traffic noise, green space, heat, and transport-related physical activity. Second, they calculated attributable health impact fractions for premature mortality, life expectancy, and economic impact of the policy. While research question 2b was less frequently observed in the literature, it is an important research question for this policy stage because it can help fine-tune design elements of the policy and identify preferred objectives. Thondoo et al. [50] provide a great example of the latter. In that paper, the authors compare the health and economic impacts of three different transport scenarios (worse, good, ideal) with the current baseline scenario in Port Louis, Mauritius. They used qualitative and quantitative methods to construct the different scenarios, focusing on potential changes in car trips, walking, motorcycle use, and public transport use, and estimated the health and economic impact of each scenario.

Pilot

In the pilot stage, government entities partially implement the policy for testing purposes. The partial implementation can be spatial (i.e., over reduced areas) or temporal (i.e., specific periods), and not all policies are piloted. Experimental data are typically gathered to answer questions regarding the actual impact of the pilot on behavior [52], environmental pathways [3], and health [38] (Fig. 1), occasionally including qualitative information on citizens' perception of the

pilot [53]. Common research questions for this policy stage include:

- (3a) What is the impact of the pilot implementation on environmental pathways and health outcomes?
- (3b) What is the public perception/acceptability of the policy?

Johansson and colleagues' analysis of Stockholm's congestion charge pilot implementation [38] in Sweden is an instructive example of question 3a. The Stockholm congestion charge was implemented at city scale for a 6-month period to assess its efficacy on reducing traffic congestion and air pollution. The authors measured and modeled changes in road traffic as a result of the pilot implementation, and then propagated traffic decreases to air quality changes using a dispersion model. Finally, the authors used ERRs to estimate health benefits [38]. Other evaluations of the Stockholm trial [3, 53] assessed citizen perspectives on acceptability and equity of the policy (question 3b).

Implementation

The implementation policy stage refers to the time period from the pilot stage (if existent) to the full implementation of the policy. This is the period during which the policy is being rolled out. Given that it can take many years for a policy to be fully implemented, studies may evaluate the impact of a policy on behavior [41], environmental pathways [39, 40••], and health outcomes even before full implementation has been completed. The results can then be used to justify continuing the rollout, halt the rollout, or may be used more like that of a pilot study, to tweak the policy as it is being implemented. Common environmental health-relevant research questions for this policy stage include:

- (4a) What is the actual efficacy of the policy on environmental pathways and health outcomes at the current implementation state?
- (4b) What is the level of implementation of the policy targets related to environmental health and how does that level impact health outcomes?

The majority of reviewed studies evaluating policies at this stage focused on question 4a. For example, Aldred and colleagues [39] investigated the impact of an urban design policy on active mobility levels in London, UK, using a longitudinal survey design. The authors defined control and intervention groups based on residential location to distinguish between low- and high-exposure areas. Regression models were then applied to investigate if the policy had an impact on resident behavior, attitude, and perception, adjusting for SES characteristics. Research question

4b directly incorporates implementation into the analysis, but there were far fewer studies that addressed this question. For example, Lowe and colleagues [40••] investigated the capacity of existing policies to equitably attain their livability targets using a set of spatial indicators to assess level of policy implementation in the four largest Australian cities. The authors first conducted a policy review and identified policies whose implementation levels could be assessed using available spatial data. For instance, for policies targeting public open space, they analyzed distance from all residential addresses within each city to public open spaces based on street network analysis. The authors of the study recommended the creation of consistent indicators for healthy city policies, to allow for levels and inequities in policy implementation to be assessed across different geographic locations.

Operation

At this stage, the policy is under full operation and evaluation studies aim to understand its actual impact on environmental pathways [26, 43] and health [6, 17, 42, 44••]. The counterfactual represents what would have happened had the policy not occurred, assuming that everything else continued to be the same. Simulated or measured exposures from the observed reality (i.e., with policy) are compared to counterfactual scenarios (Fig. 2, fifth panel). Common research questions for this policy stage include:

- (5a) What are the impacts of the policy on urban design, behavior, pathways, and/or health?
- (5b) Are the impacts of the policy on exposures or health distributed in an equitable manner?
- (5c) What is the cost-effectiveness of the policy?

Most studies we reviewed addressed research questions 5a and 5b, often together. As an example, we highlight an analysis by Cesaroni and colleagues [54] of the air quality and health effects of two low-emission zones in Rome, Italy. The study used emission and dispersion models combined with existing ERRs from the literature and evaluated the policy impact across SES levels to check for potential social inequalities. For an instructive example of cost-effectiveness analysis (question 5c) that incorporates multiple pathways and health outcomes, we refer the readers to Gu et al. [55], who evaluated the cost-effectiveness of bike lane construction in New York City, US, in two phases. First, they estimated the impact of increasing bike lane miles on bike ridership using regression analysis. In the second phase, the authors assessed the cost-effectiveness of bike lane construction by simulating the injury risk, ridership, physical activity, and air pollution under a no-construction scenario via a Markov model. As health-related effects, the authors considered impacts on risk of injury, ridership, physical activity, and air pollution.

Dismantling

In the dismantling stage, a policy is no longer in effect or has been removed. However, despite no longer being in effect, the policy might still have residual impacts on human health even many years after its deployment and dismantling [9]. It should be noted that not all policies have a dismantling stage. Evaluation studies at this stage aim to investigate the remaining impact of a dismantled policy. A common research question for this stage is:

- (6a) What is the impact of a historic policy on present day urban design, environmental pathways, and health?

For example, the historical redlining classification (a racist mortgage appraisal process that systematically denied loans to people of color) of more than 200 cities in the US during the 1930s was investigated by Nardone and colleagues [45, 56]. The authors evaluated the impact of redlining on current urban design (green space) [56] and health outcomes (birth weight) [45] using propensity score matching and regression models. They found that redlining exacerbated racial segregation and was associated with reduced greenspace and adverse health outcomes [45, 56].

For more examples of studies that evaluate the health impact of policies, at each of the six policy stages, please see Supplemental Table 1.

Main Challenges and Future Directions

This review has described some of the methods and research questions used to evaluate environmental health impacts of urban policies at different stages. Building on this structure and the recommendations from previous reviews (Table 1), we draw on the most recent studies to discuss the main challenges and future directions, exemplified by significant developments addressing these challenges.

We identified several challenges related to the scope of the formulated research questions. A common challenge derived from the life cycle perspective of the policy process in this review is the scarcity of studies assessing the health impacts of policies using a multi-stage approach. Most studies statically investigate impacts at a single stage of the policy, during a specific time frame. Urban infrastructure, population, and behavior in cities change over time [10] and, as a consequence, the impact of a policy is not static. Thus, the results of policy impact evaluations depend on the time frame considered [9]. In addition, impact evaluations would benefit from connecting evaluations of the

impacts at several stages to develop comprehensive evaluations across the policy process. As an illustrative example of this need, Holman and colleagues [4] reviewed studies analyzing the intended impacts of low-emission zones on air quality in European cities. The authors concluded that simulation studies conducted during the design stage of the policy estimated much larger benefits than observed during the operation stage. In this case, connecting both design and operation impact evaluations by updating the design-stage modeling efforts whenever relevant information becomes available (e.g., real-world emission factors from diesel vehicles [30]) may be helpful to refine the estimated impacts of the implemented policy. This iterative process, including the adjustment of policies to reflect the updated knowledge (i.e., learning phase in Fig. 1), represents an essential component of adaptive management [10].

Another challenge is the limited research into unintended impacts of urban policies. Most studies investigated the policy's intended impacts on environmental pathways and human health. Studies also assessing unintended impacts, such as green gentrification [44••] or redistribution of traffic emissions and air pollution levels that may be the consequence of traffic calming policies [57], may provide a more complete analysis of the policy impacts. Unintended impacts can particularly affect environmental justice communities [44••] that may be excluded from the initial design process. Additionally, a common challenge related to health equity is that while a number of studies measure the distribution of health impacts by socioeconomic groups, there has been limited research into root causes of inequities as identified by Buse and colleagues [20••]. These include root causes of environmental health inequities as well as inequities in potential benefits/harms from a policy. Buse and colleagues [20••] concluded that such research can inform how a policy can exacerbate or ameliorate environmental health disparities. We recommend that researchers broaden their research questions to develop comprehensive evaluations of the full policy life course, including unintended impacts.

A common methodological challenge of simulation and observational studies is to assess health impacts due to changes in multiple exposures and environmental pathways. The reviewed simulation studies addressing multiple exposures and pathways give a broad view of the potential impacts in isolation, without including interactions between exposures and pathways, so they should be interpreted with caution (e.g., [27, 37, 46]). Some observational studies are starting to consider multiple exposures in relation to air pollution, as exemplified by Mason and colleagues [6], but still tend to exclude other environmental pathways. Hybrid approaches, using both simulations and observational methods, are a promising approach to account for synergistic effects of multiple pollutants. For example, Nethery and colleagues [18] combined air pollution levels from a process-based atmospheric chemistry model with matching and machine learning methods to estimate the cumulative health impacts attributable to policy (i.e., the 1990 Clean Air Act Amendment) changes on several air pollutant concentrations in the US. Their approach, which assumes that the entire study area is affected by the policy, may be useful for evaluations dealing with policies affecting an entire city or metropolitan area, where the chances of finding an un-intervened population are limited [58].

A second methodological challenge is the need to characterize the spatio-temporal uncertainty associated with exposure assessment and propagate it into the estimated health effects. This challenge increases in complexity with the increasing number of exposures considered in a health impact study. Errors in exposure levels might depend on spatio-temporal conditions (e.g., variation across streets or seasons), which could be wrongly attributed to other factors that vary in a similar manner. Recent works estimating health burden and assessing a policy's efficacy in cities have characterized the confidence of health impacts [33••, 59]. For instance, Khomenko and colleagues [33••] compared current air pollution levels with levels complying with recommended guidelines in 969 European cities and propagated uncertainty estimates from the input variables (e.g., ERRs, exposure levels) to the health impact analysis using Monte Carlo simulations. A complementary robustness test for policy evaluation is to conduct sensitivity analyses aiming to understand how environmental health impacts are affected when certain parameters (e.g., physical activity [27]) are perturbed from a set of potential values.

Another methodological challenge when assessing the impact of existing policies is to build counterfactuals avoiding confounding (i.e., rival explanations). Recent developments in the causal inference field tackle this challenge, allowing more robust impact evaluations. For instance, in both DiD [60, 61] and typical regression models [45], researchers have matched units (e.g., census tracts) on their propensity scores, which capture the probability that a unit would receive the intervention [62]. Researchers can protect against misspecification of the propensity score model by using doubly robust methods, i.e., adjusting for confounders in the main model, after matching by propensity score [61]. Machine learning approaches, such as neural networks and regression trees, can be used to flexibly model the relationship between predictors and treatment likelihood [63]. Traditional DiD involves comparing a treated unit to one or more control units, though none of the control units may be exactly comparable to the treated unit. One approach to loosen the assumptions of the matching is to create synthetic controls, which are a weighted average of the potential controls, weighted by their similarity to the treated unit [64]. Ben-Michael and colleagues [65] recently proposed the ridge augmented synthetic control approach to address

residual bias from imprecise matching on units, which was subsequently applied by Cole and colleagues [44••] to estimate the air quality and health impacts of the COVID-19 lockdown in Wuhan, China.

For policies acting upon reduced spatio-temporal contexts, conducting impact evaluations may be challenging. In these cases, using robust designs and evaluation methods has been useful to investigate small-scale policy effectiveness [66, 67]. For instance, Benton and colleagues [66] conducted an observational study to evaluate the impact of green space enhancements along an urban canal on physical activity and other wellbeing behaviors in Manchester, UK. The authors matched two comparison sites to the intervention site using a five-step process, based on eight physical activity variables at both the site and neighborhood levels. Then, the authors compared intervened and un-intervened matched outcomes using multilevel mixed-effects regression models. In addition, collaborations between policymakers and researchers in the design of the policy and its evaluation component can enhance the learning process about the policy effectiveness and potentially inform extension of these practices, as exemplified by Macmillan and colleagues [68]. The authors developed a controlled before-after intervention to investigate the physical activity and equity impacts of an urban design's intervention with the aim of fostering physical activity in deprived areas in Auckland, New Zealand.

Another challenge is that many of the reviewed articles analyzed a policy's impacts on environmental pathways without assessing health impacts. For example, applications using process-based simulations (e.g., air pollution from dispersion models) and new causal inference methods, such as Bayesian structural time series models [69], are growing but often do not continue to quantify health impacts [70, 71]. In general, for a comprehensive evaluation of a policy, the impact on several health outcomes should be evaluated. We encourage researchers to incorporate more environmental health analyses in policy impact evaluation. Integrating team members with a solid background on health impact evaluations would be beneficial in those cases. In addition, implementing data sharing standards (e.g., FAIR protocol [72]) could allow for reusing the results of impacts on environmental pathways for health analysis in subsequent studies if adequately documented and, ideally, accompanied with uncertainty estimates. For instance, this could enable intercomparison of health impact evaluations from different research groups and over different spatio-temporal contexts answering different research questions (e.g., impact of policy-derived changes on chronic vs. acute health outcomes).

Lastly, a general challenge that may limit the usefulness of evaluation studies is that many cities often do not include evaluation efforts as part of the policies aiming to promote public health [1••, 73, 74]. Such policies require an evaluation component that explicitly incorporates estimation of environmental health impacts at multiple policy process stages aiming to monitor and improve the effectiveness of the policies [75]. Enabling collaborations between policymakers and researchers embedded within the policy process over the long term may contribute to developing more effective interventions, as suggested by Lowe and colleagues [76]. For instance, the INTERACT program is a collaboration of researchers, urban planners, and citizens assessing the effectiveness of built environment changes on health in four Canadian cities through observational studies [77, 78].

Conclusions

Policies are required to enhance equitability and health in cities; impact evaluations can be useful to monitor and improve the effectiveness of urban policies. In this narrative review, we investigated the most recent methods and research questions from health impact evaluations of policies at different stages, from the diagnosis of environmental problems affecting human health to the design, pilot, implementation, operation, and dismantling of the policy. This life cycle perspective allowed us to identify the scarcity of studies assessing the health impact of policies at multiple stages. The predominant approach in the literature of statically assessing impacts at single stages of the policy, during a specific time frame, is insufficient for fully understanding the impact of a policy. We recommend that researchers connect evaluations of the impacts of a policy at multiple stages in order to develop comprehensive impact evaluations of the full policy life course.

Declarations

Conflict of Interest The authors declare no competing interests.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1.•• Nieuwenhuijsen MJ. Urban and transport planning pathways to carbon neutral, liveable and healthy cities; a review of the current evidence. Environ Int. 2020;140:105661. **A review**

of research on quantitative evidence of how cities can improve environmental health through urban and transport planning**

2. Baghestani A, Tayarani M, Allahviranloo M, Gao HO. Evaluating the traffic and emissions impacts of congestion pricing in New York City. Sustain. 2020;12(9):1–16.

3. Eliasson J, Hultkrantz L, Nerhagen L, Rosqvist LS. The Stockholm congestion - charging trial 2006: overview of effects. Transp Res Part A Policy Pract. 2009;43(3):240–50.

4. Holman C, Harrison R, Querol X. Review of the efficacy of low emission zones to improve urban air quality in European cities. Atmos Environ. 2015;111:161–9.

5. Cole HVS, Lamarca MG, Connolly JJT, Anguelovski I. Are green cities healthy and equitable? Unpacking the relationship between health, green space and gentrification. J Epidemiol Community Health. 2017;71(11):1118–21.

6. Mason TG, Schooling CM, Chan KP, Tian L. An evaluation of the air quality health index program on respiratory diseases in Hong Kong: an interrupted time series analysis. Atmos Environ. 2019;211:151–8.

7. Dianati K, Zimmermann N, Milner J, Muindi K, Ezeh A, Chege M, et al. Household air pollution in Nairobi's slums: a long-term policy evaluation using participatory system dynamics. Sci Total Environ. 2019;660:1108–34.

8. Talen E. Do plans get implemented? A review of evaluation in planning. J Plan Lit. 1996;10:248–59.

9. Briggs DJ. A framework for integrated environmental health impact assessment of systemic risks. Environ Heal A Glob Access Sci Source. 2008;7:1–17.

10. Corburn J, Cohen AK. Why we need urban health equity indicators: integrating science, policy, and community. PLoS Med. 2012;9(8):1–6.

11. Giles-Corti B, Vernez-Moudon A, Reis R, Turrell G, Dannenberg AL, Badland H, et al. City planning and population health: a global challenge. Lancet. 2016;388(10062):2912–24.

12. Boogaard H, van Erp AM, Walker KD, Shaikh R. Accountability studies on air pollution and health: the HEI experience. Curr Environ Health Rep. 2017;4(4):514–22.

13. Knol AB, Briggs DJ, Lebret E. Assessment of complex environmental health problems: framing the structures and structuring the frameworks. Sci Total Environ. 2010;408(14):2785–94.

14. Health Effects Institute. C O M M U N I C A T I O N 11 Assessing health impact of air quality regulations : concepts and methods for accountability research. HEI Commun. 2003;11.

15. Ding P, Li F. Causal inference: a missing data perspective. Stat Sci. 2018;33(2):214–37.

16. Dickerman BA, Hernán MA. Counterfactual prediction is not only for causal inference. Eur J Epidemiol. 2020;35(7):615–7.

17. Benmarhnia T, Schwarz L, Nori-Sarma A, Bell ML. Quantifying the impact of changing the threshold of New York City heat emergency plan in reducing heat-related illnesses. Environ Res Lett. 2019;14(11).

18. Nethery RC, Mealli F, Sacks JD, Dominici F. Evaluation of the health impacts of the 1990 Clean Air Act Amendments using causal inference and machine learning. J Am Stat Assoc. 2020;0(0):1–28.

19. Jack DW, Kinney PL. Health co-benefits of climate mitigation in urban areas. Curr Opin Environ Sustain. 2010;2(3):172–7.

20.●● Buse CG, Lai V, Cornish K, Parkes MW. Towards environmental health equity in health impact assessment: innovations and opportunities. Int J Public Health. 2019;64(1):15–26. **A review summarizing the evidence to incorporate equity in health impact assessment research, highlighting the need for investigating the root causes of inequities**

21. Henneman LRF, Liu C, Mulholland JA, Russell AG. Evaluating the effectiveness of air quality regulations: a review of accountability studies and frameworks. J Air Waste Manag Assoc. 2017;67(2):144–72.

22. Münzel T, Sørensen M, Lelieveld J, Hahad O, Al-Kindi S, Nieuwenhuijsen M, et al. Heart healthy cities: genetics loads the gun but the environment pulls the trigger. Eur Heart J. 2021;2422–38.

23. Rich DQ. Accountability studies of air pollution and health effects: lessons learned and recommendations for future natural experiment opportunities. Environ Int. 2017;100:62–78.

24. Bell ML, Morgenstern RD, Harrington W. Quantifying the human health benefits of air pollution policies: review of recent studies and new directions in accountability research. Environ Sci Policy. 2011;14(4):357–68.

25 Cimorelli AJ, Perry SG, Venkatram A, Weil JC, Paine RJ, Wilson RB, et al. AERMOD: a dispersion model for industrial source applications. Part I: general model formulation and boundary layer characterization. J Appl Meteorol. 2005;44(5):682–93.

26. Petetin H, Bowdalo D, Soret A, Guevara M, Jorba O, Serradell K, et al. Meteorology-normalized impact of the COVID-19 lockdown upon NO_2 pollution in Spain. Atmos Chem Phys. 2020;20(18):11119–41.

27. Otero I, Nieuwenhuijsen MJ, Rojas-Rueda D. Health impacts of bike sharing systems in Europe. Environ Int. 2018;115:387–94.

28. Tonne C, Beevers S, Armstrong B, Kelly F, Wilkinson P. Air pollution and mortality benefits of the London Congestion Charge: spatial and socioeconomic inequalities. Occup Environ Med. 2008;65(9):620–7.

29. Johnson S, Haney J, Cairone L, Huskey C, Kheirbek I. Assessing air quality and public health benefits of New York City's climate action plans. Environ Sci Technol. 2020;54(16):9804–13.

30. Benavides J, Guevara M, Snyder MG, Rodríguez-Rey D, Soret A, García-Pando CP, et al. On the impact of excess diesel NOX emissions upon NO_2 pollution in a compact city. Environ Res Lett. 2021;16(2).

31. 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.

32. He G, Pan Y, Tanaka T. The short-term impacts of COVID-19 lockdown on urban air pollution in China. Nat Sustain. 2020;3(12):1005–11.

33.●● Khomenko S, Cirach M, Pereira-Barboza E, Mueller N, Barrera-Gómez J, Rojas-Rueda D, et al. Premature mortality due to air pollution in European cities: a health impact assessment. Lancet Planet Heal. 2021;5(3):e121–34. **A study examining the effects of air pollution exposure estimated using LUR models on premature mortality for adult residents in 969 cities in Europe**

34. Khomenko S, Nieuwenhuijsen M, Ambròs A, Wegener S, Mueller N. Is a liveable city a healthy city? Health impacts of urban and transport planning in Vienna. Austria Environ Res. 2020;183:109238.

35 Tong R, Liu J, Wang W, Fang Y. Health effects of $PM_{2.5}$ emissions from on-road vehicles during weekdays and weekends in Beijing. China Atmos Environ. 2020;223:117258.

36.●● Mueller N, Rojas-Rueda D, Khreis H, Cirach M, Andrés D, Ballester J, et al. Changing the urban design of cities for health: the superblock model. Environ Int. 2020;134:105132. **An assessment of the potential health impacts of the superblock model in Barcelona, Spain, using a multi-exposure approach**

37. Rossi IA, Vienneau D, Ragettli MS, Flückiger B, Röösli M. Estimating the health benefits associated with a speed limit reduction to thirty kilometres per hour: a health impact assessment of noise and road traffic crashes for the Swiss city of Lausanne. Environ Int. 2020;145.

38. Johansson C, Burman L, Forsberg B. The effects of congestion tax on air quality and health. Atmos Environ. 2009;43(31):4843–54.

39. 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. Transp Res Part A Policy Pract. 2019;123:147–69.

40.•• Lowe M, Arundel J, Hooper P, Rozek J, Higgs C, Roberts R, et al. Liveability aspirations and realities: implementation of urban policies designed to create healthy cities in Australia. Soc Sci Med. 2020;245:112713. **A study assessing the capacity of existing urban policies to equitably deliver healthy cities, including their level of implementation, in the four largest Australian cities**

41. Bian Z, Zuo F, Gao J, Chen Y, PavuluriVenkata SSC, Duran Bernardes S, et al. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Seattle. Transp Res Part A Policy Pract. 2021;145:269–83.

42. Martenies SE, Milando CW, Batterman SA. Air pollutant strategies to reduce adverse health impacts and health inequalities: a quantitative assessment for Detroit. Michigan Air Qual Atmos Heal. 2018;11(4):409–22.

43. Zeng J, Guijarro M, Carrilero-Castillo A. A regression discontinuity evaluation of the policy effects of environmental regulations. Econ Res Istraz. 2020;33(1):2993–3016.

44.•• Cole MA, Elliott RJR, Liu B. The impact of the Wuhan Covid-19 lockdown on air pollution and health: a machine learning and augmented synthetic control approach. Environ Resour Econ. 2020;76(4):553–80. **A study applying new causal inference methods based on machine learning techniques to investigate the air quality and health impacts of the Covid-19 lockdown in Wuhan, China**

45. Nardone AL, Casey JA, Rudolph KE, Karasek D, Mujahid M, Morello-Frosch R. Associations between historical redlining and birth outcomes from 2006 through 2015 in California. PLoS ONE. 2020;15:1–18.

46 Zhao B, Wang S, Ding D, Wu W, Chang X, Wang J, et al. Non-linear relationships between air pollutant emissions and $PM_{2.5}$-related health impacts in the Beijing-Tianjin-Hebei region. Sci Total Environ. 2019;661:375–85.

47. Mitsakou C, Dimitroulopoulou S, Heaviside C, Katsouyanni K, Samoli E, Rodopoulou S, et al. Environmental public health risks in European metropolitan areas within the EURO-HEALTHY project. Sci Total Environ. 2019;658:1630–9.

48. Mueller N, Rojas-Rueda D, Khreis H, Cirach M, Milà C, Espinosa A, et al. Socioeconomic inequalities in urban and transport planning related exposures and mortality: a health impact assessment study for Bradford. UK Environ Int. 2018;121:931–41.

49. Seposo X, Kondo M, Ueda K, Honda Y, Michikawa T, Yamazaki S, et al. Health impact assessment of $PM_{2.5}$-related mitigation scenarios using local risk coefficient estimates in 9 Japanese cities. Environ Int. 2018;120:525–34.

50. Thondoo M, Mueller N, Rojas-Rueda D, de Vries D, Gupta J, Nieuwenhuijsen MJ. Participatory quantitative health impact assessment of urban transport planning: a case study from Eastern Africa. Environ Int. 2020;144:106027.

51. Malmqvist E, Lisberg Jensen E, Westerberg K, Stroh E, Rittner R, Gustafsson S, et al. Estimated health benefits of exhaust free transport in the city of MalmÖ. Southern Sweden Environ Int. 2018;118:78–85.

52. De Vos J. Road pricing in a polycentric urban region: analysing a pilot project in Belgium. Transp Policy. 2016;52:134–42.

53. Eliasson J. Is congestion pricing fair? Consumer and citizen perspectives on equity effects. Transp Policy. 2016;52:1–15.

54. Cesaroni G, Boogaard H, Jonkers S, Porta D, Badaloni C, Cattani G, et al. Health benefits of traffic-related air pollution reduction in different socioeconomic groups: the effect of low-emission zoning in Rome. Occup Environ Med. 2012;69(2):133–9.

55. Gu J, Mohit B, Muennig PA. The cost-effectiveness of bike lanes in New York City. Inj Prev. 2017;23(4):239–43.

56. Nardone A, Rudolph KE, Morello-Frosch R, Casey JA. Redlines and greenspace: the relationship between historical redlining and 2010 greenspace across the United States. Environ Health Perspect. 2021;129(1):1–9.

57. Rodriguez-Rey D, Guevara M, Linares MP, Casanovas J, Armengol JM, Benavides J, et al. To what extent the traffic restriction policies applied in Barcelona city can improve its air quality? Sci Total Environ. 2021;807:150743.

58. Mudway IS, Dundas I, Wood HE, Marlin N, Jamaludin JB, Bremner SA, et al. Impact of London's low emission zone on air quality and children's respiratory health: a sequential annual cross-sectional study. Lancet Public Heal. 2019;4(1):e28-40.

59. Chart-asa C, Gibson JMD. Health impact assessment of traffic-related air pollution at the urban project scale: influence of variability and uncertainty. Sci Total Environ. 2015;506–507:409–21.

60. Stuart EA, Huskamp HA, Duckworth K, Simmons J, Song Z, Chernew ME, et al. Using propensity scores in difference-in-differences models to estimate the effects of a policy change. Heal Serv Outcomes Res Methodol. 2014;14(4):166–822.

61. Sant'Anna PHC, Zhao J. Doubly robust difference-in-differences estimators. J Econom. 2020;219:101–223.

62. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–554.

63. Westreich DJ, Lessler J, Jonsson Funk M. Classification methods as alternatives to logistic regression. J Clin Epidemiol. 2010;63(8):826–335.

64. Abadie A, Diamond A, Hainmueller AJ. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J Am Stat Assoc. 2010;105(490):493–5056.

65. Ben-Michael E, Feller A, Rothstein J. The augmented synthetic control method. J Am Stat Assoc. 2021

66. Benton JS, Cotterill S, Anderson J, Macintyre VG, Gittins M, Dennis M, et al. A natural experimental study of improvements along an urban canal: impact on canal usage, physical activity and other wellbeing behaviours. Int J Behav Nutr Phys Act. 2021;18(1):1–16.

67. Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. 2018;18(1):1–15.

68. Macmillan AK, Mackie H, Hosking JE, Witten K, Smith M, Field A, et al. Controlled before-after intervention study of suburb-wide street changes to increase walking and cycling: Te Ara Mua-Future Streets study design. BMC Public Health. 2018;18(1):1–13.

69. Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL. Inferring causal impact using Bayesian structural time-series models. Ann Appl Stat. 2015;9(1):247–74.

70. Wang S, Liu X, Zhou C, Hu J, Ou J. Examining the impacts of socioeconomic factors, urban form, and transportation networks on CO_2 emissions in China's megacities. Appl Energy. 2017;185:189–200.

71. Zhang Y, Fricker JD. Quantifying the impact of COVID-19 on non-motorized transportation: a Bayesian structural time series model. Transp Policy. 2021;103(January):11–20.

72. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. Comment: The FAIR Guiding

Principles for scientific data management and stewardship. Sci Data. 2016;3:1–9.

73. Ramirez-Rubio O, Daher C, Fanjul G, Gascon M, Mueller N, Pajín L, et al. Urban health: an example of a "health in all policies" approach in the context of SDGs implementation. Global Health. 2019;15(1):1–21.

74. Backholer K, Baum F, Finlay SM, Friel S, Giles-Corti B, Jones A, et al. Australia in 2030: what is our path to health for all? Med J Aust. 2021;214(S8):S5-40.

75. Hoffmann B, Roebbel N, Gumy S, Forastiere F, Brunekreef B, Jarosinska D, et al. Air pollution and health: recent advances in air pollution epidemiology to inform the European Green Deal: a joint workshop report of ERS, WHO, ISEE and HEI. Eur Respir J. 2020;56(5).

76. Lowe M, Hooper P, Jordan H, Bowen K, Butterworth I, Giles-Corti B. Evidence-informed planning for healthy liveable cities: how can policy frameworks be used to strengthen research translation? Curr Environ Health Rep. 2019;6(3):127–36.

77. Fuller D, Bell S, Firth CL, Muhajarine N, Nelson T, Stanley K, et al. Wave 1 results of the INTerventions, Research, and Action in Cities Team (INTERACT) cohort study: examining spatio-temporal measures for urban environments and health. Health Place. 2021:102646.

78. Kestens Y, Winters M, Fuller D, Bell S, Berscheid J, Brondeel R, et al. INTERACT: a comprehensive approach to assess urban form interventions through natural experiments. BMC Public Health. 2019;19(1):1–11.