Economics and Human Biology 22 (2016) 161–176

Contents lists available at ScienceDirect

Economics and Human Biology

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

# Decomposing racial differences in adolescent smoking in the U.S.☆

Vikesh Amin*, Aparna Lhila

*Department of Economics, Central Michigan University, Michigan, MI 48859, USA*

ARTICLE INFO

*Article history:*
Received 31 July 2015
Received in revised form 4 May 2016
Accepted 7 May 2016
Available online 11 May 2016

*Keywords:*
Smoking
Racial differences
Decomposition

ABSTRACT

Despite declining smoking rates in the U.S., a substantial fraction of adolescents still smoke. In addition, there are notable racial differences in adolescent smoking. We use Add Health data and apply a nonlinear decomposition method to determine the extent to which racial differences in observable characteristics account for (i) the racial smoking gaps in adolescent smoking (ages 12–18) and (ii) racial gaps in the probability of becoming a smoker in young adulthood (ages 18–24), conditional on being a non-smoker in adolescence. The model includes a host of explanatory factors, including individual, family socio-economics, smoke exposure, school characteristics, and county crime rate. Of the 19 (9) percentage-point gap in white-black (white-Hispanic) smoking in adolescence, these factors together account for 22–28% (39–77%) of the smoking gap; and of the 18 (13) percentage-point gap in white-black (white-Hispanic) smoking up-take in young adulthood, these factors together account for 26–50% (48–100%) of the gap, depending on which set of coefficients are used for the decomposition. The biggest drivers of racial smoking gaps in adolescence are differences in friends' smoking and school peer smoking, while only school peer smoking contributes to the explained portion of racial gaps in smoking up-take in young adulthood.



## 1. Introduction

Since the publication of the 1964 Surgeon General report highlighting the dangers of smoking, smoking rates in the United States have fallen substantially from 43% in 1965 to 18% in 2012 (USDHHS, 2013). However, smoking is still the single largest cause of preventable death and has substantial monetary and non-monetary costs. The Surgeon General estimates that smoking causes almost half a million deaths per year, and costs $289–$333 billion, including over $150 billion in lost productivity due to premature death. Motivated by this, a large literature has focused on understanding the determinants of smoking. Research has shown that smoking is inversely related to socioeconomic factors such as education (Cowell, 2006), employment (Hersch, 2000), cigarette prices and taxes (Chaloupka, 1991), and restrictive tobacco control policies (Chaloupka and Grossman, 1996); on the other hand, smoking is positively related to factors like parental smoking (Loureiro et al., 2010; Balia and Jones, 2011), school peer smoking (Gaviria and Raphael, 2001; Fletcher, 2010; Powell et al., 2005), and neighborhood crime (Virtanen et al., 2007).

Despite the now well understood dangers of smoking (Doll and Peto, 1976; Doll et al., 1994, 2004),[1] in 2013 15.7% of students in grades 9–12 smoked cigarettes in the last month (Health People, 2020), and more than 80% of established adult smokers started smoking before age 18 (USDHHS, 2012). These statistics underscore the importance of studying adolescent

☆ This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis. We acknowledge financial support from Central Michigan University Grant Number C62238. We thank William Sullivan and Arthur Nunoo for their research assistance. The authors are grateful for the comments and insights provided by four anonymous referees, and for comments from participants at the 2015 Southern Economics Association annual conference, at which a prior version of this manuscript was presented. Any errors are ours.

* Corresponding author.
*E-mail addresses:* amin1v@cmich.edu (V. Amin), lhila1a@cmich.edu (A. Lhila).

[1] A series of studies by Drs. Richard Doll, Richard Peto, and coauthors has examined the results of the British Doctors Survey, which followed the smoking status of 34,439 male doctors in 1951. These studies found that prolonged users of cigarettes lost 10 years of life expectancy on average than their non-smoking peers. And, quitting at ages 60, 50, 40, and 30 gained 3, 6, 9, and 10 years of life expectancy, respectively.

http://dx.doi.org/10.1016/j.ehb.2016.05.001
1570-677X/

smoking. One of the aims of Healthy People 2020 is to decrease the proportion of adolescents who smoked in the last month to 16%. There are also large racial disparities in adolescent smoking. Among high school students in 2013, 18.6% of whites had smoked cigarettes in the last month, compared to only 14% of Hispanics and 8.2% of Blacks (Health People, 2020). But why should one's race affect his/her decision to initiate smoking? The underlying assumption of this study is that race connotes significant differences in culture, history, socioeconomics, and opportunities that lead to racial differences in health behaviors, including smoking. It is erroneous to think that race is associated with different biology, physiology, or ability that somehow correlates with one's decision to take up smoking. In fact, our main finding supports the assumption that racial differences in smoking are attributable to the racial difference in the environments in which the adolescent resides. Decomposition of racial differences in adult smoking has received some attention in the literature (Dubowitz et al., 2011) but to the best of our knowledge there are no studies decomposing racial gaps in adolescent smoking. Understanding drivers of racial differences in adolescent smoking is important, as it is a time when individuals are likely to initiate smoking.

We use the National Longitudinal Study of Adolescent Health and the nonlinear decomposition method by Fairlie (2005) to quantify how much of the white-black and white-Hispanic gaps in (1) adolescent smoking (ages 12–18) and (2) probability of becoming a smoker in young adulthood (ages 18–24) conditional on being a non-smoker in adolescence are accounted for by racial differences in individual characteristics, family socioeconomic status, smoke exposure, school peer characteristics and county crime. It is important to note that while the majority of the literature on smoking is concerned with finding causal estimates of the determinants of smoking, this is a descriptive exercise concerned with estimating the extent to which differences in observable characteristics account for the racial gaps in adolescent smoking.

This study, like other studies that apply Oaxaca-Blinder type decompositions suffer from the same limitations. The method will not lead to irrefutable evidence of behavioral relationships and policy prescriptions because it does not uncover mechanisms underlying the relationships between independent variables and the outcome. Decomposition exercises add value to the literature as they serve to draw attention to the most important correlates of the gap they seek to study. For example, this study includes a host of factors that are found to be important correlates of smoking in the literature: parents' smoking status (e.g., Loureiro et al., 2010; Lillard 2011), parents' socioeconomic status, including education and family income (Soteriades and DiFranza, 2003), area crime (Virtanen et al., 2007), demand for weight loss (Cawley et al., 2013), gender (Lundborg and Andersson, 2008), school peer smoking (Manski 1995; Gaviria and Raphael, 2001; Lundborg, 2006), and often unobserved characteristics like ability (Farrell and Fuchs, 1982). Of these correlates, this study highlights that the racial gaps in friends' smoking status, school peer smoking, ability, and maternal smoking account for the biggest portion of the racial/ethnic gap in adolescent smoking.

We acknowledge that the results of this analysis could be different if we were able to address the potential sources of endogeneity; however, estimating such a causal model that contains a number of potentially endogenous variables would rely on using a variety of identification strategies, which would make this exercise intractable. In the discussion section, we attempt to estimate how these results would be different if we used an instrumental variable for the variable (school peer smoking) that accounts for the majority of the explained portion of the racial gaps in white-black and white-Hispanic smoking.

## 2. Methodology

The most common approach used to understand drivers of inequalities between different groups is the Oaxaca-Blinder decomposition (Blinder, 1973; Oaxaca, 1973). Recent studies have used this approach to investigate racial differences in adult binge drinking and fruit and vegetable consumption in the U.S. (Dubowitz et al., 2011), racial differences in adolescent BMI in the U.S. (Powell et al., 2012) and gender differences in adult smoking in South Korea (Chung et al., 2010). The Oaxaca-Blinder decomposition is given in Eq. (1). It decomposes the mean difference in an outcome between two groups into one part due to differences in observable characteristics and another part due to differences in the effect of those characteristics. We focus on the first part as that can potentially be affected by policy intervention.

$$\overline{Y}^1 - \overline{Y}^2 = \left[(\overline{X}^1 - \overline{X}^2)\hat{\beta}^1\right] + \left[\overline{X}^2(\hat{\beta}^1 - \hat{\beta}^2)\right] \tag{1}$$

This study uses the Fairlie decomposition method (Fairlie, 2005) rather than the standard Oaxaca–Blinder method. We do so because our dependent variable is binary and the underlying regressions in Fairlie decomposition use probit estimation. Applying the Oaxaca-Blinder method, which utilizes Ordinary Least Squares regression, poses the risk that it will produce unreasonable results that lie outside the unit interval. Further, although the linear Oaxaca-Blinder method is a good approximation for binary variables, the Fairlie method is preferred when the racial gaps are in the tails of the distribution[2] (Fairlie, 2005). This approach has been used to decompose gender differences in obesity in South Africa (Averett et al., 2014).

We first estimate reduced-form health production functions for non-Hispanic white (white henceforth), non-Hispanic black (black henceforth), and Hispanic adolescents, using a probit regression. In the exposition below, we focus on the health differences between whites and blacks.

$$H_i^w = F(X_i^w \beta w + u_i^w) \tag{2}$$

$$H_i^b = F(X_i^b \beta b + u_i^b) \tag{3}$$

In Eqs. (2) and (3) $F$ represents the standard normal cumulative density function and the $X$'s represent inputs in the health production function. We use an extensive set of characteristics to measure ability, socioeconomic status, smoke exposure, and school peer characteristics as explanatory variables, which are described in the next section.

The average health differential between whites and blacks can be decomposed into 2 parts:

$$\overline{H}^w - \overline{H}^b = \left[\sum_{\forall N^w} \frac{F(X_i^w \hat{\beta}^w)}{N^w} - \sum_{\forall N^b} \frac{F(X_i^b \hat{\beta}^w)}{N^b}\right] + \left[\sum_{\forall N^b} \frac{F(X_i^b \hat{\beta}^w)}{N^b} - \sum_{\forall N^b} \frac{F(X_i^b \hat{\beta}^b)}{N^b}\right] \tag{4}$$

The first term represents the portion of the difference in health that is due to differences in the distribution of the health inputs

[2] By the Central Limit Theorem we know that the racial gap in the rate of smoking, $(\overline{Y_W} - \overline{Y_B})$ is normally distributed with mean zero (null hypothesis), and the further away the gap is from zero the more it is said to lie in the tails. A back-of-the-envelope calculation using descriptive statistics in Table 1 shows that white-black gap and white-Hispanic gaps are respectively 24.4–30.8 and 9.7–10.6 standard deviations (z-scores under the assumptions of unequal and pooled variances, respectively) from the mean. This shows why the Fairlie method applies to the current analysis.