J. Biosoc. Sci., (2017) **49**, 744–756, © Cambridge University Press, 2017
doi:10.1017/S0021932017000050 First published online 27 Feb 2017

MODELLING THE RISK FACTORS FOR BIRTH WEIGHT IN TWIN GESTATIONS: A QUANTILE REGRESSION APPROACH

A. JOHN MICHAEL, BELAVENDRA ANTONISAMY[1], S. MAHASAMPATH GOWRI AND RAMASAMI PRAKASH

Department of Biostatistics, Christian Medical College, Vellore, India

Summary. Birth weight is used as a proxy for the general health condition of newborns. Low birth weight leads to adverse events and its effects on child growth are both short- and long-term. Low birth weight babies are more common in twin gestations. The aim of this study was to assess the effects of maternal and socio-demographic risk factors at various quantiles of the birth weight distribution for twin gestations using quantile regression, a robust semi-parametric technique. Birth records of multiple pregnancies from between 1991 and 2005 were identified retrospectively from the birth registry of the Christian Medical College and hospitals in Vellore, India. A total of 1304 twin pregnancies were included in the analysis. Demographic and clinical characteristics of the mothers were analysed. The mean gestational age of the twins was 36 weeks with 51% having preterm labour. As expected, the examined risk factors showed different effects at different parts of the birth weight distribution. Gestational age, chroniocity, gravida and child's sex had significant effects in all quantiles. Interestingly, mother's age had no significant effect at any part of the birth weight distribution, but both maternal and paternal education had huge impacts in the lower quantiles (10^{th} and 25^{th}), which were underestimated by the ordinary least squares (OLS) estimates. The study shows that quantile regression is a useful method for risk factor analysis and the exploration of the differential effects of covariates on an outcome, and exposes how OLS estimates underestimate and overestimate the effects of risk factors at different parts of the birth weight distribution.

Introduction

The birth weight of the newborn child serves as a very good predictor of a child's growth and health condition. In addition, it plays a critical role in estimating those newborns at greater risk of death and morbidity, mainly in the neonatal period (Eisner *et al.*, 1979).

[1] Corresponding author. Email: b.antonisamy@gmail.com

The study of the determinants of child birth weight is important as birth weight has been linked to a vast array of health-related complications, both short- and long-term (Abrevaya, 2001). More than half of the world's low birth weight (LBW) babies live in just ten countries, and India alone accounts for one-third (UNICEF, 2013). The weight of newborns depends on many factors, such as the socio-demographic and growth factors of the mother, the most important being gestational age and multiple births.

Much of the increase in low birth weight is attributable to an increase in the proportion of multiple births, because these infants face a much higher risk of low birth weight than do singleton infants (Child Trend Data Bank, 2014) . Twins are eight times more likely to have a low birth weight and ten times more likely to have a very low birth weight compared with singletons (Min *et al.*, 2000). Amongst other factors, shared placental masses and peripheral umbilical cord insertions, which are more common in twin pregnancies, are responsible for birth weight disadvantages in twins (Blickstein *et al.*, 2006; Gielen *et al.*, 2008). In singleton pregnancies fetal growth in the last trimester of pregnancy is linear up to 37 weeks of gestation. A twin fetus grows at the same pace as a singleton up to 32 weeks of gestation, but from then on growth levels slow down. Hence, twins weigh about 600 g less than singletons at birth; even when allowance is made for the shorter gestation a discrepancy remains (Loos *et al.*, 2005). In twin pregnancies the risk of delivery of LBW twin infants has been found to be significantly higher in primiparous than multiparous women (Blickstein *et al.*, 1995; Onyiriuka, 2009). The birth weight of twins also differs according to sex of the twins, with male twins tending to weigh more than female twins; also, same-sex female twins weigh less than same-sex male twins (Onyiriuka, 2011a). Factors related to zygosity and chorionicity also influence twins' intrauterine growth. Dichorionic (DC) twins weigh more than monochorionic (MC) twins (Grennert *et al.*, 1980).

The identification of risk factors for low birth weight is important for the mediation of its health consequences after birth, and also to reduce its prevalence. Many studies have been done at different time points to study the epidemiology of low birth weight risk factors in an effort to reduce its prevalence. Statistical analysis has been carried out using multiple linear regression taking birth weight as the dependent variable, or logistic regression keeping low birth weight indicators as the dependent variables. Some studies have used multiple linear regression to examine the factors affecting birth weight in multiple pregnancies (Ananth *et al.*, 1998; Papageorghiou *et al.*, 2008). This method gives only the mean change in birth weight conditional on other covariates, and is a pure location shift model since it assumes that covariates affect only the location of the conditional distribution of birth weight, not its scale, or any other aspect of its distributional shape. It does not give the complete picture of how the covariates affect birth weight at all parts of the distribution. The major limitation of the standard regression technique is that it assumes the effect of covariates is the same at all quantiles of the outcome.

Bivariate logistic regression analysis has been done on small-for-gestational age (SGA)/large-for-gestational (LGA) age births (Ananth & Preisser, 1999); this is often used for ease of clinical interpretation. However, by dichotomizing the outcome of interest (e.g. LGA), the impact on the full distribution is ignored and the results are not intuitive for individuals falling near to, but at opposite sides of the cut-off point. More seriously, when the outcome is dichotomized, information is lost, which can result in a loss of power, sensitivity to cut-off points and an inability to detect non-linear effects of risk factors (Ellerbe *et al*., 2013).

This study used quantile regression, which provides a more complete picture of the effects of covariates on birth weight by estimating the family of conditional quantile function. Quantile regression allows birth weight to be treated continuously as in linear regression, but allows differential interpretation for the effect at different tails of the birth weight distribution, as in logistic regression, but with the benefit that cut-offs are empirically selected (Ellerbe *et al*., 2013). Its robustness property, and lack of distribution assumptions, make quantile regression a natural choice for this analyis. The application of quantile regression has increased in recent years. Many articles have been published emphasizing its utility in risk factor analyses (Terry *et al*., 2007; Beyerlein *et al*., 2008, 2011; Fenske *et al*., 2008; Belasco *et al*., 2012) and the development of reference growth charts (Wei *et al*., 2006; Li *et al*., 2010; Mayr *et al*., 2012; Daniel-Spiegel *et al*., 2013), compared with other statistical techniques. In this study's twin study data, birth weight was considered to be correlated since newborns share the characteristics of a single mother. Each mother was considered as a cluster and the quantile regression estimates were obtained for each covariate after adjusting for the cluster effect.

Methods

Data

The study was carried out in Vellore, a city in the southern Indian state of Tamil Nadu. Births records maintained by the Christian Medical College (CMC), Vellore, a 2600 bedded academic medical centre, were used. The centre provides obstetric care to the local population of Vellore city and surrounding towns and villages and also acts as a tertiary hospital. The birth records of multiple pregnancies were identified for a 15-year period from 1st January 1991 to 31st December 2005. During the study period, approximately 9000 women delivered each year in this hospital and there were a total of 1673 multiple pregnancies. Of these, the following were sequentially removed from further analysis: 39 triplets; 141 still births of one or more fetuses; 170 with missing information on chorionicity; and 16 with missing data on birth weight and 3 with missing data on gestational age). After exclusions, data were available for 1304 twin pregnancies and these were used for further analysis. Figure 1 shows the flowchart of study participants from the total number of multiple pregnancies delivered during the study period.

Study variables

All socio-demographic, obstetric and clinical variables that determine birth weight in twin births were studied extensively. Socio-demographic variables included age of parents (continuous), education of parents (illiterate, primary, middle, high school, higher secondary and above), religion (Hindu, Muslim, Christian), place of residence (rural or urban), consanguinity (non-consanguineous or consanguineous) and gravida (primi- or multigravida). The placental variable included chorionicity (mono- or dichorionic),

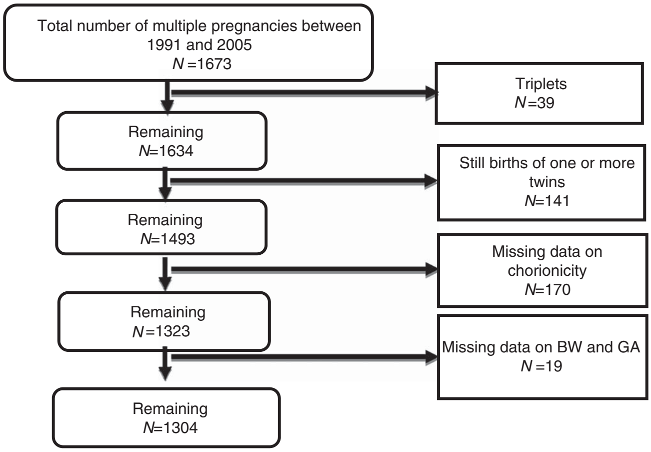

Fig. 1. Flowchart of study participants.

subtype (diamnionic dichorionic, diamnionic monochorionic, monochorionic monochorionic, conjoined), placental abruption (yes or no), placental praevia (yes or no) and abnormal cord insertion (yes or no). Obstetric and clinical variables included pregnancy-induced hypertension (yes or no), gestational diabetes (yes or no), pre-eclampsia (yes or no), birth weight (continuous) and gestational age (continuous).

Quantile regression model

Koenker and Bassett (1978) were the first to introduce the quantile (linear) regression model. Quantile regression allows the identification of heterogeneity in health outcomes from different risk factors and the assessment of the differences in sensitivity to risk factors among outcome levels. In deriving the quantile regression it is important to point out that one can obtain the median of a random variable by minimizing the sum of absolute deviations. As Koenker and Hallock (2001) pointed out, the quantile (τ) can also be obtained by minimizing the sum of asymmetrically weighted absolute residuals, where positive residuals are weighted with τ and negative residuals are weighted with $1-\tau$. This can be written as the optimization problem (Koenker, 2005) and it is given by:

$$\widehat{\beta}_\tau = \arg\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta)$$

The above minimizing equation can be solved effectively using linear programming techniques. For the θth quantile, a common way to write the quantile regession model (Buchinsky, 1998) is:

$$y_i = x_i'\beta_\theta + \mu_{\theta i}, \quad Quant_\theta(\mu_{\theta i}|x_i) = 0 \quad (i = 1, \ldots, n).$$

In the above linear model $x_i'\beta_\theta$ is the conditional quantile of y_i given x_i, which is denoted as $Quant_\theta\,(y_i|x_i)$. The important feature of this quantile regression framework is that the marginal effects of the covariates (x_i), given by β_θ, may differ over quantiles (i.e. different values of θ) (Hao, 2007). The estimates from the standard regression method (ordinary

https://doi.org/10.1017/S0021932017000050 Published online by Cambridge University Press

least squares, OLS) is said to be the same throughout the distribution of y_i. This implies that the effects of covariates on the conditional mean of the dependent variable are the same for the entire distribution.

In this study the dependent variable was birth weight of twins born to a mother. Since both twins share the same maternal characteristics, the problem of heterogeneity arises. To overcome this, the quantile regression model was adjusted for cluster effect and given by:

$$y_{gi} = x'_{gi}\beta_0 + u_{gi},$$

where g is the number of clusters (mothers) and i is the size of the cluster (i = 1, 2 (twins)). Then the quantile regression estimator for clustered data (Silva & Parente, 2013) is defined by:

$$\widehat{\beta} = \arg \min_{\beta \in \mathbb{R}^k} \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n} \rho_\tau (y_{gi} - x'_{gi}\beta)$$

The dependent variable y_{gi} is the birth weight of the ith twin born to the gth mother and covariate x_{gi} is all other variables listed in Table 1. Quantile regressions were estimated at five different quantiles (0.10, 0.25, 0.50, 0.75 and 0.90) of the birth weight distribution and compared with the OLS estimates. The asymptotic standard errors of quantile regression estimates adjusted for cluster were calculated using the Paulo and Santos estimation method (Silva & Parente, 2013) and for OLS estimates the Huber White Sandwich Estimator was used.

Results

Descriptives

The birth weight of twins ranged from 550 g to 3880 g with an overall mean of 2109 g (±537). Of the total 1304 pregnancies, 1127 (86%) gave birth to at least one LBW infant. About 65% of twin pregnancies resulted in both twins being born LBW. The mean maternal and paternal ages were 25.2 (±4.3) and 31.2 (±4.3) years, respectively. About 8% of mothers were illiterate, while 45% of fathers had completed higher secondary grade schooling. There were slightly more urban (55.3%) than rural participants (44.7%) and the predominant religion of participants was Hinduism (82.3%). Twenty-four per cent reported that their marriages were consanguineous and about 45.7% were primigravida mothers. The mean gestational age was 36.1 (±2.9) weeks with 51% being preterm labour. The most common pregnancy-related complications were pregnancy-induced hypertension (26.2%) pre-eclampsia (17.2%) and gestational diabetes (4.9%). Placental praevia (0.6%) and abruption (1.1%) were relatively less common. About one-third (35.4%) of the twin pregnancies had monochorionic presentation. Seventy per cent of twins were of the same sex with 35.8% being males (same-sex pairs).

Univariate analysis

Simple linear regression analysis was performed to identify the most significant factors associated with birth weight. Factors found to be significant ($p < 0.10$) in the

Table 1. Univariate analysis of association of socio-demographic and clinical factors with birth weight of twins (simple linear regression), Vellore, India, 1991–2005

Factor	n[a]	Mean (SD)	β (95% CI)	p-value
Gestational age (weeks)	2608		126.64 (120.7, 132.5)	<0.001
Maternal age (years)				
≤19	212	1908.9 (625.7)	−264.9 (−352.2, −171.6)	0.001
20–24	206	2053.1 (545.3)	−120.8 (−167.6, −74.0)	<0.001
25–29 (Ref.)	446	2173.9 (521.2)		
30–34	710	2206.5 (495.1)	32.5 (−33.9, 99.1)	0.337
35+	1032	2082.6 (480.5)	−91.3 (−200.5, 17.9)	0.101
Maternal education (grades)				
Illiterate	212	1982.9 (524.4)	−171.8 (−272.7, −70.8)	0.001
Primary (1–5)	204	2042.1 (555.2)	−112.5 (−222.7, −70.8)	0.404
Middle (6–8)	446	2096.6 (561.7)	−58.1 (−138.7, 22.5)	0.158
High (9–10)	714	2103.6 (531.1)	−51.1 (−117.6, 15.5)	0.133
Higher secondary+ (12+) (Ref.)	1032	2154.7 (528.4)		
Paternal education (grades)				
Illiterate	176	1953.3 (504.5)	−201.3 (−305.9, −96.6)	<0.001
Primary (1–5)	152	2040.4 (626.5)	−114.2 (−251.8, 23.3)	0.104
Middle (6–8)	306	2110.5 (550.7)	−44.1 (−133.8, 45.6)	0.335
High (9–10)	790	2084.8 (535.7)	−69.8 (−132.8, −6.8)	0.03
Higher secondary+ (12+) (Ref.)	1184	2154.6 (524.9)		
Gravida				
Primigravida	1194	2009.8 (527.3)	−181.1 (−234.8,−127.4)	<0.001
Multigravida (Ref.)	1414	2190.8 (534.5)		
Place of residence				
Rural	1170	2066.1 (535.3)	−77.7 (−132.2, −23.1)	0.005
Urban (Ref.)	1438	2142.8 (539.1)		
Consanguineous marriage				
Yes	636	2069.6 (551.1)	−50.7 (−114.9, 13.5)	0.121
No (Ref.)	1972	2120.5 (534.2)		
Complications (yes vs no [Ref.])				
Gestational diabetes	128	2281.3 (551.9)	182.3 (52.6, 312.1)	0.006
Pregnancy-induced hypertension	680	2136.7 (498.0)	38.9 (−18.8, 96.7)	0.187
Pre-eclampsia	446	2066.2 (505.5)	−50.4 (−118.1, 17.2)	0.144
Placental previa	16	1838.1 (698.9)	−271.5 (−734.5, 191.5)	0.250
Placental abruption	28	1746.4 (402.6)	−365.5 (−554.3,−176.7)	<0.001
Abnormal cord insertion	40	1967.0 (665.1)	-143.2 (−422.2, 136.1)	0.315
Infertility treatment	176	2039.2 (526.5)	−73.8 (−179.4, 31.8)	0.171
Chorionicity				
Monochorionic	914	2027.7 (546.3)	−123.4 (−181.1, −65.8)	<0.001
Dichorionic (Ref.)	1694	2153.2 (527.3)		
Sex of children				
Different sexes	812	2132.9 (547.0)	−15.8 (−66.4, 34.7)	0.539
Same sex, female	864	2040.5 (510.7)	−108.1 (−157.8, −58.4)	<0.001
Same sex, male (Ref.)	932	2148.7 (551.3)		

[a]Number of fetuses.

univariate analysis were included in the multivariable regression analysis. Gestational age, primigravida, lower maternal age, lower educational status of the parents and rural residence were the demographic factors found to be associated with birth weight. Complications such as gestational diabetes and placental abruption had a significant influence on birth weight. Further infant characteristics such as monochorionicity and same-sex females were significantly associated with birth weight. Year of birth seemed to have no impact on the birth weight of the twins.

Comparison of quantile regression estimates and OLS estimates

Table 2 gives a comparison of the coefficients obtained using quantile regression and the OLS method. The quantile regression coefficients revealed that almost all the risk factors for the birth weight of twins had dissimilar effects at different parts of the distribution. The marginal effects of all the covariates differed very much in magnitude compared with the OLS estimates at the 10th percentile of the birth weight distribution, i.e. lower tail of the distribution, which refers to the very low birth weight (VLBW) group. Gestational age plays a major role in a child's birth weight. The influence of gestational age increases monotonically moving from the lower (10th) to the upper quantile (90th). This reveals that a 1-week increment in mother's gestational age will result in an increase of 97 g at the 10th percentile and 157 g at the 90th percentile, which is highly significant. As for maternal age, the changes in birth weight for mothers aged ≤19 years and those aged 35+ years compared with the reference group (20–34 years) were not significant in all the quantiles or in the OLS. Interestingly, in VLBW and LBW quantiles (10th and 25th quantiles), it was observed that the fathers with less than high school education (less than 5th grade) had a negative effect of around 110 g on birth weight compared with others. This marginal effect was underestimated by OLS estimates. The impact of place of residence was high in the lower (10th) quantile as the babies born to mothers residing in rural areas had a decrease of 63 g in birth weight, but the OLS underestimated the decrease at 35 g.

The next most influencial factor on birth weight was gravida. Babies born to primigravida mothers had lighter birth weights with a decrease of 61 g at the 10th quantile (VLBW) and 110 g at the 90th quantile compared with the babies of multigravida mothers. Of the many possible complications for mothers during pregnancy, gestational diabetes and placental abruption had significant roles to play in birth weight. The presence of gestational diabetes was associated with an increased birth weight of babies; the increments were higher in the 10th and 75th quantiles at 102 g and 104 g respectively. Placental abruption had the reverse effect, reducing the birth weight of babies. Though a steady decrease in birth weight due to placental abruption was seen from the lower to upper quantiles, a significant decrease was observed from the 50th quantile with the estimates ranging from −107 g at the 50th to −207 g at the 90th quantiles; this was underestimated by OLS at −141.34 g.

Chorionicity of the twins was found to have differential and significant effects at all parts of the birth weight distribution. The monochorionic twins had lower birth weights compared with dichorionic twins. This change in birth weight was higher in both the lower tail (10th quantile, VLBW) and upper tail (90th quantile, HBW) at −81g and

Table 2. Multiple quantile and linear regression analysis of factors associated with birth weight in twins, Vellore, India, 1991–2005

Explanatory variable	Quantile Coeff. (SE)					OLS Coeff. (SE)
	10%	25%	50%	75%	90%	
Gestational age (weeks)	97.43 (3.87)**	115.15 (2.86)**	128.09 (2.61)**	144.85 (2.67)**	157.45 (3.09)**	123.59 (2.98)**
Maternal age (years)						
≤19	−53.08 (56.52)	−16.36 (56.60)	−1.68 (38.82)	4.48 (40.67)	22.13 (59.74)	−23.60 (42.49)
20–34 (Ref.)						
35+	56.28 (65.85)	−9.091 (53.70)	22.81 (59.61)	−20.00 (49.18)	−27.23 (53.14)	−8.52 (46.12)
Maternal education						
<5 years	−38.59 (35.94)	−64.55 (29.63)*	−102.81 (30.49)*	−53.19 (28.20)	−62.13 (41.17)	−68.42 (27.32)*
Other (Ref.)						
Paternal education						
<5 years	−110.38 (47.04)*	−106.06 (43.75)*	−43.37 (33.59)	−36.08 (28.78)	−52.34 (37.98)	−60.02 (32.06)*
Other (Ref.)						
Gravida						
Primigravida	−61.92 (27.79)*	−103.33 (22.18)**	−113.60 (19.49)**	−103.51 (19.99)**	−109.79 (24.58)**	−104.38 (18.57)**
Multigravida (Ref.)						
Residence						
Rural	−62.44 (27.38)*	−24.24 (21.04)	−39.55 (18.66)*	−27.83 (19.21)	−8.94 (26.16)	−35.51 (18.00)*
Urban (Ref.)						
Complications (yes vs no [Ref.])						
Gestational diabetes	102.44 (48.0)*	82.73 (39.01)*	83.26 (43.83)*	104.64 (51.67)*	71.06 (63.26)	114.62 (39.78)*
Placental abruption	−50.77 (104.4)	−43.03 (101.98)	−107.42 (60.99)*	−134.48 (41.34)*	−206.81 (78.85)*	−141.34 (74.27)
Chroniocity						
Monochorionic	−81.54 (34.19)*	−75.45 (24.46)*	−65.28 (24.63)*	−51.24 (22.69)*	−86.38 (30.56)*	−80.55 (20.96)**
Dichorionic (Ref.)						
Sex						
Different sex	−84.87 (34.50)*	−54.55 (27.52)*	−90.23 (24.98)**	−67.68 (25.64)*	−76.81 (38.01)*	−75.08 (22.95)**
Same sex, female	−129.23 (34.23)**	−126.67 (28.24)**	−140.00 (22.42)**	−133.87 (22.31)**	−132.98 (32.2)**	−141.75 (22.07)**
Same sex, male (Ref.)						

Standard errors adjusted for clusters are given in parentheses; $*p<0.05$; $**p<0.01$.

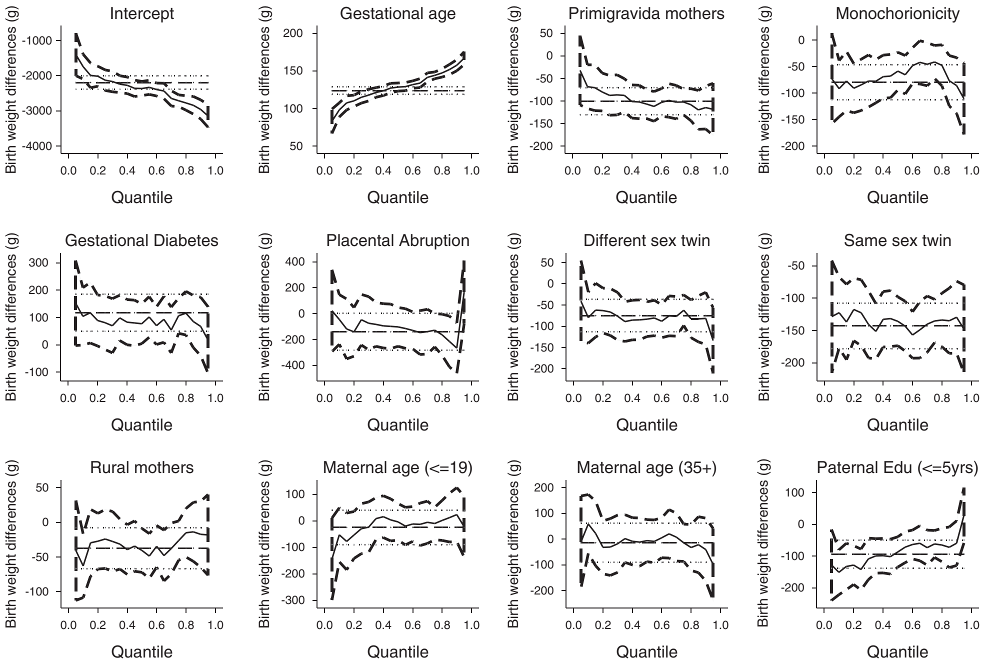

Fig. 2. Visualizing the effect of covariates on birth weight at different quantiles.

−86 g, respectively. The sex of the twin child also played a major role in determining birth weight. Both different-sex twins and same-sex female twins had lower birth weights compared with same-sex male twins. The magnitude of the difference in birth weight was higher in the same-sex female twins compared with the different-sex twins, with coefficients ranging from 130 g to 140 g. Thus the comparison of coefficients obtained using quantile regression with the OLS method revealed that the OLS method failed to describe the effects of covariates on the entire birth weight distribution.

The *F*-test was used to test the equality of coefficients across and between quantiles. With the exception of gestational age, no other covariate's estimates had statistically significant differences between successive quantiles. The test results examining coefficient equality across the quantiles was significant for all covariates, except maternal age and gestational diabetes. Thus this result confirms the differential impact of the covariates across the birth weight distribution for twins.

The quantile regression model estimates clearly expose the differential effects of covariates on the different parts of the birth weight distribution of twins. A comparison of OLS estimates, which assumes equal effects in all parts of the distribution, with quantile estimates for all covariates is shown in Fig. 2. The intercept of the model, represented graphically, can be interpreted as the estimated conditional quantile function of the birth weight distribution of dichorionic male twins born to a multigravida mother aged between 24 and 34 years with more than a high school education, residing in a rural area and with no pregnancy complications such as gestational diabetes and placental abruption. The curve in each graph represents the coefficients of the quantile regression models for the different percentiles of birth weight, while the thickened curve above and below the centre curve shows the 95% confidence band of these estimates; the straight solid line represents the OLS estimates while the dotted straight line indicates the 95% confidence intervals around it.

Discussion

There is a great need to determine the risk factors for birth weight, as low birth weight poses a big threat to the newborn and is a matter of great concern to public health. Globally, multiple births account for a large proportion of low birth weight babies. In 2012, 6.3% of singleton births were of low birth weight, compared with 56.7% of twin or higher plurality births. Similarly, 1.1% of singleton births were very low birth weight, compared with 10.7% of other births (Child Trend Data Bank, 2014). In India, the risk factors for low birth weight in multiple pregnancies have not been well studied. The present study used multiple pregnancy data from the birth records of the Christian Medical College, Vellore, to model the risk factors associated with birth weight in twin gestations.

Previous studies have found that monochorionic twins tend to be lighter in weight by 288 g (Hack *et al*., 2006) and 66.1 g (Ananth *et al*., 1998) after adjusting for gestational age. The current study found that the effect of chorionicity varied in different parts of the birth weight distribution with values ranging from 51.2 g to 86.3 g, and the effect was more prominent in the tails than in the centre. The same-sex male twins weighed more than the same-sex female twins, with average birth weights of 2148.7 g and 2040.5 g, respectively. Similar findings were reported by Onyiriuka (2011b), who found the average birth weights of same-sex male and female twins to be 3903 g and 3426 g, respectively. The next most important factor affecting birth weight is gravida of the mother. Blickstein *et al*. (1995) observed a significantly higher frequency of twins weighing 1500–2500 g and lower frequency weighing >2500 g in the parity 1 group compared with groups of higher parity. The current study found that the birth weight of twins born to primigravida mothers was significantly lower (2009.8 g) than those born to multigravida mothers (2190.8 g) ($p < 0.001$).

Many studies have used multivariable linear regression and logistic or multinomial regression to analyse the risk factors associated with birth weight (Verropoulou & Basten 2014), techniques that categorized birth weight. Such categorization leads to a loss of information, and these regression techniques only explore mean changes in the covariates; this is not adequate to explain the full effect of the risk factors on birth weight. The present study used quantile regression, a semi-parametric robust regression technique that allows birth weight to be treated continuously, as in linear regression. This does not require any transformation of the outcome to a binary variable as in logistic regression, assesses shifts in specific parts of the continuous outcome (birth weight) distribution instead of probabilities of falling into one outcome category or another (Beyerlein *et al*., 2011) and allows differential interpretation for the effect of covariates at different tails of the birth weight distribution, as in logistic regression. Quantile regression provides more flexibility than other regression methods for the identification of differing relationships at different parts of the distribution of the dependent variable.

Verropoulou and Tsimbos (2013) and Abrevaya (2001) used quantile regression effectively to determine the effects of maternal and demographic factors on the birth weight distribution. Both studies stated that quantile regression allows the study of the differential effects of covariates at different parts of birth weight distribution. They compared coefficients from quantile regression models and the OLS method. Finally, these studies concluded that OLS overestimates the effects of covariates on birth weight at the lower tail and underestimates them in the upper tail. The present study supports these studies' findings, as the effects of gestational age and primigravida mothers compared with multigravida mothers were underestimated by OLS at the upper tails and overestimated in the lower tails of the birth weight distribution.

The main advantage of quantile regression methodology is that it allows for understanding of the relationships between variables outside of the mean of the data, making it useful to understand outcomes that are non-normally distributed and that have non-linear relationships with predictor variables (Cade & Noon, 2003). Quantile regression allows the analyst to drop the assumption that variables operate in the same way at the upper tails of the distribution as at the mean, and the identification of factors that are important determinants of the outcome (Le Cook & Manning, 2013). Quantile regression is very robust against outliers, in contrast to mean and least squares regression, and a normal distribution of errors is not assumed (Eilers *et al*., 2012). The method allows the heterogeneity effect in the outcome to be studied (Belasco *et al*., 2012) and additional interpretation of the risk factors affecting only parts of the distribution (Beyerlein *et al*., 2008), making the method more useful in many realistic situations.

By using quantile regression, the cut-point is empirical made rather than clinical. As a result, care must be taken when interpreting the results across studies (Ellerbe *et al*., 2013). In addition, in contrast to linear or logistic regression, interpretation is further complicated by having to interpret the results for each quantile of the distribution rather than report a single summary measure of effect, one limitation of the technique. However, the quantile regression method provides more insight and the strengths of this analytical method outweigh its limitations. Thus quantile regression is a useful technique for the analysis of risk factors for birth weight.

Acknowledgments

The authors sincerely thank Dr Jiji Mathews, Dr Annie Regi and Dr Ruby Jose from the Department of Obstetrics and Gynaecology, CMC, Vellore, and Dr Anil Kuruvilla from the Department of Neonatology, CMC, Vellore, for their support and guidance.

References

Abrevaya, J. (2001) The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics* **26**, 247–257.

Ananth, C. V. & Preisser, J. S. (1999) Bivariate logistic regression: modelling the association of small for gestational age births in twin gestations. *Statistics in Medicine* **18**, 2011–2023.

Ananth, C. V., Vintzileos, A. M., Shen-Schwarz, S., Smulian, J. C. & Lai, Y. L. (1998) Standards of birth weight in twin gestations stratified by placental chorionicity. *Obstetrics & Gynecology* **91**, 917–924.

Belasco, E., Chidmi, B., Lyford, C. & Funtanilla, M. (2012) Using quantile regression to measure the differential impact of economic and demographic variables on obesity. *Journal of Health Behavior and Public Health* **2**, 35–45.

Beyerlein, A., Fahrmeir, L., Mansmann, U. & Toschke, A. M. (2008) Alternative regression models to assess increase in childhood BMI. *BMC Medical Research Methodology* **8**, 59.

Beyerlein, A., Toschke, A. M., Schaffrath Rosario, A. & von Kries, R. (2011) Risk factors for obesity: further evidence for stronger effects on overweight children and adolescents compared to normal-weight subjects. *PLoS One* **6**, e15739.

Blickstein, I., Mincha, S., Goldman, R. D., Machin, G. A. & Keith, L. G. (2006) The Northwestern twin chorionicity study: testing the 'placental crowding' hypothesis. *Journal of Perinatal Medicine* **34**, 158–161.

Blickstein, I., Zalel, Y. & Weissman, A. (1995) Pregnancy order: a factor influencing birth weight in twin gestations. *Journal of Reproductive Medicine* **40**, 443–446.

Buchinsky, M. (1998) Recent advances in quantile regression models: a practical guideline for empirical research. *Journal of Human Resources* **33**, 88.

Cade, B. S. & Noon, B. R. (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* **1**, 412–420.

Child Trend Data Bank (2014) Low and very low birthweight infants. *Child Trends*. URL: http://www.childtrends.org/?indicators=low-and-very-low-birthweight-infants. (accessed 9th September 2016).

Daniel-Spiegal, E., Weiner, E., Yarom, I., Doveh, E., Friedman, P., Cohen, A. & Shalev, E. (2013) Establishment of fetal biometric charts using quantile regression analysis. *Journal of Ultrasound Medicine* **2**, 23–33.

Eilers, P. H. C., Röder, E., Savelkoul, H. F. J. & van Wijk, R. G. (2012) Quantile regression for the statistical analysis of immunological data with many non-detects. *BMC Immunology* **13**, 37.

Eisner, V., Brazie, J. V., Pratt, M. W. & Hexter, A. C. (1979) The risk of low birth weight. *American Journal of Public Health* **69**, 887–893.

Ellerbe, C. N., Gebregziabher, M., Korte, J. E., Mauldin, J. & Hunt, K. J. (2013) Quantifying the impact of gestational diabetes mellitus, maternal weight and race on birth weight via quantile regression. *PLoS One* **8**, e65017.

Fenske, N., Fahrmeir, L., Rzehak, P. & Höhle, M. (2008) *Detection of Risk Factors for Obesity in Early Childhood with Quantile Regression Methods for Longitudinal Data.* URL: http://epub.ub.uni-muenchen.de/6260/ (accessed 19th September 2014).

Gielen, M., Lindsey, P. J., Derom, C., Smeets, H. J. M., Souren, N. Y., Paulussen, A. D. C. *et al.* (2008) Modeling genetic and environmental factors to increase heritability and ease the identification of candidate genes for birth weight: a twin study. *Behavioral Genetics* **38**, 44–54.

Grennert, L., Persson, P. H., Gennser, G. & Gullberg, B. (1980) Zygosity and intrauterine growth of twins. *Obstetrics & Gynecology* **55**, 684–687.

Hack, K. E. A., Derks, J. B., de Visser, V. L., Elias, S. G. & Visser, G. H. A. (2006) The natural course of monochorionic and dichorionic twin pregnancies: a historical cohort. *Twin Research and Human Genetics* **9**, 450–455.

Hao, L. (2007) *Quantile Regression*. SAGE Publications Inc., Thousand Oaks, CA.

Koenker, R. (2005) *Quantile Regression*. Cambridge University Press, Cambridge, New York.

Koenker, R. & Bassett, G. (1978) Regression quantiles. *Econometrica* **46**, 33.

Koenker, R. & Hallock, K. F. (2001) Quantile regression. *Journal of Economic Perspectives* **15**, 143–156.

Le Cook, B. & Manning, W. G. (2013) Thinking beyond the mean: a practical guide for using quantile regression methods for health services research. *Shanghai Archive of Psychiatry* **25**, 55–59.

Li, Y., Graubard, B. I. & Korn, E. L. (2010) Application of nonparametric quantile regression to body mass index percentile curves from survey data. *Statistics in Medicine* **29**, 558–572.

Loos, R. J. F., Derom, C., Derom, R. & Vlietinck, R. (2005) Determinants of birth weight and intrauterine growth in liveborn twins. *Paediatric and Perinatal Epidemiology* **19** (Supplement 1), 15–22.

Mayr, A., Hothorn, T. & Fenske, N. (2012) Prediction intervals for future BMI values of individual children: a non-parametric approach by quantile boosting. *BMC Medical Research and Methodology* **12**, 6.

Min, S. J., Luke, B., Gillespie, B., Min, L., Newman, R. B., Mauldin, J. G. *et al.* (2000) Birth weight references for twins. *American Journal of Obstetrics and Gynecology* **182**, 1250–1257.

Onyiriuka, A. N. (2009) Birthweight data of live-born twins in Benin City, Nigeria. *Sahel Medical Journal* **11**(4) doi: 10.4314/smj2.v11i4.12988.

Onyiriuka, A. N. (2011a) Low birthweight infants in twin gestation. *Current Pediatric Research* **15**(1), 37–41.

Onyiriuka, A. N. (2011b) Birthweight of full-term twin infants in relation to sex-pair. *Genomic Medicine Biomarkers and Health Sciences* **3**(3–4), 123–127.

Papageorghiou, A. T., Bakoulas, V., Sebire, N. J. & Nicolaides, K. H. (2008) Intrauterine growth in multiple pregnancies in relation to fetal number, chorionicity and gestational age. *Ultrasound Obstetrics and Gynecology* **32**, 890–893.

Silva, J. M. C. S. & Parente, P. M. D. C. (2013) Quantile regression with clustered data. *Economics Discussion Paper No. 728.* Department of Economics, University of Essex.

Terry, M. B., Wei, Y. & Esserman, D. (2007) Maternal, birth, and early-life influences on adult body size in women. *American Journal of Epidemiology* **166**, 5–13.

UNICEF (2013) Undernourishment in the womb can lead to diminished potential and predispose infants to early death. *UNICEF Data: Monitoring the Situation of Children and Women*. URL: http://data.unicef.org/nutrition/low-birthweight#sthash.BG4IvrwC.dpuf.

Verropoulou, G. & Basten, S. (2014) Very low, low and heavy weight births in Hong Kong SAR: how important is socioeconomic and migrant status? *Journal of Biosocial Science* **46**, 316–331.

Verropoulou, G. & Tsimbos, C. (2013) Modelling the effects of maternal socio-demographic characteristics on the preterm and term birth weight distributions in Greece using quantile regression. *Journal of Biosocial Science* **45**, 375–390.

Wei, Y., Pere, A., Koenker, R. & He, X. (2006) Quantile regression methods for reference growth charts. *Statistics in Medicine* **25**, 1369–1382.