Biostatistics & Epidemiology

ISSN: 2470-9360 (Print) 2470-9379 (Online) Journal homepage: https://www.tandfonline.com/loi/tbep20

Using time-varying quantile regression approaches to model the influence of prenatal and infant exposures on childhood growth

Ying Wei, Xinran Ma, Xinhua Liu & Mary Beth Terry

To cite this article: Ying Wei, Xinran Ma, Xinhua Liu & Mary Beth Terry (2017) Using time-varying quantile regression approaches to model the influence of prenatal and infant exposures on childhood growth, Biostatistics & Epidemiology, 1:1, 133-147, DOI: 10.1080/24709360.2017.1358137

To link to this article: https://doi.org/10.1080/24709360.2017.1358137

BIOSTATISTICS & EPIDEMIOLOGY, 2017
VOL. 1, NO. 1, 133–147
https://doi.org/10.1080/24709360.2017.1358137

Taylor & Francis
Taylor & Francis Group

Using time-varying quantile regression approaches to model the influence of prenatal and infant exposures on childhood growth

Ying Wei [a], Xinran Ma [b], Xinhua Liu [a] and Mary Beth Terry [b,c]

[a]Department of Biostatistics, Columbia University, New York, NY, USA; [b]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA; [c]Herbert Irving Comprehensive Cancer Center of Columbia University, New York, NY, USA

ABSTRACT

For many applications, it is valuable to assess whether the effects of exposures over time vary by quantiles of the outcome. We have previously shown that quantile methods complement the traditional mean-based analyses, and are useful for studies of body size. Here, we extended previous work to time-varying quantile associations. Using data from over 18,000 children in the U.S. Collaborative Perinatal Project, we investigated the impact of maternal pre-pregnancy body mass index (BMI), maternal pregnancy weight gain, placental weight, and birth weight on childhood body size measured 4 times between 3 months and 7 years, using both parametric and non-parametric time-varying quantile regressions. Using our proposed model assessment tool, we found that non-parametric models fit the childhood growth data better than the parametric approaches. We also observed that quantile analysis resulted in difference inferences than the conditional mean models in three of the four constructs (maternal per-pregancy BMI, maternal weight gain, and placental weight). Overall, these results suggest the utility of applying time-varying quantile models for longitudinal outcome data. They also suggest that in the studies of body size, merely modelling the conditional mean may lead to incomplete summary of the data.

ARTICLE HISTORY
Received 17 October 2016
Accepted 15 July 2017

KEYWORDS
Quantile regression; time-varying coefficient model; longitudinal methods; childhood growth

1. Introduction

Many longitudinal studies of growth and other outcomes collect measurements and/or biospecimens repeatedly over time. Such longitudinal outcome data are usually analysed by repeated measure models [1,2] or time-varying coefficient models [3,4]. Time-varying coefficient models are more flexible in capturing the changing pattern of the association between the exposure and outcome compared with classical repeated measure models. Although repeated measure models and time-varying coefficient models are very useful, they both focus on estimating the effect of the covariates on the conditional mean of outcome Y, just like standard linear regression. When data exhibit skewness or heavy-tails, a

CONTACT Ying Wei yw2148@cumc.columbia.edu
Supplemental data for this article can be accessed at: https://doi.org/10.1080/24709360.2017.1358137.

conditional quantile model based inference could uncover important features missed by a model only estimating effects based on the conditional mean [5,6]. For example, we found that maternal pregnancy weight gain had a strong impact on the upper quantiles of the body mass index (BMI) of adult women, but the association was much smaller for the lower quantiles of adult BMI [7], and even within siblings, quantile-based approaches resulted in different inferences between prenatal exposures and childhood growth compared with models estimating the conditional mean of the growth outcome [8].

Epidemiologic studies have found that prenatal factors and birth factors such as maternal pre-pregnant BMI, maternal weight gain during pregnancy, placental weight, and child's birth weight are associated with growth throughout life [9–14]. Most epidemiologic studies of growth focus on modelling the conditional mean or categories of the outcome variable based on the marginal distribution of the outcome Y. Here, we extend our prior work, which investigated quantile-specific effects of early life exposures on adult body size [7], by specifically looking at whether the effect of exposures of interest varies across time and whether this variation with time is different for each quantile of the outcome. Specifically, using prospective data from over 18,000 children collected as part of the National Collaborative Perinatal Project (NCPP), we examined the association between maternal pre-pregnancy BMI, maternal pregnancy weight gain, placental weight and birth weight, and childhood body size using both parametric and non-parametric time-varying quantile regression models.

2. Material and methods

2.1. Study population

This study uses prospectively collected data from families who were part of the NCPP [8,15]. The children were born from 1959 to 1966 at 14 centres across the United States and followed until 7 years of age. Of the 56,966 births, 20,523 children (10,327 boys; 10,196 girls) had complete information on weight and height measurements until 7 years. The remaining 36,443 births were not eligible for this study because they were missing at least one measurement for height, weight, and/or age for the following time periods (birth; 4 months; 1, 4, and 7 years) (73%), died at birth or before 7 years (8%), missing gender information (3%), missing other maternal or child covariate information (14%), or had a time measurement outside of the range considered reasonable for interpolation (2%). Further details on the NCPP are available in [8,15]. In addition, we also excluded children whose gestation length was shorter than 32 weeks and/or whose mother's pregnancy weight gain was less than 5 pounds. The final sample size is 9,362 boys and 9177 girls.

Reported characteristics were based on direct study measurement and maternal report at exam visit. Child physical measurements (weight and height) were taken at fixed intervals (birth; 4 months; 1, 4, and 7 years). The NCPP protocol also specified time for measurements. They could vary from the exact target ages. The ranges of the actual measurement times are 72–96 months at 7 years, 44–54 months at 4 years, 10–15 months at 1 year, and 3.1–5.4 months at 4 months. Birth weight was obtained within one hour of delivery by the NCPP observer of labour and delivery using calibrated scales, and birth length was obtained using a standardized procedure within 24 hours of birth and measured crown-heel. Placental weight (grams) was measured according to the Benirschke protocol [15].

2.2. Statistical considerations

2.2.1. Time-varying quantile regression model

We denote $Y(t)$ as a time-dependent outcome. In childhood growth applications, $Y(t)$ could be height or weight at the age of t. $(X_1, X_2, \ldots, X_p)$ are p covarites that may influence the outcome $Y(t)$. We then consider the following time-varying quantile regression model [16].

$$Q_\tau[Y(t)] = \beta_{\tau,0}(t) + X_1\beta_{\tau,1}(t) + \cdots + X_p\beta_{\tau,p}(t) \tag{1}$$

where $Q_\tau[Y(t)]$ stands for the τth conditional quantile of the outcome $Y(t)$ at time t. In model (1), we assume that the conditional quantile of Y is linear with covariates Xs. The coefficient $\beta_{\tau,j}(t)$ quantifies to which extent the factor X_j influences the τth quantile of Y, and how the impact varies with time t. A review on the existing literature of such varying-coefficient quantile regression model is provided in the Appendix. One can model $\beta_{\tau,j}(t)$ parametrically, assuming $\beta_{\tau,j}(t) = \beta_{\tau,j}(t, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a set of parameters determining the shape of $\beta_{\tau,j}(t)$. For example, if $\beta_{\tau,j}(t)$ is a quadratic function of time, then $\beta_{\tau,j}(t, \boldsymbol{\alpha}) = \alpha_{\tau,j,0} + \alpha_{\tau,j,1}t + \alpha_{\tau,j,2}t^2$. If no a-priori shape can be assumed, $\beta_{\tau,j}(t)$ can be well approximated using non-parametric functions such as smoothing splines and normalized B-splines [17], i.e. $\beta_{\tau,j}(t, \boldsymbol{\alpha}) \approx \boldsymbol{\pi}(t)^\top \alpha_{\tau,j}$ with $\pi(t)$ for a vector of B-spline basis functions of t, and $\alpha_{\tau,j}$ for the B-spline coefficients. The basis functions are determined by the placement of internal knots and the order of the polynomial functions (linear, quadratic, or cubic). We refer to [17] for the construction of B-spline basis functions. With these parameterizations, we can write $\beta_{\tau,j}(t)$ as $\beta(t, \boldsymbol{\alpha}_{\tau,j})$, which can be estimated using the existing quantile regression computation packages.

Suppose that $(Y_{i,j}, X_{i,j}), i = 1, \ldots, n, j = 1, \ldots m_i$ is a longitudinal sample consisting of n subjects, where m_i is the number of observations of the ith subjects, $Y_{i,j}$ is the jth outcome of the ith subject measured at time/age $t_{i,j}$, and $X_{i,j}$ is the set of covariates associated with $Y_{i,j}$. We then estimate quantile coefficient functions $\beta_{\tau,j}(t)$ by minimzing the following quantile regression loss function over $\boldsymbol{\alpha}$:

$$\sum_{i=1}^{n} \rho_\tau\left(y_{i,j} - \beta(t_{i,j}, \alpha_{\tau,0}) - X_{1,i,j}\beta(t_{i,j}, \alpha_{\tau,1}) - \cdots - X_{p,i,j}\beta(t_{i,j}, \alpha_{\tau,p})\right),$$

where $\rho_\tau(u) = u(\tau - I\{u < 0\})$ is the quantile regression loss function. Denote $\widehat{\alpha_{\tau,j}}$ as the resulting estimated coefficients, then the estimated quantile coefficient function can be written as $\widehat{\beta}_{\tau,j}(t) = \beta(t, \widehat{\alpha_{\tau,j}})$.

2.2.2. Visual assessment of model fitness

Non-parametric approach requires fewer assumptions about the shapes of coefficient functions, and will hence ensure a proper fitting. Theoretically, any smooth function can be approximated by a B-spline approximation. As a cost for the added flexibility, the estimation of $\beta_{\tau,j}(t)$ is not efficient. In contrast, parametric approaches restrict the candidate functions of $\beta_{\tau,j}(t)$ to be in a smaller family of functions. The resulting estimates are more efficient, but are more likely to be biased. In what follows, we propose a convenient visual assessment of model fitness, based on which one could select models. Several model assessment tools for quantile models have been developed [18–20], although these methods generally compare relative fitness between two nested models and are hard to compute in practice. Here, we propose a simple and quick assessment to evaluate whether a model is adequate. The underlying logic of the proposed visual assessment is as follows: if a model for τth conditional quantile fits the data well, then, for each time subinterval, we would expect the proportion of the residuals within that interval to be sufficiently close to its nominal level τ. For a comprehensive assessment of conditional model fitness, the entire time interval is partitioned into as many subintervals as possible while each subinterval has sufficient data points ($n > 50$). For example, in our study, the measurement times are clustered around birth, 4 months, 1 year, 4 years, and 7 years due to the design of the study; we define fixed time intervals between these clusters. Let n_k be the number of residuals and p_k be the proportion of negative residuals in the kth interval. We can normalize the proportion p_k by a score $z_k = \frac{p_k - \tau}{\sqrt{\tau(1-\tau)n_k}}$, which approximates a standard normal distribution when the model fits data well. Inadequate fit for a given k-interval occurs when the z_k score has an absolute value exceeding 1.96. An index plot of z_k vs. k will identify the subintervals with poor overall fit. This approach does not depend on the model format and therefore can be applied to select models of any form and importantly does not require models to be nested. For example, one can select parametric or non-parametric $\beta_{\tau,j}(t)$ by comparing the model fitness, i.e. if the model fitness from a model with parametric $\beta_{\tau,j}(t)$ is comparable to that from the model with non-parametric $\beta_{\tau,j}(t)$, then we select the parametric model because it is more efficient. Otherwise, the non-parametric model is preferred for its flexibility to reduce bias.

2.2.3. Testing hypotheses on the influence of predictors to the growth path

The primary interest under model (1) is to determine whether an exposure has any influence on the τth conditional quantile of Y at any time, i.e. to test whether $\beta_{\tau,j}\ (t) = \ 0$ for all t. We used a Rank score test [21] rather than the more computationally burdensome bootstrapping approach (see reference [22] for the construction of rank score test statistics with longitudinal outcomes).

2.2.4. Modelling childhood growth using NCPP

We assessed the impact of birth weight, placental weight, maternal pre-pregnancy BMI, and maternal pregnancy weight gain on the childhood weight measurements from 3 months to 7 years with the quantile regression model:

$$Q_\tau\ [Y_{i,j}] = \beta_{\tau,0}\ (t_{i,j}) + X_{\mathrm{i},1}\beta_{\tau,1}(t_{i,j}) + X_{\mathrm{i},2}\beta_{\tau,2}(t_{i,j}) + X_{\mathrm{i},3}\beta_{\tau,3}(t_{i,j}) + X_{\mathrm{i},4}\beta_{\tau,4}(t_{i,j}) + Z_{\mathrm{i}}\gamma(t_{i,j}) \tag{2}$$

where $t_{i,j}$ is the jth measurement time of the ith child, $Y_{i,j}$ is the ith child's weight measured at age $t_{i,j}$, $X_{i,1}-X_{i,4}$ are birth weight, placental weight, maternal pre-pregnancy BMI, and maternal pregnancy weight gain for the ith child, and Z_i is the gestational age.

We applied model (2) to boys and girls separately by two approaches, parametric and non-parametric. With the parametric approach, we model $\beta_{\tau,j}(t)$ as polynomial functions of age t, i.e. $a + bt + ct^2$. With the non-parametric approach, we used linear spline with internal knots at 3.76, 11.77, 48, and 84 months. The internal knots were selected to be the medians of the four clusters of time points at which the measurements were available. We also considered spline functions with higher order (quadratic and cubic). Because the model fit was worse, we only present the linear spline. The *R* codes for model assessment and rank-type inference we used are available in supplementary online materials.

Remark: Where to place internal knots is a fundamental task in B-spline curve fitting. A common and convenient practice is to use uniformly distributed knots and use the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) to determine the number of internal knots. Such uniform knots selection may not be a good choice for human growth curves, since the underlying growth rates vary over time. Ideally, one would place more internal knots during infancy and puberty to capture the rapid and uneven growth during those times, while leave fewer knots in other growth periods. In the NCPP data, the body size measures are clustered around birth, 4 months, 1, 4, and 7 years by study design. The estimation algorithm will fail if there are no data between two internal knots. Hence, we choose the knots to the medians of the four clusters of time points, so that we have sufficient number of data in between internal knots. In addition, the NCPP study followed the children at the ages of birth, 4 months, 1, 4, and 7 years, since they are critical growth periods when the growth rates were expected to change. Placing the knots around those target times could also ensure a good capture of the changings of growth rates.

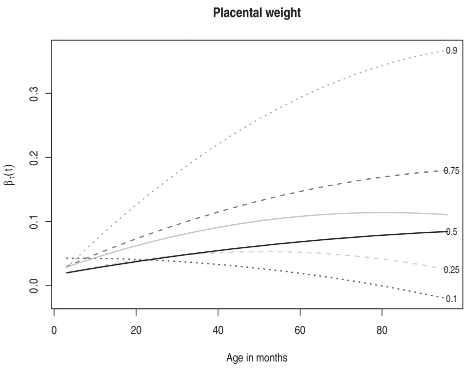
3. Results

Descriptive statistics are summarized in Table 1. The sample size (boys: 9362; girls: 9177) and racial breakdown for boys and girls were similar (boys: 50%, 46%, 4%; girls: 48%, 48%, 4% for white, black, and other respectively).

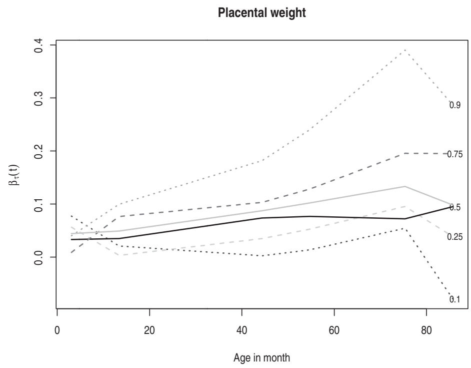
We observed considerable differences between the estimated coefficient functions from parametric and non-parametric approaches. We use the placental weight as an example to illustrate the difference. Based on the parametric model, the placental weight effect on the 90th percentile of boy's weight increased steadily with age as shown in Figure 1(a); however, the non-parametric model suggested that placental weight does not have a substantial impact on boy's childhood weight before 50 months (see Figure 1(b)). Figure 1(c)

Table 1. Descriptive statistics, NCPP sample.

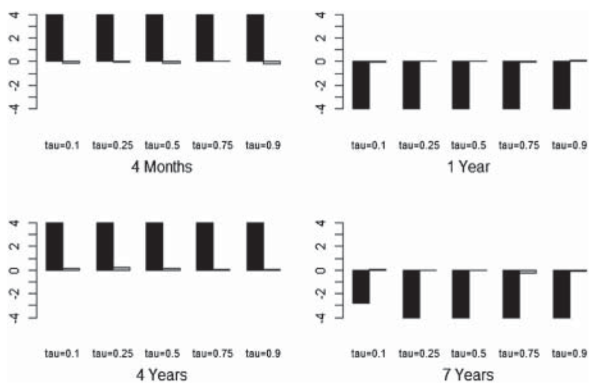
	Boys, N = 9362		Girls, N = 9177	
	Mean (SD)	Median (range)	Mean (SD)	Median (range)
Maternal pre-pregnant BMI	22.55 (3.93)	21.76 (11.86–47.75)	22.59 (3.95)	21.80 (13.17–49.92)
Maternal weight Gain (lb)	22.75 (8.68)	22.00 (6–50)	22.26 (8.71)	21 (6–50)
Gestational age (week)	39.64 (2.46)	40.00 (33–51)	39.78 (2.47)	40 (33–51)
Placental weight (g)	441.48 (94.77)	434 (120–1090)	436.47 (96.61)	425 (28–999)
Birth weight (kg)	3.27 (0.51)	3.26 (1.25–5.56)	3.14 (0.50)	3.15 (1.08–5.56)
Weight at 4 months (kg)	6.59 (0.87)	6.58 (3.35–10.49)	6.08 (0.80)	6.04 (2.75–10.12)
Weight at 1 year (kg)	10.08 (1.18)	10.00 (5.6–16.8)	9.47 (1.14)	9.40 (5.5–15.0)
Weight at 4 years (kg)	16.77 (2.15)	16.60 (9.1–36.1)	16.21 (2.22)	15.90 (9.8–35.4)
Weight at 7 years (kg)	23.88 (3.94)	23.30 (12.2–61.4)	23.46 (4.20)	22.70 (13.5–54.9)



(a) Estimated placental weight effect on boy's weight using parametric model at quantile levels 0.1, 0.25, 0.5, 0.75 and 0.9.



(b) Estimated placental weight effect on boy's weight using non-parametric model at quantile levels 0.1, 0.25, 0.5, 0.75 and 0.9.



(c) Comparison of model fitness between parametric and nonparametric models.

Y-axis represents the Z values of the model fitness scores. Black bars denote parametric method and grey bars denote non-parametric method (Z values above 4 and below -4 are truncated at 4 and -4, respectively).

Figure 1. Comparison of the estimated placental effects from parametric and non-parametric models.

reports the comparison of goodness of fit for the parametric model (black) and non-parametric (grey) models for boys. We partitioned the seven-year study period into four time intervals starting at 2 months, 7 months, 2.5 years, and 4.5 years, so that each interval contains the measures around one of the four target times. In each interval, we calculated the goodness-of-fit scores for both parametric and non-parametric models. The goodness-of-fit scores from the non-parametric coefficient model were all bounded within a window of $(-1, 1)$, indicating a reasonably good model fit in all the time intervals. In contrast, the goodness-of-fit scores from the parametric models went beyond the range of (−4, 4), indicating lack-of-fit in all the time intervals. In particular, the parametric models overestimated the associations for quantiles at years 1 and 7, and underestimated the associations for quantiles at 4 months and 4 years.

Given its model superiority, the non-parametric model should be used to estimate the impact of the four main exposures on childhood weight. Figures 2 (boys) and 3 (girls) describe the estimated coefficient functions over time from the non-parametric models at the 0.1th, 0.25th, 0.5th, 0.75th, and 0.9th quantiles. For comparison, the time-varying coefficient function for the conditional mean model is highlighted in solid grey.

Moreover, Table 2 lists the estimated covariate effects on the selected quantiles of childhood weight at ages 4 month, 1 year, 4 years, and 7 years, together with their p values from testing whether the covariate has any association with the τth quantile of childhood weight at any time, i.e. testing $\beta_{\tau,j}(t) = 0$ for all t.

Combining on Figures 2 and 3 and Table 2, birth weight is significantly associated with the childhood weight, and such effect was seen for both boys and girls at all quantile levels with very little variation in the effect size by quantile. It suggests that birth weight only impacts the mean of the childhood weight, but does not affect the variance and the distribution of childhood weight. For the three other variables of interest, we did observe quantile-specific differences, and the time at which these differences emerge varied by exposure. For example, placental weight has strong association with the upper quantiles of boys' weight, and the association is more evident after 50 months. Increasing the placental weight by 0.1 kg will result in an increase of 0.19 kg in the 0.9th quantile of boy's weight at age 4 years, and an increase of 0.45 kg in the 0.9th quantile of boy's weight at age 7. That indicates boys born with high placental weight is more likely to have heavy weight between ages 4 and 7, even though their weight in average is comparable to general population. A similar pattern was also found among girls, although the impact is less significant, also less evident until later childhood around year 7. The quantile-specific differences in maternal pre-pregnancy weight emerged in early childhood (by 4 years of age) for both boys and girls. The association is much stronger with upper quantiles of weight than the lower quantiles. Note that the quantiles classify the population by weight, and the observed quantile-varying association pattern indicates that maternal pre-pregnancy BMI has stronger impact in children of larger body size. Finally, we did not observe quantile-specific differences in the associations between maternal weight gain during pregnancy and childhood weight gain until later in childhood (approximately 70 months).

4. Discussion

We formally compared quantile regression models with time varying coefficients for longitudinal outcomes with conventional approaches estimating the effects on the conditional mean over time. We applied both models to a prospectively collected data with over 18,000 children. Overall, we found that maternal pre-pregnancy body size, maternal pregnancy weight gain, birth weight, and placental weight are significantly associated with childhood body size. That is consistent with other studies and points to the importance of earlier life influences in predicting childhood weight [22–25], but quantile-based approaches provide more comprehensive pictures on these associations. For three of the

Table 2. Estimate covariate effect at ages 4 months, 1 year, 4 years, and 7 years based on non-parametric quantile time-varying coefficient model, and their p values from testing the null hypothesis of no association ($\beta_{T,j}(t) = 0$).

				Placental weight		
		Estimated covariate effect (per 0.1 kg)				
Quantile level	Quantile level	t = 4 months	t = 1 year	t = 4 years	t = 7 years	p Value
Girls	0.1	−0.02	0.04	0.07	0.05	0.0355
	0.25	−0.11	0.03	0.08	0.07	0.0015
	0.5	−0.01	0.01	0.09	0.11	0.0103
	0.75	0.00	0.03	0.07	0.13	0.2383
	0.9	0.01	0.07	0.10	0.15	0.0566
Boys	0.1	0.02	0.02	0.00	0.07	0.07
	0.25	0.02	0.00	0.04	0.11	0.1161
	0.5	0.02	0.03	0.08	0.07	0.0147
	0.75	0.04	0.08	0.11	0.22	<0.0001
	0.9	0.03	0.10	0.19	0.45	<0.0001
				Birth weight		
		Estimated covariate effect (per kg)				
	Quantile level	t = 4 months	t = 1 year	t = 4 years	t = 7 years	p Value
Girls	0.1	0.85	0.86	1.25	1.63	<0.0001
	0.25	0.84	0.87	1.20	1.67	<0.0001
	0.5	0.87	0.97	1.27	1.71	<0.0001
	0.75	0.90	0.95	1.33	1.62	<0.0001
	0.9	0.91	0.95	1.34	1.36	<0.0001
Boys	0.1	0.85	0.85	1.23	1.39	<0.0001
	0.25	0.85	0.93	1.20	1.47	<0.0001
	0.5	0.86	0.95	1.29	1.49	<0.0001
	0.75	0.86	0.90	1.17	1.48	<0.0001
	0.9	0.90	0.78	1.15	1.23	<0.0001
				Maternal pre-pregnancy BMI		
		Estimated covariate effect				
	Quantile level	t = 4 months	t = 1 year	t = 4 years	t = 7 years	p Value
Girls	0.1	0.00	0.00	0.03	0.07	<0.0001
	0.25	0.00	0.00	0.04	0.12	<0.0001
	0.5	0.00	0.01	0.06	0.16	<0.0001
	0.75	0.01	0.02	0.09	0.27	<0.0001
	0.9	0.01	0.03	0.15	0.51	<0.0001
Boys	0.1	0.00	0.00	0.03	0.07	<0.0001
	0.25	0.00	0.01	0.04	0.11	<0.0001
	0.5	0.00	0.01	0.06	0.14	<0.0001
	0.75	0.01	0.02	0.08	0.22	<0.0001
	0.9	0.01	0.01	0.12	0.37	<0.0001
				Weight gain		
		Estimated covariate effect (per kg)				
	Quantile level	t = 4 months	t = 1 year	t = 4 years	t = 7 years	p Value
Girls	0.1	0.01	−0.00	0.01	0.03	0.0037
	0.25	0.00	0.01	0.02	0.04	<0.0001
	0.5	0.00	0.01	0.03	0.05	<0.0001
	0.75	0.01	0.01	0.02	0.08	<0.0001
	0.9	0.01	0.01	0.03	0.12	<0.0001
Boys	0.1	0.01	0.00	0.01	0.02	0.1860
	0.25	−0.00	0.00	0.02	0.04	0.0005
	0.5	0.00	0.01	0.02	0.04	<0.0001
	0.75	0.00	0.01	0.03	0.04	<0.0001
	0.9	0.00	0.01	0.03	0.07	0.0014

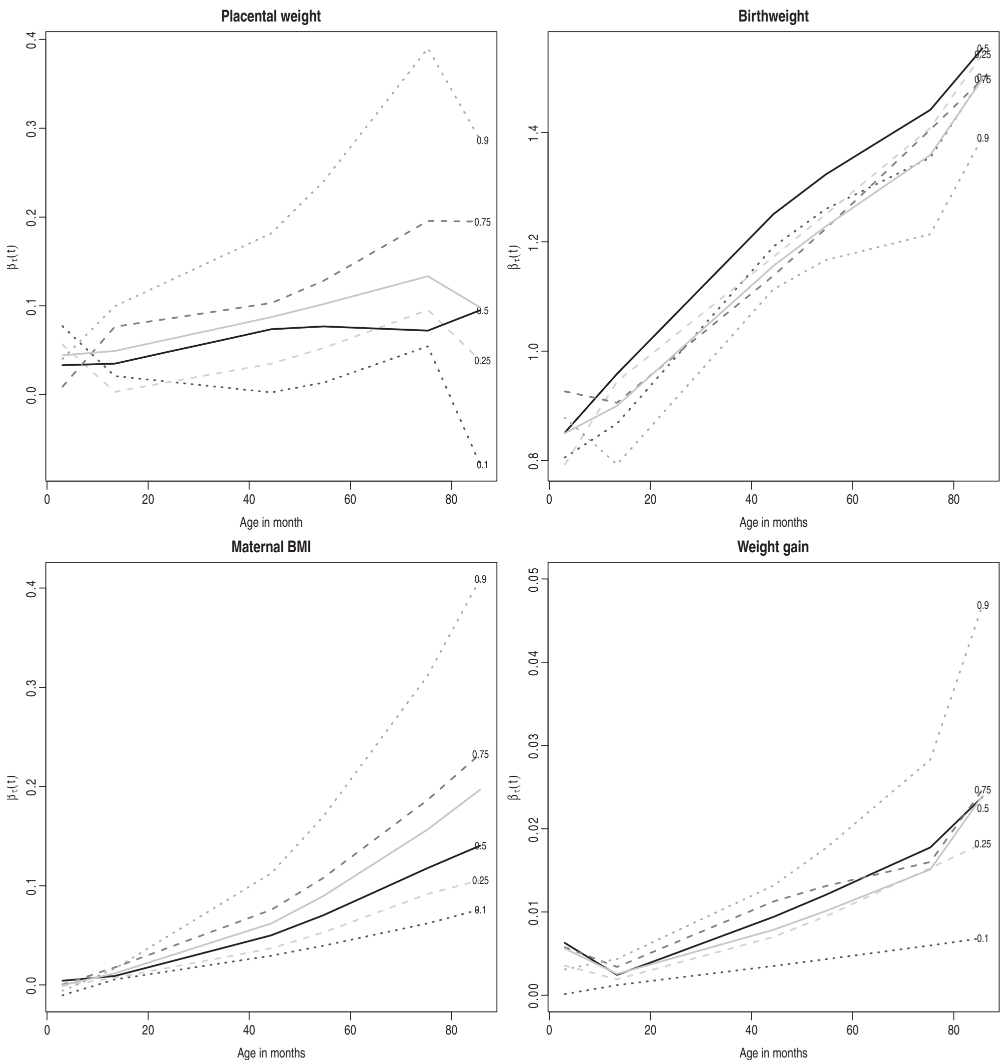


Figure 2. Estimated coefficient functions for boys based on non-parametric models.

four constructs (maternal pre-pregnancy BMI, maternal pregnancy weight gain, and placental weight), quantile regression models revealed differences in the magnitude of the association by quantile of childhood weight with the strongest associations for the 90th quantile of childhood weight and weaker, sometimes negative associations for the lower quantiles of childhood weight. This suggests that studies that focus on modelling the conditional mean for longitudinal data may miss differences that exist across the entire continuum of the outcome. For example, maternal pre-pregnancy BMI has a strong effect on the upper quantiles of childhood weight. However, if one focuses only on the conditional mean models, the impact of maternal pre-pregnancy BMI on childhood body size would miss this heterogeneity by quantile and result in underestimation of the effect of maternal BMI for some children.

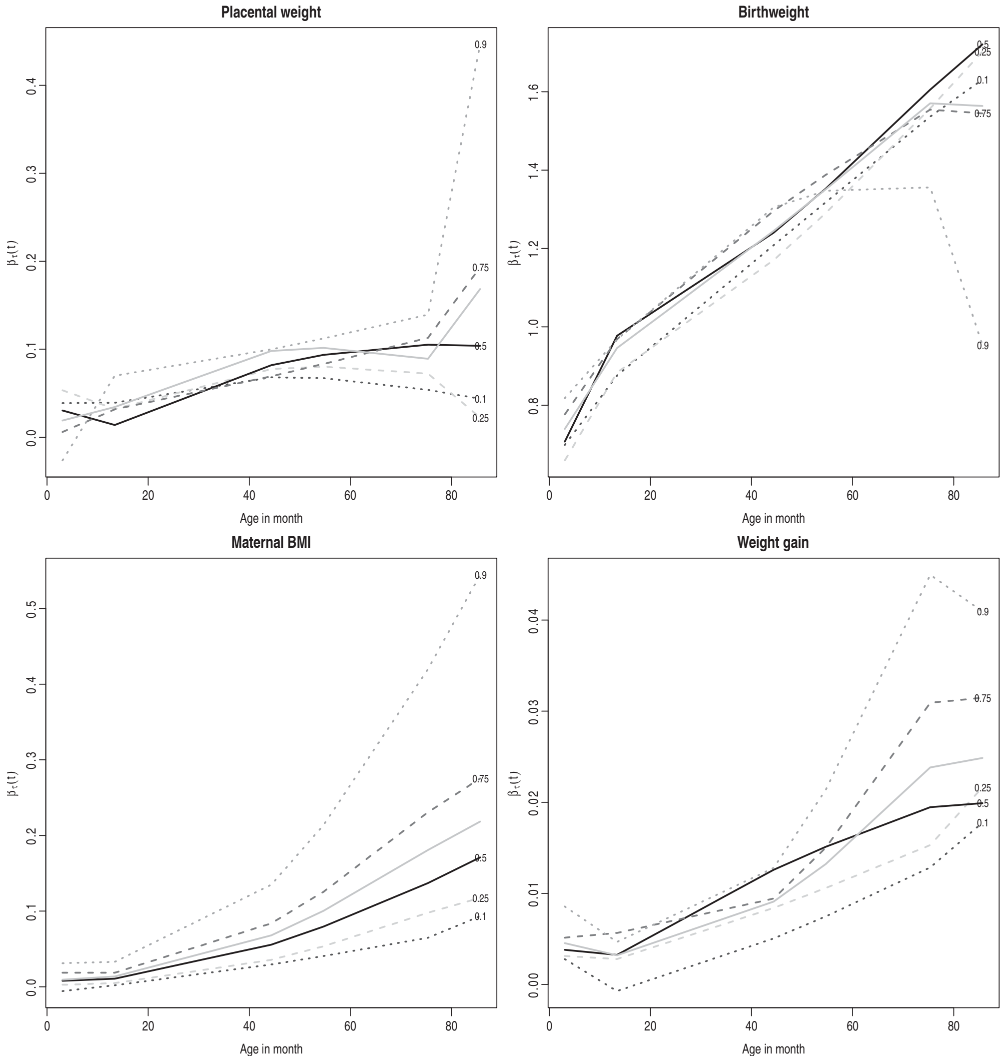


Figure 3. Estimated coefficient functions for girls based on non-parametric models.

We further compared non-parametric versus parametric approaches for time-varying quantile regression models and observed better overall fit with non-parametric based models. The parametric model in our setting is commonly used in repeated measure studies, often in the mean regression framework. Our results suggest that overall non-parametric quantile models fit better than parametric quantile models as well as the mean models that typically apply parametric approaches in estimating the model coefficients.

Time-varying coefficient models are suitable to use for analysis of data that have at least three repeated measurements over certain time interval. The model is particularly useful when covariate effects on the tails of the outcome are of interest, and when one expects the covariate effects to change over time. We were interested in modelling childhood body size where the lower and upper quantiles are of interest in addition to the overall mean. Biologically, it is entirely plausible for exposures to influence body size and other related outcomes only at the extreme and thus overall inference is enhanced through estimation of the magnitude of the time-varying association with quantiles of Y rather than just whether an association between X and Y exists.

Overall, we found that quantile regression approaches had sufficient flexibility to explore the association between prenatal factors and longitudinal measures of childhood weight. In comparison, with time-varying models of the conditional mean, quantile approaches have the added benefit of providing estimates across the continuum of the outcome variable. When the outcome is skewed or the data exhibits heteroscedasticity, the association between the outcome and exposures often varies across quantile levels. That is why we observed quantile-specific heterogeneity that differed by each exposure of interest in magnitude as well as time when it was observed. For example, when estimating the association between maternal pre-pregnancy BMI and childhood weight, quantile regression approaches revealed: (1) that maternal pre-pregnancy BMI had a stronger impact on the upper quantiles of childhood weight than on the lower quantiles; (2) that maternal pre-pregnancy BMI had stronger impact on childhood weight at a later stage of childhood than earlier stages of childhood. Moreover, the importance of maternal pre-pregnancy BMI increases more rapidly for upper quantiles than for the lower quantiles. These additional observations through time-varying quantile regression approaches provide a more comprehensive picture on how the prenatal factors are associated with childhood growth. As a price for this additional flexibility, the model is more complex than the classical growth models, and a large sample is needed for precise estimates at each quantile. In the end, multiple approaches may be useful to compare whether associations between exposures and longitudinal outcome differ by level of the outcome; time-varying quantile regression approaches can complement more conventionally used methods and enhance our understanding of associations between exposures and outcomes of interest.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Funded in part by the National Cancer Institute [grant number R01 CA 104842-03], [grant number K07CA90685]; National Institute of Child Health and Development [grant number P01 AG 023028-01]; National Science Foundation [grant number DMS-0906568], [grant number DMS-120923]; National Institute of Environment Health Sciences [grant number P30-ES09089]; National Human Genome Research Institute [grant number R03 HG 007443-01], [grant number R01 HG008980-01].

Notes on contributors

Ying Wei is currently a professor of biostatistics at Columbia University. She received her PhD degree in statistics from University of Illinois at Urbana-Champaign in 2004. Her substantive area of research is quantile regression, longitudinal data analysis, and semi-parametric modelling, with a special focus on their applications in medical, clinical, and public health studies.

Xinran Ma, MS, graduated from the Department of Biostatistics at Columbia University with Master of Science in biostatistics in 2016, and has been working as data analyst in the Department of Epidemiology at Columbia University since then.

Xinhua Liu, PhD, is currently a professor of biostatistics at Columbia University Medical Center and the senior biostatistician at Northern Manhattan Center of Environment Health Sciences of Columbia University. Her interdisciplinary collaborative research includes studies on metal and environmental health in children and adults. The recent work has focused on investigating impact of various environmental exposures on health outcomes and possible genetic modifications.

Mary Beth Terry, PhD, focuses her research on breast cancer and in the molecular epidemiology and life course methods of the disease, in particular. Dr. Terry lead family-based and birth cohort studies and has specific expertise in the area of early life exposures, including environmental exposures and disease susceptibility.

ORCID

Ying Wei http://orcid.org/0000-0002-7164-8109
Xinhua Liu http://orcid.org/0000-0001-9458-3427

References

[1] Fitzmaurice G, Davidian M, Verbeke G, et al., editors. Longitudinal data analysis. Boca Raton (FL): Chapman and Hall/CRC; 2008.
[2] Vonesh EF, Chinchilli VG. Linear and nonlinear models for the analysis of repeated measurements. London: Chapman and Hall; 1997.
[3] Hastie T, Tibshirani R. Varying-coefficient models. J R Stat Soc Ser B. 1993;55:757–796.
[4] Hoover DR, Rice JA, Wu CO, et al. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika. 1998;85(4):809–822.
[5] Koenker R. Quantile regression. New York (NY): Cambridge University Press; 2005.
[6] Koenker R. Quantile regression for longitudinal data. J Multivariate Anal. 2004;91(1):74–89.
[7] Terry MB, Wei Y, Essenman D. Maternal, birth, and early life influences on adult body size in women (with comments). Am J Epidemiol. 2007;166(1):5–13.
[8] Terry MB, Wei Y, Esserman D, et al. Pre- and postnatal determinants of childhood body size. J Dev Orig Health Dis. 2011;2(2):99–111.
[9] Strauss RS. Effects of the intrauterine environment on childhood growth. Br Med Bull. 1997;53:81–95.
[10] Oken E, Gillman MW. Fetal origins of obesity. Obesity Research. 2003;11:496–506.
[11] Whitaker RC, Dietz WH. Role of the prenatal environment in the development of obesity. J Pediatr. 1998;132:768–776.
[12] Monteiro PO, Victora CG. Rapid growth in infancy and childhood and obesity in later life – a systematic review. Obes Rev. 2005;6:143–154.
[13] Ong KK. Size at birth, postnatal growth and risk of obesity. Horm Res. 2006;65(Suppl 3):65–69.
[14] Gillman MW, Kleinman K. Invited commentary: antecedents of obesity-analysis, interpretation, and use of longitudinal data. Am J Epidemiol. 2007;166:14–16.
[15] Broman S. The collaborative perinatal project: an overview. In: Mednick SA, Harway M, Finello KM, editors. Handbook of longitudinal research. Vol. I. New York (NY): Praeger Publishers; 1984. p. 185–227.
[16] Kim MO. Quantile regression with varying coefficients. Ann Stat. 2007;35(1):92–108.
[17] De Boor C. On calculating with B-splines. J Approx Theory. 1972;6(1):50–62.

[18] Koenker R, Machado JA. Goodness of fit and related inference processes for quantile regression. J Am Stat Assoc. 1999;94:1296–1310.
[19] He X, Zhu LX. A lack-of-fit test for quantile regression. J Am Stat Assoc. 2003;98:1013–1022.
[20] Salibián-Barrera M, Wei Y. Weighted quantile regression with nonellipically structured covariates. Can J Stat. 2008;36:1–17.
[21] Gutenbrunner C, Jureckova J. Regression quantile and regression rank score process in the linear model and derived statistics. Ann Stat. 1991;20:305–330.
[22] Wang H. Inference on quantile regression for heteroscedastic mixed models inference on quantile regression for heteroscedastic mixed models. Stat Sin. 2009;19:1247–1261.
[23] Binkin NJ, Yip R, Fleshood L, et al. Birth weight and childhood growth. Pediatrics. 1988;82:828–834.
[24] Li C, Goran MI, Kaur H, et al. Developmental trajectories of overweight during childhood: role of early life factors. Obesity. (Silver Spring). 2007;15:760–771.
[25] Salsberry PJ, Reagan PB. Dynamics of early childhood overweight. Pediatrics. 2005;116:1329–1338.
[26] Chiang CT, Rice JA, Wu CO. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. J Am Stat Assoc. 2001;96(454):605–619.
[27] Fan J, Zhang W. Statistical estimation in varying coefficient models. Ann Stat. 1999;27: 1491–1518.
[28] Fan J, Zhang W. Simultaneous confidence bands and hypothesis testing in varying−coefficient models. Scand J Stat. 2000;27(4):715–731.
[29] Honda T. Quantile regression in varying coefficient models. J Stat Plan Inference. 2004;121 (1):113–125.

Appendix: Introduction to time-varying coefficient quantile regression

Since proposed in [3], time-varying coefficient models have been an appealing modelling tool for analysing longitudinal data due to its flexibility in capturing the dynamic associations between exposures and outcome. Suppose $Y(t)$ is an outcome of interest measured at time t, and $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ is a p-dimensional covariate. A time-varying coefficient model can be written as

$$Y(t) = \beta_0(t) + X_1\beta_1(t) + \cdots + X_p\beta_p(t) + \varepsilon(t)$$

where $\varepsilon(t)$ is the random error with $E\{\varepsilon(t)\} = 0$, and $\beta_k(t)$ measures the impact of an unit increase of X_k on the mean of Y at time t. The estimations of $\beta_k(t)'s$ are often achieved through non-parametric approximations. We refer to [3,26–28] for various versions of estimation algorithms and the asymptotic properties of their resulting estimations.

In recent years, quantile regression emerged as an appealing alternative to least squares in many applications, especially when outcomes are non-normally distributed, and data are heteroscedastic. Time-varying coefficient models have been widely applied to analysis of conditional means, but have only recently been applied to analysis of conditional quantiles. Honda [29] and Kim [16] extended varying-coefficient model to quantile regression, and assumed that

$$Q_\tau\{Y(t)\,|\,\mathbf{X}\} = \beta_{\tau,0}(t) + X_1\beta_{\tau,1}(t) + \cdots + X_p\beta_{\tau,p}(t), \tag{2}$$

where $Q_\tau\{Y(t)|\mathbf{X}\}$ denotes the τth conditional quantile of Y given $\mathbf{X}$ at time t, and $\beta_{\tau,k}(t)$ is the time-varying quantile coefficient of the covariate X_k. A unit increase of X_k will result in an increase of the τth quantile of Y by $\beta_{\tau,k}(t)$ at time t. It is often assumed that $\beta_{\tau,k}(t)$ are unknown smooth functions of t, and can be estimated either parametrically or non-parametrically. Parametric approach offers an efficient way to estimate the coefficients, which are polynomial functions of time. Estimation of parameters in such models can be easily done in existing computational software, and the resulting estimated quantile coefficients enjoy the root-n consistency and asymptotic normality [5].

However, possible bias could be induced due to the underlying assumption of distribution. To overcome the limitation of parametric method, non-parametric approximations are often used to estimate the coefficient functions since it does not require assumptions of the shape of β. [16] considered normalized B-spline to approximate the quantile coefficient functions. That is also the approach used in the paper. Suppose that $(Y_{i,j}, X_{i,j}), i = 1, \ldots\ldots, n,\ j = 1, \ldots m_i$ is a longitudinal sample consisting of n subjects, where m_i as the number of observations of the ith subjects, $Y_{i,j}$ is the jth outcome of the ith subject measured at time/age $t_{i,j}$, and $X_{i,j}$ is the set of covariates associated with $Y_{i,j}$. We approximate the quantile coefficient functions $\beta_{\tau,j}(t)$ by $\pi(t)'\alpha_{\tau,j}$, where $\pi(t)$ is a set of B-spline basis functions with appropriately selected internal knots and order of spline. We then estimate spline coefficients $\boldsymbol{\alpha_\tau} = (\alpha_{\tau,0}, \alpha_{\tau,1}, \ldots, \alpha_{\tau,\mathrm{p}})$ by minimizing the following quantile regression loss function:

$$\widehat{\alpha}_\tau = \arg\min_\alpha \sum_{i=1}^{n} \sum_{j=1} \rho_\tau(y_{i,j} - \pi(t_{i,j})'\alpha_{\tau,0} - X_{1,\mathrm{i,j}}\pi(t_{i,j})'\alpha_{\tau,1} - \cdots - X_p\pi(t_{i,j})'\alpha_{\tau,\mathrm{p}})$$

Here, $\rho_\tau(u) = u(\tau - I\{u < 0\})$ is the quantile regression loss function. The estimated quantile coefficient function can then be written as $\widehat{\beta}_{\tau,j}(t) = \pi(t)'\widehat{\alpha_{\tau,j}}$. [16] showed that the estimated $\widehat{\beta}_{\tau,j}(t)$ has a convergence rate of $O_p\left(n^{\left\{-\frac{2(r-1)}{2r+1}\right\}}\right)$, where r is degree of smoothness of $\beta_{\tau,j}(t)$. If piecewise linear splines are used and $\beta_{\tau,j}(t)$ has bounded second derivative, then the rate of convergence is $n^{\left\{-\frac{2}{5}\right\}}$. Honda [29] considered same model, but used kernel smoothing to estimate model (A1). The coefficient functions $\beta_{\tau,j}(t)$ are estimated by

$$\left\{\widehat{a}(t), \widehat{b}(t)\right\} = \arg\min_{\{a,b\}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \rho_\tau\left(y_{i,j} - \boldsymbol{X}_{i,j}\left(a + b\,\frac{t_{i,j} - t}{h}\right)\right)K\left(\frac{t_{i,j} - t}{h}\right),$$

where $K(\cdot)$ is a kernel function and h is a bandwidth. Honda [29] shows that the estimated quantile coefficient function converges at the rate of $n^{\{-2/5\}}$ when $h = cn^{\{-1/5\}}$.