Journal of Transport & Health ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Journal of Transport & Health

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

The relationship between active transportation and health

Stephanie A. Schauder [a,1], Mark C. Foley [b,*]

[a] Davidson College, P.O. Box 5024, Davidson, NC 28035, United States
[b] Davidson College, P.O. Box 6941, Davidson, NC 28035, United States

ARTICLE INFO

Article history:
Received 10 March 2015
Received in revised form
24 June 2015
Accepted 30 June 2015

Keywords:
Active transportation
Bicycling
Walking
Health
BMI
Cholesterol

ABSTRACT

Active transportation has received attention for its environmental and health benefits. In this paper we use individual-level survey data from NHANES III to investigate the extent to which the number of minutes of bicycling and walking for transportation are associated with 10 health outcomes. We use instrumental variables to address the endogeneity caused by the complex relationship between exercise and health. We find mixed results indicating that active transportation is associated with improvements in some health outcomes such as weight and cholesterol, but not others such as systolic blood pressure and glycohemoglobin.

1. Introduction

Many studies have shown the positive effects of exercise on various health outcomes (Contoyannis and Jones, 2004; Humphreys et al., 2014; Irwin et al., 2003). Recently, active transportation (bicycling and walking) has garnered attention as a feasible way to increase exercise and thereby positively affect health (Cooper et al., 2008; de Hartog et al., 2010; Rojas-Rueda et al., 2011; Wen and Rissel, 2008). Among Swedish children, regular cycling to school has been associated with improved fitness, measured by maximal oxygen consumption (VO_2max), while there was no significant improvement in fatness or cardiometabolic risk factors (Chillón, et al., 2012). A study of Danish children concluded that those who commuted to school via cycle had significantly better cardiovascular risk profiles (Andersen et al., 2011). Among adults, studies have also found cycling is associated with reduced obesity, better fitness, and reduced risk of all-cause mortality (Gordon-Larsen et al., 2009; Oja et al., 1991; Andersen et al., 2000). Furthermore, active transportation's lessened impact on the environment can make it additionally attractive (Larouche, 2012). This has led some researchers to assert that building infrastructure for active transportation can be profitable because the benefits of avoided environmental damage and healthcare expenditures are greater than the infrastructure costs (Gronau et al., 2013; Wang et al., 2004).

Public health is a concern in any country because it would benefit all citizens to have a healthy population as this would decrease healthcare expenditures. Healthcare spending in the United States is expected to grow faster than Gross Domestic Product for at least the rest of this decade. As the Affordable Care Act expands insurance coverage, Medicaid spending is projected to grow by 8.2% per year on average, and even private spending is expected to grow by 5.3% (Centers for Medicare and Medicaid Services, 2014). Data from the Canadian Community Health Survey reveal that physically inactive individuals have been shown to place a larger burden on the healthcare system spending on average 38% more time in the hospital (Sari, 2009). Also, a study of Dutch workers found that those who commuted by bicycle were significantly less likely to miss work due to sickness (Hendriksen et al., 2010). Another study found that connected street networks, which have been shown to increase walking and bicycling, are correlated with reduced rates of obesity, heart disease, high blood pressure, and diabetes (Marshall et al., 2014). This suggests those who use active transportation could be more productive in addition to placing a smaller burden on the healthcare system. If active transportation is effective in improving health, the resulting pecuniary benefits should be weighed against the costs of incentivizing bicycling and walking in order to determine if promoting active transportation would be a cost effective policy.

---

[*] Corresponding author. Tel.: +1 704 894 2248.
E-mail addresses: stschauder@davidson.edu (S.A. Schauder), mafoley@davidson.edu (M.C. Foley).
[1] Tel.: +1 704 778 2280.

http://dx.doi.org/10.1016/j.jth.2015.06.006

However, this clear correlation between active transportation and health does not necessarily mean that active transportation causes good health. When there are endogeneity concerns about measures of active transportation, there is less reason to believe that encouraging bicycling and walking for transportation will improve the health of a population. The aim of this study is to address this endogeneity by using instrumental variables and consequently shed light on the causal effect of active transportation on health.

2. Methods and data

The data come from the National Health and Nutrition Examination Survey (NHANES). NHANES III is designed to evaluate the health and nutrition of residents of the United States through both examination and interview (National Health and Nutrition Examination Survey 2013–2014 Overview, 2014). Each respondent (for whom NHANES has complete data) participated in an in-house interview followed by a physician examination, which included laboratory work, so the data is a mix of self-reported (questionnaire), physician reported (examination), and laboratory data (National Health and Nutrition Examination Survey Questionnaires, 2014; National Health and Nutrition Examination Survey 1999–2014 Survey Content Brochure, 2014). NHANES randomly selects counties across the U.S. from which to sample, but in some years demographic groups of particular interest are over-sampled. The years chosen for analysis were 2007–2010. Respondents were only sampled in one year, so additional years were included to increase the sample size (National Health and Nutrition Examination Survey Sample Design 2007–2010, 2013). The most recent year in which all relevant data were available was 2010 and no data before 2007 were used because some important survey questions were not asked in prior years. In the survey cycle 2007–2010, Hispanics were oversampled (Curtin et al., 2013).

We defined the following 10 health outcome variables: BMI, Overweight, Obese, Systolic blood pressure, FEV1/FVC ratio, Total cholesterol level, HDL cholesterol level, TC/HDL cholesterol ratio, Glycohemoglobin level, and Good self-reported health. The variables are in their natural units unless otherwise indicated. Overweight and Obese are dichotomous variables created from the continuous BMI measure. For BMI of 25 or greater, Overweight was coded as 1, and 0 otherwise, and for a BMI of 30 or greater, Obese was coded as 1, and 0 otherwise (World Health Organization, 2014). Self-reported health was reported on a scale from 1 to 5, with 1 indicating “excellent,” 2 was “very good,” 3 was “good,” 4 was “fair,” and 5 was “poor.” Very good self-reported health was coded as 1 if a respondent answered “very good” or “excellent,” and 0 otherwise.

The variable of interest was the survey respondent's reported daily minutes of active transportation. The survey respondent was first prompted to think about the way he/she traveled to and from places such as work, shopping, or school. Then the respondent was asked, if he/she walked or used a bicycle for at least 10 min continuously to get to and from places. If the respondent answered no, the number of minutes of active transportation was coded as zero. If the respondent answered yes, he/she was further asked to report the number of minutes spent walking or bicycling for transportation purposes on a typical day (National Health and Nutrition Examination Survey 2013–2014 Overview, 2014). This variable is considered to be endogenous. An endogenous variable is one that is jointly determined with the outcome variable, a health outcome in this case, or is otherwise correlated with variables not explicity included in the regression model but that influence the outcome. This is in contrast to an exogenous variable which is one that is assumed to not be correlated with variables not explicitly included in the regression model.

NHANES also supplied variables of PA and nutrition. A series of questions identically structured to those about active transportation were asked to ascertain the number of minutes the survey respondent spent on work physical activity (such as lifting heavy loads) or recreational physical activity (such as playing sports). These three categories were mutually exclusive. These are control variables, which are variables that must be held constant in order to estimate accurately the relationship between the variable of interest, minutes of active transportation, and the outcome. The Healthy Eating Index was constructed based on a 24 dietary recall questionnaire that was asked to all survey respondents (Healthy Eating Index, 2013). Additional control variables include age, income to poverty ratio, gender, education level, race, capital income, health insurance status, marital status, survey year, smoking behavior, whether the respondent was born in the U.S., occupation, and a measure of food security. The income-to-poverty ratio was used instead of income because it takes into account the respondents' family size and requirements to qualify for financial assistance in his or her state (National Health and Nutrition Examination Survey 2013–2014 Overview, 2014). Capital income was included to control, at least in part, for wealth. The omitted referent categories are indicated in Table 2 of the Section 2. All of the control variables listed in this paragraph are included in each regression analysis.

3. Theory and statistical analysis

The influence of physical activity on health is thought to have diminishing marginal returns, meaning that the effect of exercise on health will be largest for those with the smallest amount of baseline activity (Humphreys et al., 2014). For this reason, we chose a linear-log functional form.

$$H_i = \beta_0 + \beta_1 Ln(AT_i) + \beta_2 X_i + \varepsilon_i \tag{1}$$

Eq. (1) was used to obtain the ordinary least squares estimate of β_1, where H_i represents the total stock of health for person i, $ln(AT_i)$ is the natural logarithm of the minutes of active transportation of person i, and X_i signifies all the control variables noted above. In practice there is no one variable for H_i, so health stock is approximated by the 10 aforementioned health outcomes. Because the natural log of zero is undefined, if minutes of active transportation was equal to zero, it was recoded as 0.25.

However, when estimating the effect of active transportation on health, it is likely that there is either reverse causality (good health makes individuals more likely to use active transportation), or at least one underlying unmeasurable variable, such as internal motivation or persistence, that is affecting both propensity to use active transportation and various health outcomes. In these situations, active transportation is called an endogenous independent variable, which is another way of saying that there is a non-zero correlation between active transportation and the error term in Eq. 1, after controlling for the other independent variables. Such a correlation leads Ordinary Least Squares (OLS) coefficient estimates to be biased (Wooldridge, 2010). Many studies that examine active transportation do not account for this endogeneity of the independent variable of interest and, consequently, the resulting estimates could be biased. However, in order to evaluate the cost effectiveness of incentivizing walking and cycling, it is important to accurately predict the effect of active transportation on health. Indeed Cavill et al. (2008) note a lack of transparency and consistency of results in studies on this subject, which makes it difficult for policy makers to understand the true costs and benefits of active transportation.

Yet some related studies have used instrumental variables to examine the effect of exercise on health. An instrumental variable is a variable, call it Z_i, which is not among the set of control variables X_i above, and which satisfies two conditions: relevance and exogeneity (Wooldridge, 2010). First, relevance means that, in order to conduct accurate hypothesis tests, it must be strongly correlated with the endogenous independent variable: in this case $corr(Z_i, Ln(AT_i) \neq 0)$. Second, exogeneity means it must be uncorrelated with the error term, that is, variables not explicitly controlled for in the model: in this case, $corr(Z_i, \varepsilon_i = 0)$. One can think of instrumental variable estimation as a two-stage process in which the first stage is a regression of the endogeneous independent variable on the instrumental variable and the control variables. Predicted values from this regression would represent the variation of $Ln(AT_i)$ that is uncorrelated with the error term. The second stage is a regression equivalent to Eq. (1) but with $Ln(AT_i)$ replaced by those predicted values from the first stage.[2] Intuitively, the instrumental variable isolates the variation in $Ln(AT_i)$ that is *not* due to reverse causality or omitted variables, which is called

---

[2] Estimation is not actually carried out in two stages so that standard errors are correctly calculated.

Please cite this article as: Schauder, S.A., Foley, M.C., The relationship between active transportation and health. Journal of Transport & Health (2015), http://dx.doi.org/10.1016/j.jth.2015.06.006