Environmental Research 137 (2015) 108–119

Contents lists available at ScienceDirect

Environmental Research

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

Quantile regression in environmental health: Early life lead exposure and end-of-grade exams

Sheryl Magzamen [a,*], Michael S. Amato [b], Pamela Imm [c], Jeffrey A. Havlena [d], Marjorie J. Coons [c], Henry A. Anderson [c], Marty S. Kanarek [e,f], Colleen F. Moore [b,g]

[a] Department of Environmental and Radiological Health Sciences, Colorado State University, 1681 Campus Delivery, Fort Collins, CO 80523-1681, United States
[b] Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, WI 53706, United States
[c] Bureau of Environmental and Occupational Health, Wisconsin Department of Health Services, 1 West Wilson Street, Madison, WI 53703, United States
[d] Department of Surgery, University of Wisconsin, 600 Highland Ave, Madison, WI 53792, United States
[e] Department of Population Health Sciences, University of Wisconsin, 707 WARF, 610 Walnut Street, Madison, WI 53726, United States
[f] Nelson Institute for Environmental Studies, University of Wisconsin, 550 North Park Street, 122 Science Hall, Madison, WI 53706, United States
[g] Department of Psychology, Montana State University, PO Box 173440, Bozeman, MT 59717, United States

ARTICLE INFO

Article history:
Received 7 October 2014
Received in revised form
14 November 2014
Accepted 2 December 2014

Keywords:
Education
Lead exposure
Pediatrics
Quantile regression
Socioeconomic status
Urban health

ABSTRACT

Conditional means regression, including ordinary least squares (OLS), provides an incomplete picture of exposure–response relationships particularly if the primary interest resides in the tail ends of the distribution of the outcome. Quantile regression (QR) offers an alternative methodological approach in which the influence of independent covariates on the outcome can be specified at any location along the distribution of the outcome. We implemented QR to examine heterogeneity in the influence of early childhood lead exposure on reading and math standardized fourth grade tests. In children from two urban school districts ($n = 1{,}076$), lead exposure was associated with an 18.00 point decrease (95% CI: −48.72, −3.32) at the 10th quantile of reading scores, and a 7.50 point decrease (95% CI: −15.58, 2.07) at the 90th quantile. Wald tests indicated significant heterogeneity of the coefficients across the distribution of quantiles. Math scores did not show heterogeneity of coefficients, but there was a significant difference in the lead effect at the 10th ($\beta = -17.00$, 95% CI: −32.13, −3.27) versus 90th ($\beta = -4.50$, 95% CI: −10.55, 4.50) quantiles. Our results indicate that lead exposure has a greater effect for children in the lower tail of exam scores, a result that is masked by conditional means approaches.

1. Introduction

Epidemiological research often focuses on inference for high or low values in a population distribution, e.g. high body mass index, low birth weight, high blood pressure, or reduced lung function. However, the preponderance of statistical approaches in epidemiology employ conditional-mean modeling, such as ordinary least squares regression (OLS), which summarizes the relationship between the dependent variable and explanatory variables by describing the mean of the response for each fixed value of the predictors (Hao and Naiman, 2007). Conceptually, the primary limitation of conditional-mean modeling is that it cannot be easily extended to describe relationships for non-central locations, such as in the upper or lower tails of a distribution. Statistically, a main concern with conditional-means approaches is that these models provide an incomplete picture of the exposure-outcome function for regression models with heterogeneous variances. In many cases there may not be one, unique slope that effectively characterizes the changes across the probability distribution (Cade and Noon, 2003). In these instances, the assumptions for conditional-means models like ordinary least squares or other generalized linear model techniques are violated. Focusing primarily on changes in the mean effect may underestimate, overestimate, or fail to distinguish real changes in other locations of the distribution (Terrell et al., 1996; Cade et al., 1999; Cade and Noon, 2003).

Quantile regression, formalized by Koenker and Bassett (1978), can be considered a logical extension of linear regression. Rather than focusing on the change in the conditional mean of the dependent variable associated with a change in the explanatory variables, quantile regression models specify changes in the conditional quantile, or at any point along the distribution of the outcome (Hao and Naiman, 2007). Further, a set of equally spaced quantiles (e.g. every 5% of the population) can describe the shape of the distribution in addition to its central location (Hao and

* Corresponding author. Fax: +1 970 491 2940.
E-mail address: sheryl.magzamen@colostate.edu (S. Magzamen).

http://dx.doi.org/10.1016/j.envres.2014.12.004
0013-9351/

Naiman, 2007). Though quantile regression has been used extensively in the economics literature (Buchinsky, 1998; Eide and Showalter, 1998, 1999; Martins and Pereira, 2004; Machado and Mata, 2005), it is not a common analytic strategy in the epidemiologic literature (Beyerlein, 2014). We present an application of quantile regression to better understand the effect of early childhood lead exposure on elementary school end-of-grade examinations. We hypothesized that slope estimates obtained using ordinary least squares (OLS) methods would not fully describe the relationship between lead exposure and test scores at all points of the distribution of reading and math scores on a standardized end-of-grade exam. As socioeconomic status is considered to modify lead neurotoxicity (Bellinger, 2008), and is strongly associated with test scores (Duncan and Magnuson, 2005), we hypothesized that the effect of lead exposure on test scores would be greater for children scoring at lower quantiles, compared to children scoring at higher quantiles.

2. Method

2.1. Statistical approach

Quantile regression (QR) is a method for estimation of the functional relationships between outcomes and covariates at arbitrary quantiles of a conditional probability distribution (e.g. WKCE scores given gender, race, parental education). QR and linear regression can both be considered solutions to specific minimization problems. For example, OLS regression finds the center of the distribution (i.e. mean) by locating the point where the average squared deviance from data points of Y is minimized, and means are then interpreted as predicted values for individuals (Hao and Naiman, 2007). The median has a similar minimizing property, but a distance of Y to m is measured in absolute terms ($|Y-m|$) rather than squared distance. QR uses a generalized absolute minimization procedure to estimate predicted values for individuals that are conditional not only on explanatory variables, but also on the location of that individual in the distribution of outcomes scores Y. This approach is particularly relevant if the relationship between outcome and explanatory variables varies across the distribution of Y, or if the distribution of unobserved variables varies across Y.

As an illustrative case, if the outcome variable y represents exam score, and x_1 and x_2 represent gender and race, and e represents the error term, a general linear model to describe the relation between y and x for each child i can be written as

$$y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$$

However, if unmeasured or unobserved confounding results in heterogeneity of exam scores across race and gender, then for every point i in a distribution (e.g., 50% quantile, or the median), the outcome is not only conditioned on the values of the known parameters at that point in the distribution, but also the parameter estimates of the unknown or unmeasured covariates. Dividing the distribution of Y into discrete quantiles p, the observed value for each individual i at a specific quantile p is captured by the conceptual expression

$$y_i = \beta_{p0} + \beta_{p1} x_{i1} + \beta_{p2} x_{i2} + (\gamma_{po} + \gamma_{p1} x_{i1} + \gamma_{p2} x_{i2}) XU_i$$

in which the effect of each known variable X depends on both the quantile of the outcome, and also the distribution of unobserved variables U scaled by the parameters γ (Van Sickle et al., 2011). In other words, rather than a single summary statistic, QR provides parameter estimates for the outcome y conditioned on covariates X (both observed and unobserved) for every quantile specified (p). The estimates are considered to be semiparametric; the random error of the model has no parametric distributional form, though a parametric distribution is assumed for the deterministic part of the model (Cade and Noon, 2003).

2.2. Study population

The Wisconsin Children's Lead Levels and Educational Outcomes Project (CLLEO) is a collaborative study among the University of Wisconsin, the Wisconsin Childhood Lead Poisoning Prevention Program (WCLPPP), and the Wisconsin State Department of Public Instruction (DPI) to investigate the ongoing effects of early childhood lead exposure (Amato et al., 2012, 2013; Magzamen et al., 2013). The study protocol was approved by the University of Wisconsin Education Institutional Review Board. All parents/guardians signed informed consent documents prior to participation.

The study design and sample recruitment strategies have been described elsewhere (Magzamen et al., 2013). Briefly, the Wisconsin Knowledge and Concepts Exam (WKCE) records from DPI were merged with early childhood blood lead level (BLL) data from WCLPPP to define the target population: 1) born between January 1, 1996 and December 31, 2000; 2) record of BLL <5 μg/dL (not exposed group) or between 10 μg/dL and 20 μg/dL (exposed group) before the child's third birthday; 3) confirmed by DPI to have taken a 4th grade WKCE and 4) Milwaukee or Racine address at time of BLL testing. Milwaukee and Racine are the 1st and 5th largest cities in Wisconsin, respectively, and are high-risk communities for lead exposure (Wisconsin Department of Health and Family Services, 2008). The exposure categories were defined to compare children with BLLs below the level of quantification at the time of their testing with children whose BLLs were elevated but not sufficiently elevated to receive state-mandated home remediation or education interventions. These exposure classifications were developed prior to the 2012 Centers for Disease Control (CDC) policy change which adopted a new reference value of 5 μg/dL for blood lead levels, but still have relevance for state and local interventions which do not necessarily begin at the CDC reference values.

2.3. Data

Survey packets were mailed to all parents/guardians of children who met eligibility criteria and for whom we could obtain current contact information (Magzamen et al., 2013). Current mailing addresses were obtained either from Racine Unified School District or with assistance from the University of Wisconsin Survey Center. Survey packets included an invitational letter, two consent forms to allow release of educational data from DPI to the study, and a four-page questionnaire on demographics, educational experiences, health information, and environmental exposures. Upon return of the completed survey and consent form, parents were mailed a $5 monetary incentive. With parental consent, DPI provided exam scores, classifications for disability, designation as an English Language Learner, record of an Individualized Education Plan (IEP) (a tailored educational program for students with a special education designation), and enrollment in the federal free/reduced price lunch program (FRLP).

2.4. Analysis

Scaled scores on the math and reading components of the 4th grade WKCE were the primary outcomes. As described in Section 2.2, the study sample was selected such that BLLs fell either below quantification or in an elevated range that was not sufficient to receive lead education and abatement. Due to this selection strategy, and for consistency with other project analyses, lead