Journal of Health Economics 30 (2011) 647–659

Contents lists available at ScienceDirect

Journal of Health Economics

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

Genetic lotteries within families

Jason M. Fletcher[a,d,*], Steven F. Lehrer[b,c]

[a] Yale University, School of Public Health, 60 College Street, #303, New Haven, CT 06520, United States
[b] School of Policy Studies and Department of Economics, Queen's University, Kingston, Ontario K7L 3N6, Canada
[c] NBER, United States
[d] Columbia University, Robert Wood Johnson Foundation Health & Society Scholar, 420 W 118th Street 8th Floor, Mail Code 3355, New York, NY 10027, United States

ARTICLE INFO

Article history:
Received 12 May 2010
Received in revised form 2 March 2011
Accepted 20 April 2011
Available online 17 May 2011

JEL classification:
I12
I21

Keywords:
Genetics
Sibling fixed effects
Childhood health
Mental health

ABSTRACT

Drawing on findings from the biomedical literature, this paper introduces the idea that specific exogenously inherited differences in the genetic code between full biological siblings can be used to test within-family estimators and potentially improve our understanding of economic relationships. These points are illustrated with an application to identify the causal impact of several poor health conditions on academic outcomes. We present evidence that family fixed effects estimators by themselves cannot fully account for the endogeneity of poor health when estimating education production functions. Further, our analysis elucidates the situations under which genetic markers can serve as instrumental variables for specific health conditions.

1. Introduction

One of the most controversial debates in academic circles concerns the relative importance of an individual's innate qualities ("nature") versus environmental factors ("nurture") in determining individual differences in physical and behavioral traits. For many years, researchers in the social sciences could only examine the relative importance of a multitude of environmental factors on various individual outcomes, as data on molecular genetic variation between individuals was unavailable. Yet, with the decoding of the human genome, this limitation no longer exists, and recent years have been characterized by substantial amounts of research in the biomedical literature examining whether specific point mutations in genetic code (aka single nucleotide polymorphisms (SNPs)) between dizygotic twins (among other family-based samples) are associated with specific diseases and outcomes. Findings from these studies have not only led to new drug discoveries but also improved diagnostic tools, therapies, and preventive strategies for a number of complex medical conditions. As clinical researchers identify unique molecular genetic bases for many complex health behaviors, diseases and other outcomes,[1] opportunities arise for social scientists to exploit this knowledge and use differences in specific sets of genetic information to gain new insights into a variety of questions as well as allow explicit tests of economic models.

In this paper, we exploit differences in genetic inheritance among children within the same family to first test whether family fixed effects estimators by themselves can fully solve the endogeneity problem when estimating the impact of several poor health conditions on academic outcomes.[2] Family fixed effects estimators allow researchers to simultaneously control (assuming constant impacts between family members) for both many common genetic factors and parental characteristics/behaviors, but does not provide any guidance as to why, within a twin pair, the subjects differed in explanatory characteristics and outcomes. Differences in genetic inheritance occur at conception and remain fixed between family members at every point in the lifecycle, irrespective of all nurture investments an individual faces (even those that occur in utero).[3] Since a great deal of variation in characteristics and out-

[1] Using similar methodologies, economists including Dreber et al. (2009) have begun to explore whether specific genetic loci are associated with financial risk preferences.
[2] Within economics, the use of family fixed effects regressions has a long history (e.g. Behrman and Taubman, 1976; Taubman, 1976).
[3] Genes consist of two alleles, and a child randomly inherits one of the two alleles from each parent at the time of conception. The child's genome consists of approx-

* Corresponding author. Tel.: +1 203 785 5760.
E-mail addresses: jason.fletcher@yale.edu (J.M. Fletcher), lehrers@post.queensu.ca (S.F. Lehrer).

0167-6296/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhealeco.2011.04.005

comes is found within families, exploiting the genetic processes that affect development (but are not self-selected by the individuals themselves) presents a potential strategy to identify differences within families.[4]

This study also contributes to a burgeoning literature that investigates what are the consequences of growing up in poor health for adolescent development. Prior empirical research in this area has attempted to estimate a causal link have either used a within-family strategy (i.e. Currie and Stabile, 2006; Fletcher and Wolfe, 2008, 2009; Fletcher, 2010) or instrumental variables approach (i.e. Ding et al., 2006, 2009 as well as Glewwe and Jacoby, 1995), and in general researchers find large negative impacts of poor health on academic outcomes. The genetic lottery empirical strategy combines both elements and identifies the causal impact of several poor health conditions on academic outcomes by exploiting exogenous variation in genetic inheritance between full biological siblings including dizygotic twins via a family fixed effects instrumental variables estimator.[5]

The genetic lottery identification strategy relies on knowledge of how specific genetic markers affect health and academic outcomes in adolescence. Despite a voluminous literature in several biomedical literatures, a consensus on how specific genetic markers operate still has not been reached. Thus, concerns could exist that, despite no detectable evidence in the biomedical literature to date,[6] the specific genetic markers we use in our analysis are not only related to poor health in adolescence but also to genetic factors that directly impact education outcomes. Further, we argue that theoretical channels from the science underlying the developmental process that threaten the exclusion restriction assumption are less severe when we exploit variation within families as opposed to variation across the full population, even if one were to proxy for population stratification. In our analysis, we find that the impact of poor mental health outcomes on academic achievement is substantial. Our preferred estimates examine the relationship with a sample consisting only of same sex dizygotic twins, and they indicate that inattention leads on average to a one standard deviation decrease in academic performance.[7]

Our empirical analysis reaches three major conclusions for the health economics literature. First, we conduct a variety of specification tests which indicate that family fixed effects estimators by themselves cannot fully account for the endogeneity of poor health. Specifically, with multiple samples and specifications we find that Hausman tests between the family fixed effects and family fixed effects IV model reject the consistency of the former. This rejects the assumption that the observed differences in health and education outcomes between full biological siblings can be treated as random within families.

Second, our analysis examines the potential of using genetic markers as instrumental variables when estimating the causal impact of specific health conditions. Intuitively, by accounting for family fixed effects, the genetic lottery strategy uses the random allocation of genes at conception in an analogous way to treatment assignment in a randomized experiment. We discuss that if researchers ignore common unobserved family factors in their analyses, the requirements for the genetic marker to serve as a plausibly random source of variation are not only theoretically more severe, but also that the interpretation of the estimates may differ. Empirically, we demonstrate that the sensitivity of our empirical results to the degree in which the exclusion restriction assumption is potentially violated also depends upon whether family factors are controlled for in the analyses. While we find that our preferred results using a subsample of same-sex full biological siblings are not sensitive to the plausibility of the instruments at reasonable levels, we find that when allowing for potential exogeneity error that the length of the confidence intervals for every health condition we investigate increases at a larger rate when family fixed effects are excluded from the specification; with the sole exception of inattention. Taken together, these results first caution against the use of genetic markers as instruments in the absence of family fixed effects. Second and more importantly it also stresses the importance of conducting sensitivity analysis with respect to the plausibility of the exclusion restriction assumption.

Last, we present evidence that measurement error is a serious challenge facing researchers interested in estimating the impact of specific health conditions. Measurement error arises not only because of potential errors in self reporting of one's health status, but also due to the presence of potentially unobserved comorbid conditions. In our empirical investigation, we find that if one were to include only one poor health condition in the specification that the sensitivity analysis fails at very mild levels of exogeneity error. The large difference in results from the sensitivity analyses as we move from our complete specification of the health vector to less rich subsets increases our confidence in both our main results and specification of the health vector. Since poor health conditions often occur simultaneously and it is hard to identify a unique source of genetic or environmental variation to identify the impact of specific disorders, the potential presence of unmeasured comorbid conditions has implications for any use of the instrumental variable strategies within the health economics literature.

The rest of the paper is organized as follows. In Section 2, we provide an overview of the data we employ in the study. We also review the scientific literature linking the genes in our dataset to health behaviors and health outcomes. The empirical framework that guides our investigation and our identification strategy is described in Section 3. The empirical results are presented and discussed in Section 4. A concluding section summarizes our findings and discusses the potential challenges of using genetic markers in future research.

2. Data

This project makes use of the National Longitudinal Study of Adolescent Health (Add Health), a nationally representative

imately 3.2 billion base pairs, along which there are 9.2 million candidate SNPs, which are specific locations where a mutation in the genetic code may influence an individual's susceptibility to various developmental outcomes such as developing an illness. In other words, our empirical strategy exploits these strictly exogenous differences in the coding of a specific marker between full siblings and can intuitively be viewed as an experiment in "nature".

[4] Ding et al. (2006, 2009) was the first empirical study within economics to explicitly use differences in genetic information across individuals as an instrumental variable in estimating the effects of poor health on high school grade point average. Norton and Han (2008) use genetic information to attempt to estimate the impact of obesity on employment. Neither study exploited variation in genetic inheritance within families (the "genetic lottery"), which we show to be important empirically and improves the plausibility of the exclusion restriction.

[5] In a companion paper Fletcher and Lehrer (2009b) illustrate the empirical strategy developed in this paper with an application to understanding the causal determinants of years of schooling completed.

[6] Plomin et al. (2006) and de Quervain and Papassotiropoulos (2006) present recent surveys on which genes are believed to be directly associated with intelligence and memory ability respectively. Using maps of the location between these genes and the specific genetic markers in our study, we find no evidence suggesting that linkage in inheritance is unlikely. Researchers have found no direct links between several of the genes in this study and intelligence (i.e. Moises et al., 2001) or cognitive ability (e.g. Petrill et al., 1997), and we hypothesize that if a link exists, it operates through specific health measures.

[7] Similarly large negative impacts of poor health on measures of later cognitive achievement have been found in studies that exploit shocks to an individual's prenatal conditions such as in utero exposure to the flu (Almond, 2006) and low levels of radiation (Almond et al., 2009).