Health & Place 27 (2014) 106–111

Contents lists available at ScienceDirect

Health & Place

journal homepage: www.elsevier.com/locate/healthplace

Viewpoint

Stepping towards causation in studies of neighborhood and environmental effects: How twin research can overcome problems of selection and reverse causation

Glen E. Duncan [a,b,*], Brianna Mills [a], Eric Strachan [c], Philip Hurvitz [d], Ruizhu Huang [e], Anne Vernez Moudon [a,d], Eric Turkheimer [f]

[a] Department of Epidemiology, University of Washington, Seattle, WA 98195, USA
[b] Interdisciplinary Graduate Program in Nutritional Sciences, University of Washington, Seattle, WA 98195, USA
[c] Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195, USA
[d] Department of Urban Design and Planning, Urban Form Lab, University of Washington, Seattle, WA 98195, USA
[e] Interdisciplinary PhD Program in Urban Design and Planning, Urban Form Lab, University of Washington, Seattle, WA 98195, USA
[f] Department of Psychology, University of Virginia, Charlottesville, VA 22904, USA

ARTICLE INFO

Article history:
Received 3 October 2013
Received in revised form
3 January 2014
Accepted 14 February 2014
Available online 3 March 2014

Keywords:
Causality
Environment design
Lifestyle risk reduction
Social and built environments
Twin studies

ABSTRACT

No causal evidence is available to translate associations between neighborhood characteristics and health outcomes into beneficial changes to built environments. Observed associations may be causal or result from uncontrolled confounds related to family upbringing. Twin designs can help neighborhood effects studies overcome selection and reverse causation problems in specifying causal mechanisms. Beyond quantifying genetic effects (i.e., heritability coefficients), we provide examples of innovative measures and analytic methods that use twins as quasi-experimental controls for confounding by environmental effects. We conclude that collaboration among investigators from multiple fields can move the field forward by designing studies that step toward causation.

1. Introduction

The role of built and social environments in supporting healthy lifestyles has garnered increasing attention as research paradigms shift from a focus on individuals to an assessment of macro-level influences operating within ecological models of health. Researchers have long recognized that both built and social environments affect individual health. However, studies that seek to specify the causal mechanisms linking “neighborhood effects” to health outcomes suffer from two intractable problems: selection and reverse causation.

To illustrate the complexity of these problems, we present Fig. 1 as a simplified example of the multiple levels of influence that interact to produce the current obesity epidemic in the U.S. (National Heart Lung and Blood Institute, 2004). It is important to appreciate that these influences have an effect not only on behavior, but also on each other through interaction, reinforcement, and even antagonism. Past approaches to slow the obesity epidemic have been less than effective (Flegal et al., 2010; Ogden et al., 2007), primarily targeting individual-level influences (Fig. 1, upper two boxes). Experts increasingly recognize that outcomes in a population are a function of macro-level influences (Fig. 1, bottom two boxes) that affect the lives of all people. To have the greatest population-level effect, multi-level interventions should start with environmental changes and end with individual-level programming, rather than the reverse. New methods make it possible to measure energy balance behaviors within the social and physical environments that support or hinder them. Objective macro-environmental data help formulate policies that regulate large-scale environmental influences on healthy lifestyles, such as funding for urban design and infrastructure that support active modes of transportation. A focus on the built and social environment and its implications for policy and legislation is thus a necessary step if we are to manage the epidemics of obesity and related chronic diseases.

In the following sections we outline the problems of selection and reverse causation and argue for the utility of twin study designs in overcoming them. We present examples from our research with the University of Washington Twin Registry (UWTR),

* Corresponding author at: University of Washington, Box 353410, Seattle, WA 98195-3410, USA. Tel.: +1 206 616 2680; fax: +1 206 685 1696.
E-mail address: duncag@u.washington.edu (G.E. Duncan).

http://dx.doi.org/10.1016/j.healthplace.2014.02.008
1353-8292

including novel analytic approaches, and conclude with suggestions for future research collaborations.

2. The selection problem

Imagine a study in which participants are randomly assigned to residential locations and followed longitudinally to investigate the effects of various social and built environments on their health. For ethical, legal, and practical reasons such a study is impossible. In the real world choices of residential neighborhood are nonrandom, and studies have yet to adequately control for this (Diez Roux, 2002; Oakes, 2004, 2006; Subramanian, 2004). Most existing statistical models simply adjust for individual characteristics associated with residential location, such as age, ethnicity/race, and socio-economic status. More recently, reasons for choosing a neighborhood have been measured in surveys and subsequently controlled for in analyses (Frank et al., 2007). Even studies that adjust for such characteristics inevitably include unrecognized or difficult to measure factors. Such factors are left uncontrolled and the resulting bias reduces inferential validity (Oakes, 2006). This, in a nutshell, is the selection problem.

Selection problem can also refer to the possibility that individuals select residential environments based on genetics or family upbringing (Duncan et al., 2012; Whitfield et al., 2005; Willemsen et al., 2005). The selection variables, rather than any putative environmental effects, may be responsible for findings that link environmental characteristics to health outcomes. Twins are ideal study subjects for overcoming this problem. Twins are always the same age and ethnicity/race, effectively controlling for those key demographic factors. Monozygotic (MZ) or identical twins are matched genetically and always of the same sex, so within-pair variation in health and environmental risk cannot be attributed to differences in genetics or sex. Among dizygotic (DZ) or fraternal twins, confounding by sex can be eliminated by studying same-sex pairs. Further, because MZ and DZ twins share a common family of origin (twins are almost invariably reared together) differences within pairs cannot be attributed to family background or childhood exposures. In cross-sectional studies, twins can be used as quasi-experimental controls for confounding by genetic, family, and early life environmental effects on outcomes that cannot be held constant by random assignment (Turkheimer, 2008). Because twins often become discordant in neighborhood residence and lifestyle later in life, we can detect environmental effects on health while reducing the confounding inherent in studies of unrelated individuals in nonrandom environments (Moffitt, 2005).

3. The reverse causation problem

The problem of reverse causation (or direction of causation) refers to the possibility that behaviors and health play causal roles in the choice of residential location, rather than the converse. Once again, it is neither ethical nor practical to randomize people to different residential environments and follow them prospectively to determine environmental effects on outcomes of interest. The cross-lagged panel design is a method commonly used in psychology and behavioral research for testing spuriousness by comparing cross-lagged correlations in cases in which the independent variable cannot be experimentally manipulated (Kenny, 1975), such as arises with respect to residential environments. Below we describe how this model has been adopted by twin researchers to supports a genetically informed, cross-lagged panel design that obtains environmental and health data at two or more time points to conduct analyses that support directional causal inferences while controlling for covariation that originates in either genetic or shared environmental variation.

In a cross-lagged design (Fig. 2) two traits, a purported cause (e.g. walkability) and outcome (e.g. walking level), are measured at two time points. The outcome at Time 2 is regressed on the outcome at Time 1 (stability), and on the purported cause at Time 1 (cross-lagged regression). The cross-lagged regression represents the effect of the purported cause at Time 1 on the change in outcome between Time 1 and 2. This coefficient can be compared to its converse, the regression of the purported cause at Time 2 on the purported effect at Time 1, conditional on the stability of the purported cause between Times 1 and 2. If the relationship between Time 1 and 2 is causal, the cross-lagged coefficient is significant, and significantly larger than the regression of the cause on the effect. Using twins, cross-lagged regression coefficients are estimated while controlling for covariation between the purported cause and outcome that originates in either genetic (A) or shared environmental variation (C) that is shared between cause and effect, known as a genetically informed cross-lag design. The residual variation in the purported effect is also decomposed into genetic and environmental components. The partitioning of variance into genetic, shared, and non-shared environmental variation (E) is accomplished with the usual classical twin methods (i.e., based on the correlation between the genetic backgrounds of a pair of twins being 1.0 in MZs and 0.5 in DZs). To the extent there is a causal relationship between the predictor and outcome, the cross-lagged regression will remain significant even after the association arising from familial background variables has been accounted for.

Although the twin design controls for the possibility that individuals select residential environments based on genetics or

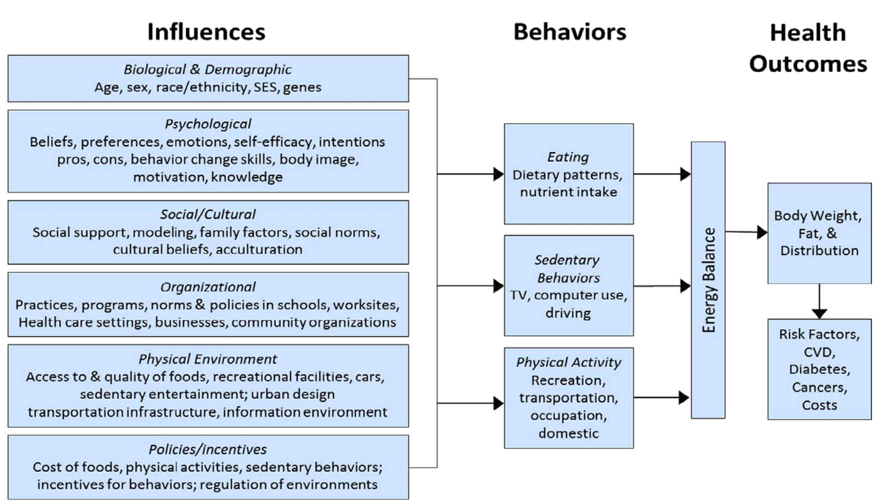

Fig. 1. An ecological model of diet, physical activity, and obesity. Abbreviation: CVD=cardiovascular disease. Developed for the NHLBI workshop on predictors of obesity, weight gain, diet, and physical activity; August 4–5, 2004, Bethesda, MD.