Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2016, Vol. 25(6) 2781–2810

DOI: 10.1177/0962280214532745
smm.sagepub.com
SAGE

Predicting birth weight with conditionally linear transformation models

Lisa Möst,[1] Matthias Schmid,[2] Florian Faschingbauer[3] and Torsten Hothorn[4]

Abstract

Low and high birth weight (BW) are important risk factors for neonatal morbidity and mortality. Gynecologists must therefore accurately predict BW before delivery. Most prediction formulas for BW are based on prenatal ultrasound measurements carried out within one week prior to birth. Although successfully used in clinical practice, these formulas focus on point predictions of BW but do not systematically quantify uncertainty of the predictions, i.e. they result in estimates of the conditional mean of BW but do not deliver prediction intervals. To overcome this problem, we introduce conditionally linear transformation models (CLTMs) to predict BW. Instead of focusing only on the conditional mean, CLTMs model the whole conditional distribution function of BW given prenatal ultrasound parameters. Consequently, the CLTM approach delivers both point predictions of BW and fetus-specific prediction intervals. Prediction intervals constitute an easy-to-interpret measure of prediction accuracy and allow identification of fetuses subject to high prediction uncertainty. Using a data set of 8712 deliveries at the Perinatal Centre at the University Clinic Erlangen (Germany), we analyzed variants of CLTMs and compared them to standard linear regression estimation techniques used in the past and to quantile regression approaches. The best-performing CLTM variant was competitive with quantile regression and linear regression approaches in terms of conditional coverage and average length of the prediction intervals. We propose that CLTMs be used because they are able to account for possible heteroscedasticity, kurtosis, and skewness of the distribution of BWs.

Keywords

conditional transformation models, component-wise boosting, prediction intervals, conditional coverage

[1]Institut für Statistik, Ludwig-Maximilians-Universität, München, Germany
[2]Institut für Medizinische Biometrie, Informatik und Epidemiologie, Bonn, Germany
[3]Frauenklinik, Geburtshilfe und Pränataldiagnostik, Universitätsklinikum Erlangen, Universitätsstraße, Erlangen, Germany
[4]Institut für Epidemiologie, Biostatistik und Prävention, Abteilung Biostatistik, Universität Zürich, Zürich, Switzerland

Corresponding author:
Torsten Hothorn, Institut für Epidemiologie, Biostatistik und Prävention, Abteilung Biostatistik, Universität Zürich, Hirschengraben 84, CH-8001 Zürich, Switzerland.
Email: Torsten.Hothorn@uzh.ch

1 Introduction

Birth weight (BW) is among the most important risk indicators for neonatal morbidity and mortality.[1,2] As shown in numerous studies, high BW is associated with serious maternal trauma after vaginal and surgical delivery and shoulder dystocia with fetal brachial plexus paralysis and/or clavicular fracture[3,4] and low BW increases the risk of neurological and developmental deficits during childhood.[5,6] The accurate estimation of BW is challenging for gynecologists who need to plan the mode of delivery and organize obstetric management.

Fetal ultrasound examinations have become routine during the last 40 years[7] and result in readily available two-dimensional measurements highly correlated with BW. Most prediction formulas for BW incorporate biometric parameters, such as biparietal diameter (BPD), fronto-occipital diameter (FOD), head circumference (HC), abdominal transverse diameter (ATD), anterior-posterior abdominal diameter (APD), abdominal circumference (AC), and femur length (FL). Here we focus on the *statistical* aspects of prediction formulas for BW. Our analysis is based on prenatal ultrasound measurements recorded within seven days before delivery of $N = 8712$ babies at the Perinatal Centre of the University Clinic Erlangen, Germany, in 2003–2011.

Statistically, the development of a prediction formula for BW is a regression modeling task that involves the accurate estimation of ultrasound predictor effects on BW:

(1) Many traditional prediction formulas for BW have been derived by applying linear regression models with Gaussian errors.[7–10] Only little attention has been given to the frequent departure of the distribution of BW from the normal distribution, which could make relying on a Gaussian model suboptimal. For example, if a high percentage of the newborns are very small, the distribution of BW would not be normal but rather right skewed. A suitable approach to model BW should take this skewness into account.
(2) A thorough investigation of the accuracy of the prediction formulas is essential for clinical practice because, as stated by e.g. Scioscia et al.,[7] many prediction formulas show the same tendency to under- and over-estimate BW at the extremes, regardless of the ultrasound parameters relied upon. To assess the performance of new prediction formulas, measures such as the relative percentage error (defined as $(\mathrm{BW} - \widehat{\mathrm{EW}})/\mathrm{BW}$) and the absolute percentage error (defined as $|\mathrm{BW} - \widehat{\mathrm{EW}}|/\mathrm{BW}$) have been commonly used, where $\widehat{\mathrm{EW}}$ denotes estimated fetal weight.[7,11,12] As the traditional formulas for predicting BW estimate only the conditional mean, the aforementioned performance measures focus on the quality of the point estimates for the actual BW, and an appropriate measure of prediction *uncertainty* is missing. An easy-to-interpret measure of prediction accuracy accompanied with some measure of uncertainty is interval estimates that cover the true weights of newborns with a high probability. Although it is possible to construct prediction intervals around the point estimates obtained from the Gaussian modeling approach mentioned earlier, these intervals are subject to potential bias. First, intervals obtained from Gaussian models are always symmetric around the conditional mean. Consequently, these intervals might be suboptimal because the distribution of BW (and possibly also the distribution of the residuals in linear regression) is skewed. Second, Gaussian prediction intervals all have the same length owing to a constant residual variance term, regardless of the ultrasound measurements. This assumption is often inappropriate as the prediction accuracy may depend on the actual BW (via the ultrasound measurements), e.g. larger fetuses might have wider prediction intervals than smaller fetuses.

To address these issues, we propose conditionally linear transformation models (CLTMs) as a novel approach to predict BW. Instead of considering the conditional mean only (as traditional Gaussian

regression does), CLTMs model the whole conditional distribution function of BW given prenatal ultrasound parameters. Consequently, each quantile of the BW distribution can be predicted by a single CLTM. This implies that the CLTM approach not only results in point predictions of BW (i.e. in predictions of the median), but additionally results in fetus-specific prediction intervals (whose boundaries are given, e.g. by the predicted 10% and 90% quantiles). The interval estimates obtained from CLTMs represent an easy-to-interpret measure of prediction accuracy and allow identification of fetuses subject to high prediction uncertainty. Moreover, interval lengths obtained from the CLTM approach depend on individual ultrasonic measurements of each fetus. This strategy results in ''personalized'' prediction intervals for each fetus and clearly provides more information than classical point predictions alone.

The CLTM approach is a special case of the recently proposed conditional transformation modeling (CTM) approach.[13] Compared to the CTM approach, the CLTM methodology proposed herein has the advantage that the underlying modifications lead to model results that are easier to interpret, and a closer insight into model structure can be gained.

In Section 2, we review common prediction formulas for BW and associated traditional methods of estimation. We also introduce the Perinatal Database Erlangen and discuss prediction intervals for BW. A thorough introduction to CLTMs, including some comments on interpretability and estimation, is given in Section 3. We present the results of the analysis of the Perinatal Database Erlangen in Section 4 and discuss the results in Section 5.

2 Prediction of birth weight

2.1 Review of common prediction formulas for BW

Since the 1970s, gynecologists have developed numerous formulas to predict BW based on prenatal ultrasound measurements. Summaries of these formulas are given in the literature.[7,8,14] A well-established prediction formula commonly used in clinical practice is that proposed by Hadlock et al.[15]

$$\log_{10}(\widehat{\mathrm{EW}}) = 1.304 + 0.05281 \times \mathrm{AC} + 0.1938 \times \mathrm{FL} - 0.004 \times \mathrm{AC} \times \mathrm{FL},$$

where biometric parameters are measured in centimeters and estimated fetal weight ($\widehat{\mathrm{EW}}$) is measured in grams. In addition to classical prediction formulas based on 2-D ultrasound measurements, other formulas incorporate clinical parameters[16] or 3-D ultrasound measurements,[17] or focus on high-risk deliveries.[11,12,18] Choi et al.[19] suggest a model with spatio-temporally varying coefficients for low BWs. Because 3-D ultrasound measurements do not seem to improve many predictions and are poorly suited for every-day clinical practice,[7] we focused on routinely measured 2-D biometric parameters in our study. The traditional prediction formulas for BW that we are aware of were derived using linear regression approaches with Gaussian errors.

2.2 Perinatal Database Erlangen

Our analysis is based on data of $N = 8712$ singleton pregnancies with a complete ultrasound examination within seven days before delivery. Biometric parameters included BPD, FOD, HC, ATD, APD, AC, and FL. Additionally, the mother's body mass index (BMI) was measured. In cases in which fetus growth was followed serially, we used measurements only from the last examination before delivery. All ultrasound measurements were made by experienced examiners who underwent

extensive training at University Clinic Erlangen. BW was measured by the nursing staff at Erlangen University Hospital within 1 h after delivery. Children with chromosomal or structural malformations and intrauterine deaths were excluded from analysis.

2.3 Prediction intervals

Since we are interested in some measure that quantifies the uncertainty of predictions for BWs, we considered fetus-specific *prediction intervals*.[20] These intervals result in a range of predicted values that cover the BW with high probability $1-\alpha$, where α is a pre-specified error level.

A common way to define the boundaries of a prediction interval is to use the $\alpha/2$ quantile and the $(1-\alpha/2)$ quantile of the conditional distribution of BW given ultrasound measurements

$$\widehat{\mathrm{PI}}_{1-\alpha}(\mathbf{x}) = \left[\hat{q}_{\alpha/2}(\mathbf{x}), \hat{q}_{1-\alpha/2}(\mathbf{x})\right] \tag{1}$$

Here, $\mathbf{x}$ denotes the ultrasound measurements of a new fetus, and $\hat{q}_{\alpha/2}$ and $\hat{q}_{1-\alpha/2}$ the $\alpha/2$ and the $(1-\alpha/2)$ quantile, respectively, of the corresponding conditional distribution of BW. Since the estimated prediction intervals depend on the ultrasound measurements, the interval lengths and interval borders are fetus-specific. In other words, depending on the ultrasound measurements, accurate or inaccurate predictions can be made, which results in narrow or wide prediction intervals, respectively.[20] Nevertheless, the underlying assumptions of the regression model used (e.g. normally distributed responses and homoscedasticity for linear regression models) in equation (1) influence the form of the resulting prediction intervals. For example, the resulting prediction intervals may differ in symmetry assumptions and methods for boundary estimation. Common methods for the calculation of prediction intervals are, e.g., linear regression or quantile regression approaches.

2.4 Existing approaches for calculation of prediction intervals

If linear regression models are used for BW prediction, the conditional mean of BW is modeled as a linear function of the (possibly transformed) prenatal ultrasound measurements. After estimation of the model parameters, symmetric prediction intervals are constructed around the point predictions based on the assumptions of homoscedasticity and normality.[21] Hence, the resulting symmetric prediction intervals are inadequate if the BW's distribution is skewed and if the residual variance depends on ultrasound measurements.

The use of linear or additive quantile regression approaches to determine prediction intervals for BW conveniently solves these problems. With quantile regression,[22,23] one directly estimates the boundaries of the prediction intervals by using separate regression models for the quantiles $q_{\alpha/2}$ and $q_{1-\alpha/2}$ (equation (1)).[24] The influence of the ultrasound parameters on the respective quantiles is assumed to be additive. Although this approach avoids any distributional assumptions, a non-trivial problem associated with quantile regression is quantile crossing.[25] The logical monotonicity requirements of the probability p ($p = q^{-1}$) are not fulfilled, and neighboring quantile curves may cross because they are estimated independently.

To avoid quantile crossing (and also the aforementioned problems associated with linear regression), we propose CLTMs to estimate intervals for the prediction of BW. In contrast to quantile regression approaches, CLTMs model all conditional quantiles simultaneously by estimating the whole conditional distribution function, and the relevant quantiles are extracted afterwards. Thereby, inconsistencies between neighboring quantiles are avoided.

3 Conditionally linear transformation models

3.1 Model class

CLTMs are a special case of CTMs that model the conditional distribution function of a response $Y_{\mathbf{x}} = (Y|\mathbf{X} = \mathbf{x})$ depending on explanatory variables $\mathbf{x}$. Most common regression models model only the conditional mean $\mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ of the response $Y \in \mathbb{R}$ as a function of the explanatory variables $\mathbf{X} = \mathbf{x}$. This is due to the underlying assumption of additivity of signal and noise, which is relaxed by considering CTMs.[13] Therefore, not only the conditional mean but also higher moments of the conditional distribution function may depend on explanatory variables.

We used the CTM approach to model the conditional distribution function of BW depending on prenatal ultrasound measurements

$$\mathbb{P}(\mathrm{BW} \leq \upsilon|\mathbf{X} = \mathbf{x}) = F_{\mathrm{BW}|\mathbf{X}=\mathbf{x}}(\upsilon) = F(h(\upsilon|\mathbf{x})) \tag{2}$$

The conditional distribution function is modeled in terms of the monotone transformation function $h : \mathbb{R} \to \mathbb{R}$, which depends on ultrasound measurements $\mathbf{x}$. Moreover, $\upsilon \in \mathbb{R}$ denotes some arbitrary BW and F denotes an absolute continuous distribution function $F : \mathbb{R} \to [0, 1]$ with corresponding quantile function $Q = F^{-1}$. The transformation function h transforms the BWs conditionally on $\mathbf{x}$, so that the distribution of the transformed BWs follows the distribution function F. When fitting CTMs, we assume that such a montone transformation function h exists. CTMs can be understood as the inverse of a quantile regression model, since we do not model the conditional quantile function, but we model the conditional distribution function of the BWs directly. Thereby, we are able to estimate all quantiles simultaneously in a joint model and do not need to fit separate models for all quantiles like in quantile regression. When CTMs are estimated, the monotone transformation function h is estimated, but the continuous distribution function F is chosen arbitrarily. A common choice is the standard normal distribution function $F = \Phi$ with corresponding quantile function $Q = \Phi^{-1}$. Hence, model characteristics must be defined in terms of characteristics of the transformation function h.

As we modeled the whole conditional distribution function, higher moments (e.g. the variance) may also depend on ultrasound measurements. In addition, further moments of the prediction distribution of the BWs can be modeled flexibly, e.g. kurtosis and skewness. When functionals of the conditional distribution function, such as prediction intervals, are calculated, it is important to note that variance and skewness may depend on explanatory variables; otherwise, heteroscedasticity and varying higher moments are ignored.

Nevertheless, the CTMs presented in Hothorn et al.[13] define a very complex and general class of transformation models, and therefore model interpretations can be challenging. Moreover, a lack of orthogonality of the model components constricts insights into model structure. As a consequence, direct interpretations of the relationship between the explanatory variables and certain moments of the distribution function of the response are difficult to obtain because these effects usually cannot be separated. Since we are interested in a more easily interpretable version of CTMs in this application, we reduced the model complexity by imposing restrictions on CTMs and introducing CLTMs. The model class of CLTMs can be described by the following linear transformation conditional on $\mathbf{x}$

$$\begin{aligned} &h(Y_{\mathbf{x}}|\mathbf{x}) = Z \ \sim F, \text{ with} \\ &h(Y_{\mathbf{x}}|\mathbf{x}) = h_0(Y_{\mathbf{x}}) \cdot \beta(\mathbf{x}) + \alpha(\mathbf{x}) \end{aligned} \tag{3}$$

Here, h denotes a monotone transformation function that depends on explanatory variables. The random variable Z is a transformation of the responses $Y_{\mathbf{x}}$ depending on explanatory variables $\mathbf{x}$ and follows the known distribution function F. In CLTMs, we modeled only linear functions of the transformed responses to reduce model complexity (equation (3)). Hence, we considered a flexible and possibly unknown response transformation $h_0(Y_{\mathbf{x}})$ that depends only on the response values $Y_{\mathbf{x}}$. The response transformation itself was transformed by the explanatory variables via a linear function, where the coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ depend on the explanatory variables. The coefficients $\alpha(\mathbf{x})$ induce shifts of the response transformation $h_0(Y_{\mathbf{x}})$, and the coefficients $\beta(\mathbf{x})$ induce shifts and scalings of the response transformation $h_0(Y_{\mathbf{x}})$ depending on the respective explanatory variables.

Owing to the restriction of the transformation function h to linear functions of the response transformation $h_0(Y_{\mathbf{x}})$, the explanatory variables $\mathbf{x}$ can only influence the conditional mean and conditional variance of the response transformation. This follows directly from calculating the conditional mean and conditional variance in equation (3) and solving the equation for both $\mathbb{E}(h_0(Y_{\mathbf{x}})|\mathbf{x})$ and $\mathbb{V}(h_0(Y_{\mathbf{x}})|\mathbf{x})$:

$$\begin{aligned}\mathbb{E}(h_0(Y_{\mathbf{x}})|\mathbf{x}) &= \frac{\mathbb{E}(Z) - \alpha(\mathbf{x})}{\beta(\mathbf{x})} \\ \mathbb{V}(h_0(Y_{\mathbf{x}})|\mathbf{x}) &= \frac{\mathbb{V}(Z)}{\beta(\mathbf{x})^2}\end{aligned} \tag{4}$$

If we assume that the transformed responses Z follow a standard normal distribution $Z \sim \mathcal{N}(0,1)$, we get $\mathbb{E}(Z) = 0$ and $\mathbb{V}(Z) = 1$, and equation (4) simplifies accordingly. The coefficients $\alpha(\mathbf{x})$ influence only the conditional mean of the response transformation, whereas the coefficients $\beta(\mathbf{x})$ influence its conditional mean and its conditional variance. Hence, the influence of the explanatory variables on the conditional mean and conditional variance of the response transformation can be formulated in CLTMs, whereas such a formulation cannot be given in CTMs. This difference can also be seen by looking at the conditional quantile functions implied by CTMs and CLTMs

$$\begin{aligned}Q_{\text{CTM}}(p|\mathbf{x}) &= h^{-1}(F^{-1}(p)|\mathbf{x}) \\ Q_{\text{CLTM}}(p|\mathbf{x}) &= h_0^{-1}\left(\frac{F^{-1}(p) - \alpha(\mathbf{x})}{\beta(\mathbf{x})}\right)\end{aligned}$$

For CTMs, the effect of the explanatory variables on the conditional quantile may vary with p, whereas in CLTMs, the conditional quantile is a nonlinear transformation of a linear function of $F^{-1}(p)$, where the coefficients of the latter do not depend on p. Because only the mean and the variance may depend on explanatory variables in CLTMs, we can only model constant kurtosis and skewness in contrast to quantile regression. A possible influence of the explanatory variables on higher moments can only be estimated in the more complex model class of CTMs.

Furthermore, we assumed additivity on the scale of the transformation function; therefore, we decomposed the monotone transformation function h into $J+1$ partial transformation functions, given the explanatory variables[13] (equation (3)):

$$\begin{aligned}Z = h(Y_{\mathbf{x}}|\mathbf{x}) &= \sum_{j=0}^{J} h_j(Y_{\mathbf{x}}|\mathbf{x}) = \sum_{j=0}^{J}\big(h_0(Y_{\mathbf{x}}) \cdot \beta_j(\mathbf{x}) + \alpha_j(\mathbf{x})\big) \\ &= h_0(Y_{\mathbf{x}}) \cdot \sum_{j=0}^{J} \beta_j(\mathbf{x}) + \sum_{j=0}^{J} \alpha_j(\mathbf{x})\end{aligned} \tag{5}$$

Despite this decomposition, the random variable Z still remains a linear function of the response transformation $h_0(Y_\mathbf{x})$.

Prominent members of the family of linear transformation models, most importantly the proportional hazards and the proportional odds model, can be connected by restricting the aforementioned CLTMs to the case where only shifts of the response transformation that depend on explanatory variables are allowed

$$h(Y_\mathbf{x}|\mathbf{x}) = h_0(Y_\mathbf{x}) + \sum_{j=0}^{J} \alpha_j(\mathbf{x}) = h_0(Y_\mathbf{x}) + \alpha(\mathbf{x}) \tag{6}$$

In this model, the explanatory variables can only influence the mean $-\alpha(\mathbf{x})$ of the transformed response $h_0(Y_\mathbf{x})$. The transformation functions of the proportional hazards model and the proportional odds model result if we choose a CLTM (equation (5)) with $\beta(\mathbf{x}) \equiv 1$ and an appropriate response transformation $h_0(Y_\mathbf{x})$, which is treated as a nuisance parameter in classical formulations of the proportional hazards model and proportional odds model. For linear shift functions $\alpha(\mathbf{x})$, a unified estimation framework has been proposed by Cheng et al.[26]

We assumed that the response transformation $h_0(Y_\mathbf{x})$ is unknown. In the first step, we decomposed the response transformation into one part consisting only of linear functions and a more complex part representing deviations from linearity:

$$h_0(Y_\mathbf{x}) = \underbrace{\alpha_0 + \beta_0 \cdot Y_\mathbf{x}}_{\text{linear part}} + \underbrace{\tilde{h}_0(Y_\mathbf{x})}_{\text{deviations from linearity}} \tag{7}$$

The decomposition in equation (7) is reasonable since the model component $\tilde{h}_0(Y_\mathbf{x})$ can be used to decide whether the response variable follows a normal distribution or not, if we additionally set the link function to $F = \Phi$. If the model component $\tilde{h}_0(Y_\mathbf{x})$ is missing, we only observe a linear transformation of the conditional response, and hence we cannot leave the class of normal distributions because the normal distribution is invariant towards linear transformations. Consequently, by estimating the more complex deviations from linearity $\tilde{h}_0(Y_\mathbf{x})$, we are able to leave the class of normal distributions and model other classes of distribution functions as well.

Combining equation (7) with the definition of CLTMs in equation (3) leads to

$$h(Y_\mathbf{x}|\mathbf{x}) = (Y_\mathbf{x} + \tilde{h}_0(Y_\mathbf{x})) \cdot \beta(\mathbf{x}) + \alpha(\mathbf{x}) = Y_\mathbf{x} \cdot \beta_{\text{lin}}(\mathbf{x}) + \tilde{h}_0(Y_\mathbf{x}) \cdot \beta_c(\mathbf{x}) + \alpha(\mathbf{x})$$

where $\beta_{\text{lin}}(\mathbf{x})$ denotes the part of $\beta(\mathbf{x})$ influencing the linear part of the response transformation $h_0(Y_\mathbf{x})$, and $\beta_c(\mathbf{x})$ denotes the part of $\beta(\mathbf{x})$ influencing the more complex deviations from linearity $\tilde{h}_0(Y_\mathbf{x})$.

We furthermore assumed that the more complex deviations $\tilde{h}_0(Y_\mathbf{x})$ do not depend on any explanatory variables; therefore, we set $\beta_c(\mathbf{x}) \equiv 1$. This is a strong assumption, but since we are interested in an interpretable model class, this is a necessary restriction of model complexity. The transformation function h with an unknown and decomposed response transformation at the start results in

$$h(Y_\mathbf{x}|\mathbf{x}) = Y_\mathbf{x} \cdot \beta_{\text{lin}}(\mathbf{x}) + \tilde{h}_0(Y_\mathbf{x}) + \alpha(\mathbf{x})$$

Then we included the decomposition of the monotone transformation function h into $J+1$ partial transformation functions (equation (5)):

$$h(Y_{\mathbf{x}}|\mathbf{x}) = \tilde{h}_0(Y_{\mathbf{x}}) + Y_{\mathbf{x}} \cdot \sum_{j=0}^{J} \beta_{j,\text{lin}}(\mathbf{x}) + \sum_{j=0}^{J} \alpha_j(\mathbf{x}) \tag{8}$$

We furthermore set $\alpha_0(\mathbf{x}) \equiv \alpha_0$ and $\beta_{0,\text{lin}}(\mathbf{x}) \equiv \beta_0$, which we already implicitly did in equation (7). By introducing the scalars α_0 and β_0, the transformation function h can be decomposed into an unconditional part (not depending on any explanatory variables) and a conditional part (depending on explanatory variables), which facilitates model interpretations. The resulting structure of the monotone transformation function is still consistent with the model class of CLTMs:

$$h(Y_{\mathbf{x}}|\mathbf{x}) = \underbrace{\alpha_0 + \beta_0 \cdot Y_{\mathbf{x}} + \tilde{h}_0(Y_{\mathbf{x}})}_{\text{unconditional part}} + \underbrace{Y_{\mathbf{x}} \cdot \sum_{j=1}^{J} \beta_j(\mathbf{x}) + \sum_{j=1}^{J} \alpha_j(\mathbf{x})}_{\text{conditional part}} \tag{9}$$

Hence, in this model, only the linear part of the response transformation ($= Y_{\mathbf{x}}$) may depend on explanatory variables, whereas the function representing deviations from linearity $\tilde{h}_0(Y_{\mathbf{x}})$ is flexible and depends only on the response values $Y_{\mathbf{x}}$. In accordance with the definition of CLTMs, the explanatory variables solely influence the mean and variance of the transformed responses. We denote the coefficients $\beta_{j,\text{lin}}(\mathbf{x})$, $j = 1, \ldots, J$ (equation (8)) simply by $\beta_j(\mathbf{x})$ as we no longer need to distinguish the linear and the more complex part of the coefficient vector. In this model, we can estimate further characteristics of the conditional distribution function of the response (e.g. skewness and kurtosis) in terms of $\tilde{h}_0(Y_{\mathbf{x}})$.

By further differentiating between linear and flexible explanatory variable effects, we get

Linear CLTM

$$h(Y_{\mathbf{x}}|\mathbf{x}) = \alpha_0 + \beta_0 \cdot Y_{\mathbf{x}} + \tilde{h}_0(Y_{\mathbf{x}}) + Y_{\mathbf{x}} \cdot \sum_{j=1}^{J} \beta_j \cdot \mathbf{x}_j + \sum_{j=1}^{J} \alpha_j \cdot \mathbf{x}_j$$

where α_j and β_j, $j = 1, \ldots, J$ are regression coefficients, and therefore the explanatory variables have a linear influence on the response transformation.

Additive CLTM

$$h(Y_{\mathbf{x}}|\mathbf{x}) = \alpha_0 + \beta_0 \cdot Y_{\mathbf{x}} + \tilde{h}_0(Y_{\mathbf{x}}) + Y_{\mathbf{x}} \cdot \sum_{j=1}^{J} \beta_j(\mathbf{x}) + \sum_{j=1}^{J} \alpha_j(\mathbf{x})$$

where $\alpha_j(\mathbf{x})$ and $\beta_j(\mathbf{x})$, $j = 1, \ldots, J$ denote smooth functions. Hence, the explanatory variables have a flexible influence on the response transformation.

3.1.1 Introduction of specific CLTMs for the analysis of the Perinatal Database Erlangen

For the analysis, we chose six variants of CLTMs with unknown response transformation CLTM 0 (linear) and CLTM 0–CLTM 4, in which the models are ordered with increasing model complexity

Table 1. Overview: Relevant CLTMs (conditionally linear transformation models) and CTM (conditional transformation model).

	Linear Expl. variable effects		Flexible Expl. variable effects				
Model	$\alpha(x)$	$\beta(x)$	$\alpha(x)$	$\beta(x)$	Linear uncond. Transf. function	Flexible uncond. Transf. function	Higher moments Depend on expl. variables
CLTM 0 (linear)	×					×	
CLTM 0			×			×	
CLTM 1	×	×			×		
CLTM 2	×	×				×	
CLTM 3			×	×	×		
CLTM 4			×	×		×	
CTM			×	×			×

(Table 1). For comparison, we used the common conditional transformation model CTM as a reference model representing the most complex modeling approach.

3.1.2 CLTM 0 (linear): linear transformation model

$$h(Y_{\mathbf{x}}|\mathbf{x}) = Y_{\mathbf{x}} + \tilde{h}_0(Y_{\mathbf{x}}) + \sum_{j=1}^{J} \alpha_j \cdot \mathbf{x}_j \overset{\text{Equation 7}}{=} h_0(Y_{\mathbf{x}}) + \sum_{j=1}^{J} \alpha_j \cdot \mathbf{x}_j$$

CLTM 0 (linear) is denoted *Linear Transformation Model* because it belongs to the class of well-known linear transformation models (equation (6)). The transformation function h is decomposed into a flexible function $h_0(Y_{\mathbf{x}})$ depending only on the response values $Y_{\mathbf{x}}$ and a part depending only on the explanatory variables. The coefficients α_j induce linear shifts of the response transformation depending on the explanatory variables $\mathbf{x}_j$, $j = 1, \ldots, J$. The flexible response transformation $h_0(Y_{\mathbf{x}})$ is restricted to monotone functions. The transformation function results from a linear CLTM if we set $\alpha_0 = 0$, $\beta_0 = 1$ and $\beta_j = 0$, $j = 1, \ldots, J$.

In the conditional distribution function of BW, these definitions result in fetus-specific means that depend linearly on the ultrasound measurements. Beyond that, the BWs might follow some arbitrary distribution function because higher moments are modeled flexibly. The corresponding class of distribution functions is the same for all fetuses because the deviations from the normal distribution are not influenced by any ultrasound measurements.

3.1.3 CLTM 0: linear transformation model with flexible explanatory variable effects

$$h(Y_{\mathbf{x}}|\mathbf{x}) = Y_{\mathbf{x}} + \tilde{h}_0(Y_{\mathbf{x}}) + \sum_{j=1}^{J} \alpha_j(\mathbf{x}) \overset{\text{Equation 7}}{=} h_0(Y_{\mathbf{x}}) + \sum_{j=1}^{J} \alpha_j(\mathbf{x})$$

CLTM 0 also represents a linear transformation model, but the influence of the explanatory variables is modeled in terms of smooth functions $\alpha_j(\mathbf{x})$, $j = 1, \ldots, J$. This results in flexible shifts

of the response transformation depending on the explanatory variables. The flexible response transformation $h_0(Y_\mathbf{x})$ is again restricted to monotone functions. This transformation function results from an additive CLTM if we set $\alpha_0 = 0$, $\beta_0 = 1$ and $\beta_j = 0$, $j = 1, \ldots, J$.

Based on CLTM 0, fetus-specific means result that depend flexibly on the ultrasound measurements. Moreover, the BWs may follow some arbitrary distribution, but the corresponding class of distribution functions is again the same for all fetuses. Thus, model CLTM 0 describes a very general but easy interpretable set of distributions. The explanatory variables have an additive influence only on the conditional mean and the response distribution belongs to the rich set of distributions that can be generated form the normal distribution via a monotone transformation.

3.1.3 CLTM 1: CLTM with linear explanatory variable effects and linear unconditional response transformation

$$h(Y_\mathbf{x}|\mathbf{x}) = \alpha_0 + \beta_0 \cdot Y_\mathbf{x} + Y_\mathbf{x} \cdot \sum_{j=1}^{J} \beta_j \cdot \mathbf{x}_j + \sum_{j=1}^{J} \alpha_j \cdot \mathbf{x}_j$$

This is a linear CLTM in which $\tilde{h}_0(Y_\mathbf{x})$ is cancelled, and, therefore, the unconditional part of the response transformation is linear in $Y_\mathbf{x}$. Hence, conditional on the explanatory variables $\mathbf{x}$, the whole conditional transformation function $h(Y_\mathbf{x}|\mathbf{x})$ is linear in $Y_\mathbf{x}$. As we cancelled the deviations from linearity $\tilde{h}_0(Y_\mathbf{x})$, we assumed that the response has a normal distribution function if we additionally set the link function to $F = \Phi$ in equation (2). This is due to the underlying assumption that the coefficients α_j and β_j, $j = 0, \ldots, J$ influence only the mean and variance of the response. These definitions result in normal distribution functions for all fetuses with fetus-specific means and variances that depend linearly on the ultrasound measurements.

3.1.4 CLTM 2: CLTM with linear explanatory variable effects and unconditional response transformation with monotone constraints

$$h(Y_\mathbf{x}|\mathbf{x}) = \underbrace{\alpha_0 + \beta_0 \cdot Y_\mathbf{x} + \tilde{h}_0(Y_\mathbf{x})}_{\text{uncond. trans. function}} + Y_\mathbf{x} \cdot \sum_{j=1}^{J} \beta_j \cdot \mathbf{x}_j + \sum_{j=1}^{J} \alpha_j \cdot \mathbf{x}_j$$

CLTM 2 is also a linear CLTM but is more complex than CLTM 1 as the unconditional response transformation is a flexible monotone function. We suggest that the distribution function of the response possibly does not belong to the class of normal distributions if we additionally set the link to $F = \Phi$. This is due to the term describing deviations from linearity $\tilde{h}_0(Y_\mathbf{x})$, which is able to affect higher moments of the distribution function of the response.

Hence, the BWs follow some arbitrary distribution function because higher moments are modeled flexibly. Nevertheless, the corresponding class of distribution functions is again identical for all fetuses as the deviations from linearity are not influenced by any ultrasound measurements. Moreover, the resulting fetus-specific means and variances depend linearly on the ultrasound measurements.

3.1.5 CLTM 3: CLTM with flexible explanatory variable effects and linear unconditional response transformation

$$h(Y_\mathbf{x}|\mathbf{x}) = \alpha_0 + \beta_0 \cdot Y_\mathbf{x} + Y_\mathbf{x} \cdot \sum_{j=1}^{J} \beta_j(\mathbf{x}) + \sum_{j=1}^{J} \alpha_j(\mathbf{x})$$

This model is an additive CLTM with $\tilde{h}_0(Y_\mathbf{x}) = 0$. Again, the unconditional response transformation is a linear function (compare CLTM 1), and we therefore implicitly assumed that the response follows a normal distribution. Therefore, these definitions result in normal distribution functions for all fetuses with fetus-specific means and variances that depend flexibly on the ultrasound measurements.

3.1.6 CLTM 4: CLTM with flexible explanatory variable effects and unconditional response transformation with monotone constraints

$$h(Y_\mathbf{x}|\mathbf{x}) = \underbrace{\alpha_0 + \beta_0 \cdot Y_\mathbf{x} + \tilde{h}_0(Y_\mathbf{x})}_{\text{uncond. trans. function}} + Y_\mathbf{x} \cdot \sum_{j=1}^{J} \beta_j(\mathbf{x}) + \sum_{j=1}^{J} \alpha_j(\mathbf{x})$$

Also, this model is an additive CLTM and is the most complex CLTM considered. Comparable to CLTM 3, the influence of the explanatory variables on the linear response transformation is modeled flexibly. Additionally, the unconditional response transformation is a flexible monotone function (compare CLTM 2), in which we implicitly assumed that the response may not follow a normal distribution.

Hence, we assumed fetus-specific means and variances that depend flexibly on the ultrasound measurements. Again, BWs for all fetuses follow some arbitrary distribution because higher moments are modeled flexibly, but the corresponding class of distribution functions is the same for all fetuses.

3.1.7 CTM: Conditional transformation model

$$h(Y_\mathbf{x}|\mathbf{x}) = \sum_{j=1}^{J} h_j(Y_\mathbf{x}|\mathbf{x}) \tag{10}$$

We define the common CTM[13] as our reference model because it represents a more general and more complex model class than the considered CLTMs. The transformation function $h(Y_\mathbf{x}|\mathbf{x})$ is decomposed additively into J partial transformation functions without any further restrictions. Thereby, we assume additivity on the scale of the transformation function, which is fundamentally different to additive mean or quantile regression, where additivity is assumed on the scale of the conditional mean or quantile function. Simulation results presented in Hothorn et al.[13] show a better performance of CTMs compared to the parametric generalized additive models for location, scale and shape (GAMLSS) and to nonparametric kernel estimators. Since CTMs are an alternative to quantile regression models, the authors also compared the two approaches and assessed that both model classes are equally flexible. Nevertheless, CTMs have the advantages of being based on differentiable and convex proper scoring rules as risk functions that allow relatively easy optimization algorithms to be applied, the simultaneous estimation of all quantiles in a joint model, and the dependency on only one hyperparameter (the number of boosting iterations), compared to additive quantile regresssion. Based on this CTM, we defined the model class of CLTMs and finally the special cases of CLTMs presented above.

3.2 Model estimation

First, we will briefly describe the model estimation in CTMs (equation (10)) and then present the necessary adaptations for CLTMs. In Hothorn et al.,[13] a parametrization of the partial

transformation functions h_j, $j = 1, \ldots, J$ in CTMs via basis functions is presented and illustrates the high flexibility of the partial transformation functions in both the response variable and the explanatory variables. For example, the j-th partial transformation function is parametrized as follows:

$$h_j(Y_{\mathbf{x}}|\mathbf{x}) = \left(\mathbf{b}_j(\mathbf{x})^\top \otimes \mathbf{b}_0(Y_{\mathbf{x}})^\top\right)\boldsymbol{\gamma}_j \tag{11}$$

where $\mathbf{b}_0$ is a basis along the grid of response values $Y_{\mathbf{x}}$, and $\mathbf{b}_j$ is a basis along a grid of explanatory variables $\mathbf{x}$. The two sets of basis functions are connected via a Kronecker product, thereby establishing an interaction surface between the basis for the response and the basis for the explanatory variables. The basis $\mathbf{b}_0$ defines the functional form of the response transformation (i.e. a linear or flexible response transformation), and the functional form of $\mathbf{b}_j$ defines how this response transformation is influenced by the explanatory variables (i.e. the response transformation varies linearly or flexibly with varying explanatory variables).[13] For example, if one chooses linear basis functions for $\mathbf{b}_0$, one gets a linear response transformation, and if one chooses B-spline basis functions for $\mathbf{b}_0$, one gets a flexible response transformation. Hence, the user is free to choose a very complex and general model framework (e.g. by choosing a B-spline basis for $\mathbf{b}_0$ and $\mathbf{b}_j$) in CTMs, which often ends up in a lack of interpretability (see Section 3.1). In CTMs, one aims at obtaining an estimate for each partial transformation function h_j that is smooth in both the response and the explanatory variables, which is achieved by imposing an appropriate penalty on the Kronecker product of basis functions in equation (11). For further details on parametrization and penalty specification, see Hothorn et al.[13]

In CTMs, model estimation is based on the minimization of the log score

$$LS = -\frac{1}{N \cdot n}\sum_{i=1}^{N}\sum_{\iota=1}^{n} I(\mathrm{BW}_i \leq \upsilon_\iota)\log(F(h(\upsilon_\iota|\mathbf{x}_i))) + I(\mathrm{BW}_i > \upsilon_\iota)\log(1 - F(h(\upsilon_\iota|\mathbf{x}_i))) \tag{12}$$

which is a proper scoring rule.[13,27] The log score measures the mismatch between the individual empirical distribution functions of subjects $i = 1, \ldots, N$ and the corresponding probabilities of the conditional distribution function $F(h(\upsilon_\iota|\mathbf{x}_i))$ resulting from the CTM in terms of the negative binomial log-likelihood. The score is evaluated on a grid of BWs $\upsilon_1, \ldots, \upsilon_n$ covering their range. As CLTMs are a special case of CTMs, we used the same approach for model estimation. All we had to adapt is the parametrization of the partial transformation functions in equation (11), which is straightforward.

The choice of the functional form of $\mathbf{b}_0(Y_{\mathbf{x}})$ and $\mathbf{b}_j(\mathbf{x})$, $j = 1, \ldots, J$ (either linear or flexible basis functions) depends on the definition of the conditional transformation function $h(Y_{\mathbf{x}}|\mathbf{x})$. As an example, we present the parametrization of transformation model CLTM 0 given in the previous subsection. CLTM 0 can be decomposed into the unconditional transformation function $h_0(Y_{\mathbf{x}})$ that depends only on the response values and the part $\alpha(\mathbf{x}) = \sum_{j=1}^{J}\alpha_j(\mathbf{x})$ that depends only on the explanatory variables. Both parts of the transformation function are parametrized separately as special cases of equation (11). First, the unconditional transformation function is parametrized via

$$h_0(Y_{\mathbf{x}}) = \left(\mathbf{1}_N^\top \otimes \mathbf{b}_0(Y_{\mathbf{x}})^\top\right)\boldsymbol{\gamma}$$

where $\mathbf{1}_N$ denotes the one-vector whose length is equal to the number of observations N. Since the unconditional transformation function does not depend on any explanatory variables, the basis functions for the explanatory variables $\mathbf{b}_j(\mathbf{x})$ are replaced by $\mathbf{1}_N$ to maintain correct dimensions.

The basis functions for the response variables $\mathbf{b}_0(Y_\mathbf{x})$ are monotonic B-splines as $h_0(Y_\mathbf{x})$ is assumed to be a flexible monotone function in the response values. Second, the function depending on the explanatory variables is parametrized by the set of basis functions

$$\alpha_j(\mathbf{x}) = \left(\mathbf{b}_j(\mathbf{x})^\top \otimes \mathbf{1}_n^\top\right)\gamma_j,\ j = 1, \ldots, J$$

where $\mathbf{1}_n$ denotes the one-vector with length n, the number of unique υ values (a hyper parameter to the algorithm). As the functions $\alpha_j(\mathbf{x})$, $j = 1, \ldots, J$ do not depend on the response variable, the corresponding basis functions $\mathbf{b}_0(Y_\mathbf{x})$ are replaced by the one-vector to maintain correct dimensions. The basis functions $\mathbf{b}_j(\mathbf{x})$, $j = 1, \ldots, J$ are B-spline basis functions because the explanatory variables have a flexible influence on the mean of the transformed response in CLTM 0. The parametrizations of the other special cases of CLTMs result accordingly.

3.3 Computational details

All analyses were carried out in the R system for statistical computing (version 2.15.3[28]). Model estimation in CLTMs and CTMs was carried out using the R add-on package **ctm.**[29] To compare our proposed transformation models and established methods, we estimated a linear regression model, linear quantile regression model, and additive quantile regression model. To estimate the linear regression model, we used the `lm` function in the **stats** package and fitted the linear quantile regression model using the `rq` function of the **quantreg** package.[30] We used component-wise boosting for the estimation of the additive quantile regression model[31] in the **mboost** package.[32] A tutorial R example `ex_fetus_CLTM.Rnw` including the code for estimating the proposed regression and transformation models, the calculation of intervals for the BW, and the generation of Figure 1 is publicly available in the **ctm** package from the R-forge repository (https://r-forge.r-project.org/projects/ctm).

3.4 Evaluation of fetus-specific prediction formulas for BW

As we are interested in reliable prediction intervals for BWs (see Section 1), we calculated fetus-specific prediction intervals based on equation (1) with a coverage probability of 80%. A further goal was to identify the C(L)TM that described the Perinatal Database Erlangen best among the proposed C(L)TMs in Section 3.1.1. We considered certain aspects of model misspecification.

For the construction of prediction intervals, we considered the conditional median and the conditional $\alpha/2$ quantile and $1-\alpha/2$ quantile representing the point prediction for the BW and the boundaries of the fetus-specific prediction intervals in equation (1). Therefore, we used the well-known relationship between the conditional distribution function and the conditional quantile function to extract the relevant quantiles:

$$q_\tau(\mathbf{x}) = F^{-1}_{\mathrm{BW}|\mathbf{X}=\mathbf{x}}(\tau)$$

where $\tau = \{\alpha/2, 0.5, 1-\alpha/2\}$ denotes the quantiles of interest and $F_{\mathrm{BW}|\mathbf{X}=\mathbf{x}}$ is defined in equation (2).[20]

In the analysis of the Perinatal Database Erlangen, we used 10 regression or transformation models to estimate the median BW and the associated interval borders. The transformation models used encompass a standard CTM and the six CLTMs [CLTM 0 (linear) and CLTM 0–CLTM 4] presented in Section 3.1.1. For comparison, we also considered a linear regression

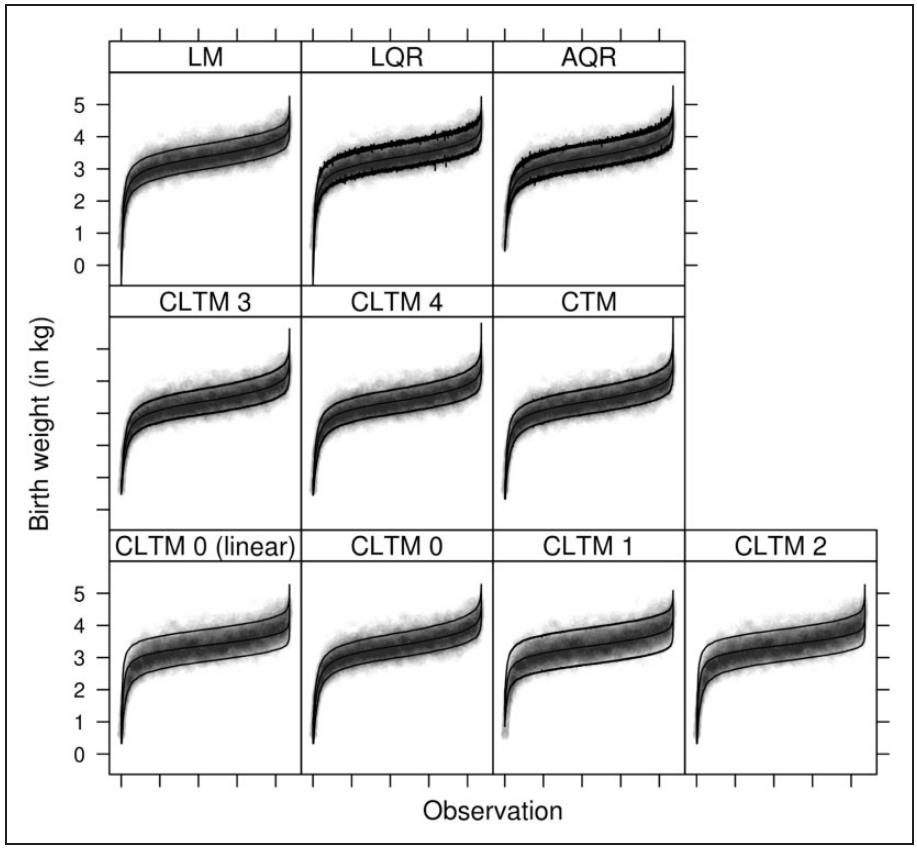

Figure 1. Birth weight prediction. Observed birth weights of 8712 newborns (dots) ordered with respect to the predicted conditional mean (LM only) or median birth weight (central black line). The lower and upper black lines display estimated 10% and 90% quantiles of birth weights, respectively. The areas in-between represent fetus-specific 80% prediction intervals. Each subplot shows the results for one of the regression or conditional transformation models. LM, linear model; LQR, linear quantile regression; AQR, additive quantile regression; CLTMs (CLTM 0–CLTM 4), conditionally linear transformation models; CTM, conditional transformation model.

model (LM), which served as a standard procedure in the past, a linear quantile regression model and an additive quantile regression model (LQR and AQR).

A common strategy to check the adequacy of prediction intervals is to check their coverage probability. When we defined prediction intervals in Section 2.3, we stated that a correctly specified prediction interval $\mathrm{PI}_{1-\alpha}(\mathbf{x})$ for a new set of ultrasound parameters $\mathbf{x}$ covers a new observation BW with high probability $1-\alpha$. The correct measure to evaluate prediction intervals adequately is the conditional coverage.[20] Therefore, we checked whether for any particular combination of ultrasound measurements $\mathbf{x}$ about $(1-\alpha)\cdot 100\,\%$ of the corresponding observations $(\mathrm{BW}_1, \mathbf{x}), \ldots, (\mathrm{BW}_M, \mathbf{x})$ were covered by the prediction interval $\mathrm{PI}(\mathbf{x})$

$$\hat{\pi}|\mathbf{x} = \widehat{\mathbb{E}}(\mathrm{BW} \in \mathrm{PI}(\mathbf{x})|\mathbf{X} = \mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M} I\{\mathrm{BW}_i \in \mathrm{PI}(\mathbf{x})\} \tag{13}$$

where I denotes the indicator function. The conditional coverage reflects what we really expect from a prediction interval because the prediction interval for a specific combination of ultrasound parameters should cover the BWs of 80% of the fetuses with exactly the same ultrasound measurements.[20]

In practice, the evaluation of the conditional coverage of prediction intervals is impossible because we usually only have one observation for each combination of ultrasound parameters $\mathbf{x}$ and more are needed with exactly the same combination of ultrasound measurements (equation (13)). Especially in a regression setting with continuous explanatory variables, multiple response values for each combination of explanatory variables are unlikely to occur. Therefore, we calculated the conditional coverage of our prediction intervals using binned observations:

(1) We used the ultrasound parameters AC and FL to divide the fetuses in the database into categories because these two parameters are essential for the prediction of BWs.[7,10,12,16] AC and FL were divided quantile-based into categories, resulting in five AC categories measured in cm ($\mathbf{1}$: (175, 316]; $\mathbf{2}$: (316, 331]; $\mathbf{3}$: (331, 343]; $\mathbf{4}$: (343, 357]; $\mathbf{5}$: (357, 428]) and five FL categories measured in cm ($\mathbf{1}$: (31.1, 69.6]; $\mathbf{2}$: (69.6, 71.7]; $\mathbf{3}$: (71.7, 73.4]; $\mathbf{4}$: (73.4, 75.4]; $\mathbf{5}$: (75.4, 86.6]).
(2) When we combined the five AC and five FL categories, we get 25 categories of fetuses, which results in good sample sizes of at least 102 observations for all groups. The distribution of the BWs in the respective categories are displayed in Figure 9 (see Appendix 1).
(3) To assess the conditional coverage, we generated a training data set by randomly choosing 90% of the fetuses in each of the 25 categories and generated a validation data set by choosing the remaining fetuses. We then estimated CLTM 0 (linear)—CLTM 4, CTM, LM, LQR, and AQR for the training data, and predicted the BWs for the validation data set for each of the models. We assessed the conditional coverage (equation (13)) for each of the regression and transformation models in each of the 25 categories.

In addition to the conditional coverage of the prediction intervals, we also checked their average interval lengths.

To identify the C(L)TM that described the Perinatal Database Erlangen best, we compared the performance among all CLTMs to the performance of the CTM and the LM. We fitted the models on a training data set and evaluated their predictive ability on an evaluation data set. Twenty-five training and evaluation data sets were generated by choosing randomly 50% of the original observations in each AC–FL category. The predictive ability was measured in terms of the log score given in equation (12), which was used to evaluate the conditional distribution function for the whole evaluation data set and for each AC-FL category separately. As the complexities of the C(L)TMs differed, this procedure could also be used to reveal model misspecifications. We were able to detect missing covariate effects on the variance (e.g. CLTM 0 against all other C(L)TMs), missing flexibility of the covariate effects on the mean or the variance (e.g. CLTM 2 against CLTM 4), and missing flexibility of the response transformation (e.g. CLTM 1 against CLTM 2). If even higher moments of the conditional distribution function were affected by the explanatory variables, it could be checked by comparing all CLTMs to the CTM, and by comparing all CLTMs to the LM if the assumption of a normal distribution with constant variance works for the database. The out-of-sample log score cannot be calculated for the quantile regression models because quantile crossing makes the inversion of the quantile function into a distribution function impossible.

4 Results

4.1 Estimated transformation and regression models

All ultrasound parameters were included as main effects in the model equations of the regression and transformation models. One exception was the interaction between AC and FL, which has been important in many earlier prediction formulas for BW.[15] Therefore, we additionally included this interaction in models CLTM 0 (linear), CLTM 1, CLTM 2, LM, LQR, and AQR; we did not include this interaction in models CLTM 0, CLTM 3, CLTM 4, and CTM because the model estimation became too complex.

The estimates of the BWs based on the prenatal ultrasound parameters are displayed in Figure 1. In model LM, symmetric intervals around the estimated conditional mean with equal interval lengths for all fetuses resulted, and possible heteroscedasticity, kurtosis, and skewness were ignored. Despite these restrictive assumptions, model LM provided satisfying and narrow intervals. We concluded that deviations from normality were small and no strong heteroscedasticity occurred. Nevertheless, we pursued further model improvements.

The quantile regression approaches (LQR and AQR) also provided satisfying results associated with narrow intervals. The wiggly estimates for the interval borders were due to the separate estimation of the quantiles. In contrast, smooth interval borders resulted for C(L)TMs because all quantiles were estimated simultaneously.

In CLTM 0 (linear), the influence of the ultrasound parameters on the conditional mean was modeled linearly, comparable to model LM. Owing to the unconditional transformation function, also a possible skewness and kurtosis of the distribution of the BWs can be modeled. This led to wider intervals for CLTM 0 (linear) compared to LM, especially for extreme BWs. In model CLTM 0, the influence of the ultrasound measurements on the conditional mean was modeled flexibly, and thus the corresponding fetus-specific intervals were narrower than with CLTM 0 (linear).

In general, a flexible inclusion of the ultrasound parameters seems advisable because the intervals with models CLTM 0, CLTM 3, and CLTM 4 were narrower than with CLTM 1 and CLTM 2. Besides, in CLTM 1–CLTM 4, the ultrasound parameters may influence the conditional mean and conditional variance. Hence, these models accounted for possible heteroscedasticity induced by the ultrasound measurements.

An additional slight improvement was gained by estimating the unconditional transformation function in terms of a flexible monotone function and thus accounting for possible kurtosis and skewness. This can be observed by direct comparison of CLTM 1 and CLTM 2 and of CLTM 3 and CLTM 4. Nevertheless, deviations from normality seemed to be small since the associated improvements were minor.

We were also interested in identifying the C(L)TM that described the Perinatal Database Erlangen best. We calculated the out-of-sample log scores based on 25 evaluation data sets for the proposed C(L)TMs and the LM to evaluate the estimated conditional distribution functions for new observations for the whole evaluation data set (Figure 2) and for each AC–FL category separately (Figures 10 and 11 in Appendix 1). The results were in accordance with those in Figure 1: the out-of-sample log scores of CLTM 0, CLTM 3, CLTM 4, CTM, and LM were similar, whereas those of CLTM 0 (linear), CLTM 1, and CLTM 2 were clearly lower. Hence, the inclusion of flexible covariate effects clearly improves the estimated conditional distribution functions. On the other hand, consideration of heteroscedasticity, deviations from the normality assumption, and higher moments depending on explanatory variables were of minor importance, which was also supported by the good performance of the LM.

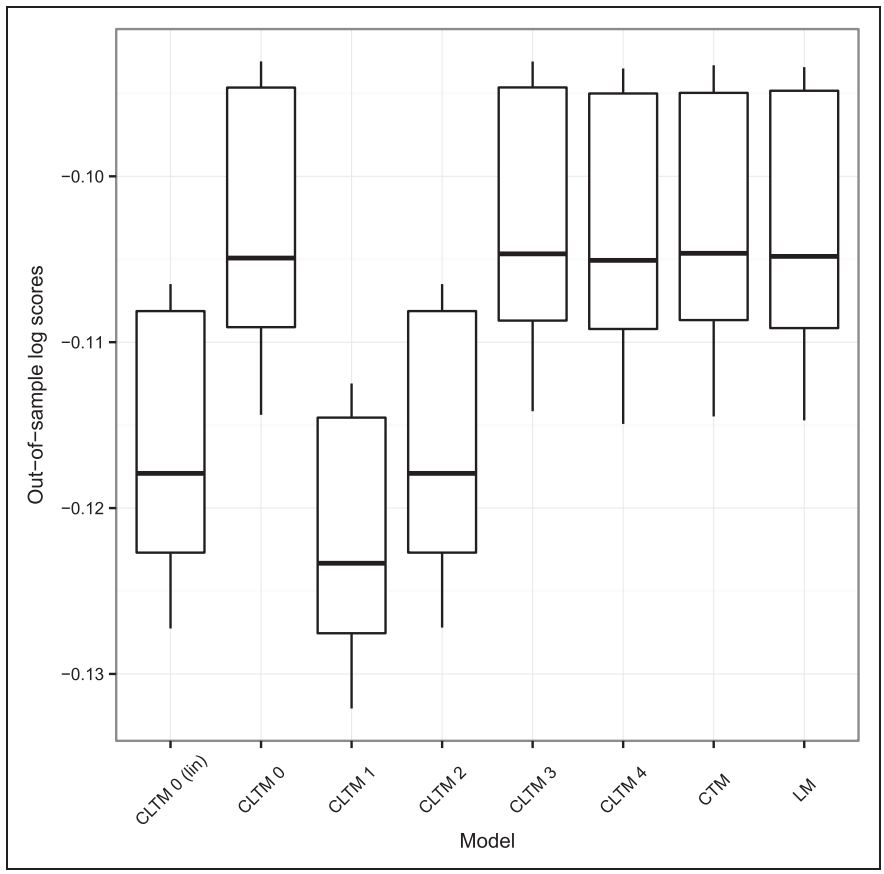

Figure 2. Out-of-sample log scores for CLTM 0–CLTM 4, LM, and CTM based on 25 randomly chosen evaluation data sets consisting of 4355 observations.

To further illustrate important characteristics of CLTMs, we more closely examined CLTM 4, which is the most flexible among all considered CLTMs. The influence of the ultrasound measurements on the conditional mean and conditional variance was modeled flexibly, and the unconditional response transformation was modeled as a flexible monotone function. We assumed that the response values most likely do not follow a normal distribution, as the following results indicated.

Low BWs did not exactly follow a normal distribution, i.e. the resulting estimated unconditional transformation function showed deviations from a linear function for low BWs (see equation (7)), whereas medium and high BWs followed a normal distribution (Figure 6 in Appendix 1). Therefore, the response values for low BWs needed to be transformed.

This conclusion can be observed clearly in normal quantile–quantile plots for original and transformed BWs resulting from model CLTM 4 (Figure 7 in Appendix 1). Low original BWs deviated from the normal distribution, but low transformed BWs approximately followed a normal distribution. A scatterplot showing the relationship between original and transformed BWs (Figure 8 in Appendix 1) also revealed similar results. Medium and high BWs scattered

unsystematically around some linear function, whereas low BWs deviated, which indicated that a non-linear transformation took place. Moreover, a kernel density plot (Figure 8 in Appendix 1) shows that the estimated density of the transformed BWs is in good accordance with the corresponding density of the normal distribution.

These results together indicated that those regression models that allow deviations from the normal distribution assumption are more reliable when original data do not entirely follow a normal distribution.

We stressed that the main advantage of CLTMs over CTMs is the improved interpretability of the estimated effects of ultrasound measurements on moments of the distribution function of BWs. The estimated effects of ultrasound parameters for model CLTM 4 (Figure 3) can be interpreted according

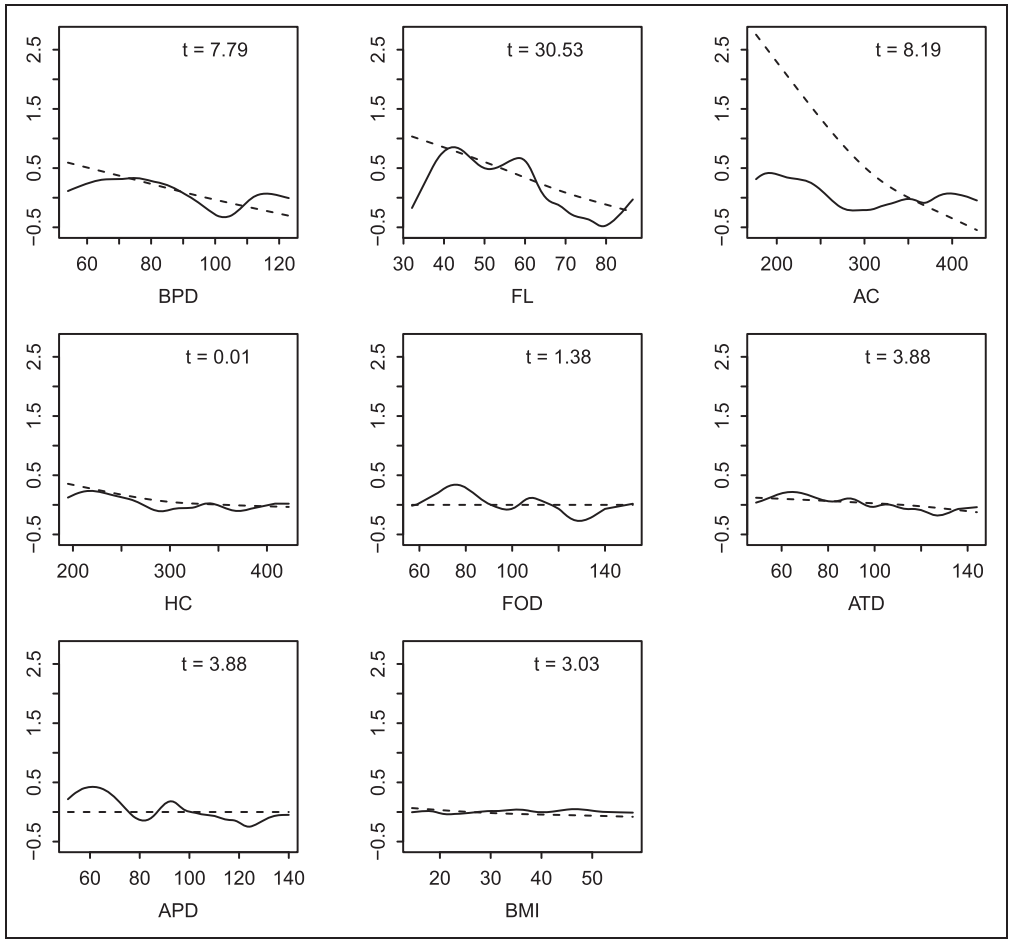

Figure 3. Estimated effects of ultrasound parameters on the conditional mean and conditional variance of transformed birth weights. Solid lines represent estimated functions $\hat{\alpha}$(ultrasonic parameter) and dashed lines represent estimated functions $\hat{\beta}$(ultrasonic parameter). The corresponding values of *t*-statistics belong to the coefficients of the ordinary linear model LM. BPD: biparietal diameter, FL: femur length, AC: abdominal circumference, HC: head circumference, FOD: fronto-occipital diameter, ATD: abdominal transverse diameter, APD: anterior–posterior abdominal diameter, BMI: mother's body mass index.

to equation (4). For almost all ultrasound parameters, estimated non-linear functions α and β resulted, which suggested that the ultrasound parameters influence both the conditional mean and conditional variance. This again argues for the presence of heteroscedasticity that increases with increasing BWs.

4.2 Assessing the accuracy of the prediction intervals

We assessed the accuracy and adequacy of the (fetus-specific) prediction intervals by calculating the conditional coverage and average interval length as quality criteria.

The conditional coverage of the prediction intervals for the BWs (Figures 4 and 5; Tables 3 and 4 in Appendix 1) is a measure to check the adequacy and correctness of estimated prediction intervals.

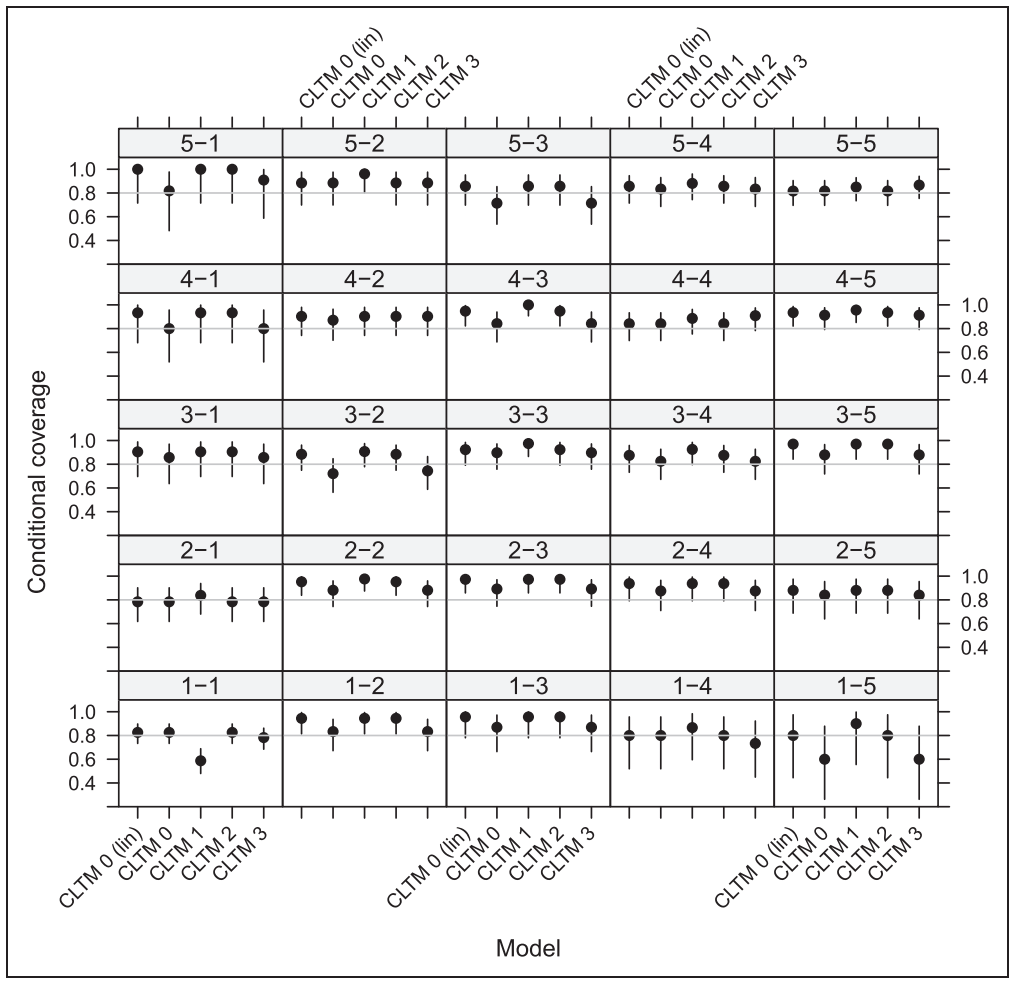

Figure 4. Conditional coverage of the prediction intervals for fetuses of the 25 AC–FL categories. Points refer to the point estimates of the conditional coverage, and error bars display corresponding Clopper-Pearson confidence intervals. Gray reference lines symbolize the postulated 80% confidence level. Model estimation was carried out with CLTM 0 (linear), CLTM 0, CLTM 1, CLTM 2, and CLTM 3.

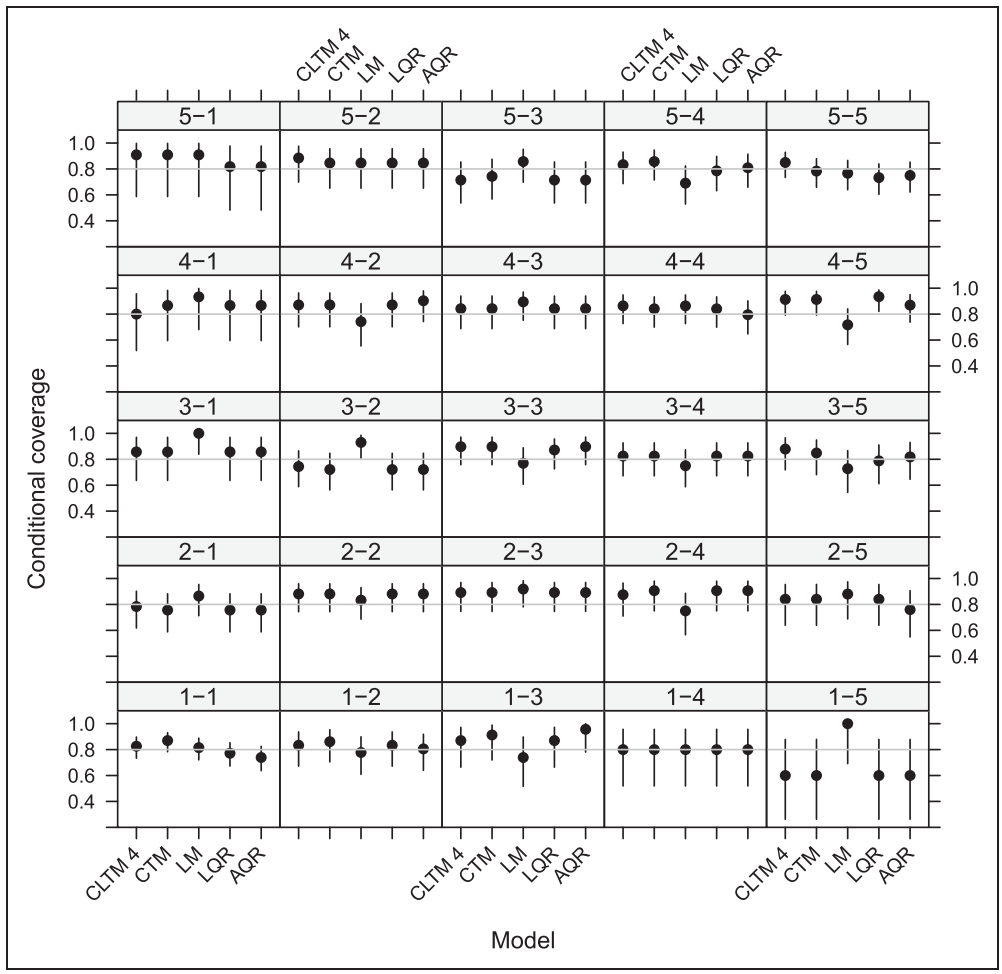

Figure 5. Conditional coverage of the prediction intervals for fetuses of the 25 AC–FL categories. Points refer to the point estimates of the conditional coverage, and error bars display corresponding Clopper-Pearson confidence intervals. Gray reference lines symbolize the postulated 80% confidence level. Model estimation was carried out with CLTM 4, CTM, LM, LQR, and AQR.

We were interested in how often the postulated coverage probability of 80% was violated in the 25 AC and FL categories (defined in Section 3.4) for the 10 regression models. Moreover, the accuracy of the prediction intervals can be measured by the average interval lengths given in Table 2.

The conditional coverage of all 10 models was satisfying. The postulated coverage probability of 80% was not significantly violated by any of the suggested models in any of the categories. The length of the corresponding error bars was mainly determined by the number of fetuses used for estimation. Hence, the length of the error bars was especially high in the categories 5–1, 4–1, 1–5, and 1–4.

The smallest associated average interval lengths were found for CLTM 3, CLTM 0, LM, CLTM 4, LQR, and AQR (Table 2). Hence, regarding the accuracy of prediction intervals, our new model class of CLTMs can compete with linear regression models and quantile regression approaches.

Table 2. Average prediction interval length.

Model	Average interval length
CLTM 0 (linear)	1.042
CLTM 0	0.785
CLTM 1	1.132
CLTM 2	1.042
CLTM 3	0.790
CLTM 4	0.776
CTM	0.807
LM	0.777
LQR	0.764
AQR	0.755

Note: Estimation is based on models CLTM 0 (linear), CLTM 0–CLTM 4, CTM, LM, LQR, and AQR. CLTM: conditionally linear transformation models; CTM: conditional transformation model; LM: linear regression model; LQR: linear quantile regression model; AQR: additive quantile regression model.

5 Discussion

Although the accurate prediction of BW is one of the most important issues in gynecology, traditional prediction formulas focus on point predictions and an easy-to-interpret, correct measure of quantifying prediction uncertainty is lacking. We therefore aimed at finding a new model-based strategy to predict BWs based on prenatal ultrasound parameters, accompanied by some measure of prediction uncertainty. We introduced CLTMs—a new model class that not only results in point estimates for the median BW but also provides a measure of uncertainty in terms of prediction intervals.

Especially BWs at the extremes have been over- or underestimated by prediction formulas presented earlier.[7] This could be due to the use of linear regression models for estimation, which are not able to deal with possible heteroscedasticity, kurtosis, or skewness of the response distribution, and are accordingly inadequate in such situations. The standard approach around this problem is the use of quantile regression approaches as no distributional assumptions are made, but one often has to deal with the problem of quantile crossing instead.[25]

In our novel approach of estimating CLTMs, we modeled the conditional distribution function of BW based on ultrasound measurements. Hence, all quantiles were estimated simultaneously, and problems such as quantile crossing were avoided. Koenker [23] already suggested the direct estimation of the conditional distribution function via transformation models as an alternative to quantile regression models. The flexibility of the influence of the ultrasound parameters on the quantiles in CLTMs is similar to the flexible influence in quantile regression, as the ultrasound measurement effects may also vary for different values of the conditional distribution function in CLTMs. The borders of the fetus-specific prediction intervals arose directly from the corresponding quantile function. In contrast to linear regression models, the fetus-specific prediction intervals showed individual interval lengths based on the ultrasound measurements and are therefore a useful measure for individual prediction accuracy. Moreover, the variance may depend on explanatory variables, and CLTMs account for possible heteroscedasticity. In addition, CLTMs can deal with skewed distributions as higher moments of the distribution of the response (e.g. kurtosis and

skewness) can be modeled flexibly in terms of the unconditional monotone transformation function. Hence, using CLTMs instead of linear regression models is advantageous in numerous situations, especially in our application of predicting BWs.

From a conceptual point of view, fetal weight estimation is fundamentally different from the construction of reference growth charts of child height and weight.[33] Growth curves are usually designed as screening tools for disease after birth (and also as reference standards for group health and economic status[34]), whereas prediction of BW is designed to estimate the risk of neonatal mortality and morbidity *before* delivery. Consequently, although similar statistical methodology may be used for both tasks, the CLTM approach proposed here specifically addresses the problem of BW prediction but not the construction of reference growth curves.

Our results suggested that the best-performing CLTM variant is able to compete with quantile regression and linear regression approaches in terms of conditional coverage and average length of the prediction intervals.

Although the differences to alternative methods were small, the estimation of C(L)TMs is advisable because of the aforementioned advantages of accounting for possible heteroscedasticity, kurtosis, and skewness. The distribution of the BWs showed deviations from a normal distribution (Figure 7 in Appendix 1), but the deviations were kept within certain limits. Therefore, the linear regression model would not be the worst choice in this application, and we would expect larger differences in favor of C(L)TMs for response variables showing more extreme deviations from normality. Consequently, our results show that prediction intervals for BWs can be derived from a relatively easy and stable model, since the medium and high BWs follow a normal distribution and only small BWs show deviations from normality (Figures 6 and 7 in Appendix 1). This conclusion is also underlined by the good performance of model CLTM 0 (Figure 2). It would have been very hard to derive such insights into the conditional distribution of BWs from alternative models, for example additive quantile regression models. In general, the remarkably good performance of CTMs compared to alternative modeling strategies has already been investigated in simulation studies and numerous applications.[13,35]

Interpretability in CLTMs is different than in linear and quantile regression models. In linear and quantile regression models, the influence of explanatory variables can be interpreted as direct effects on the conditional mean or conditional quantile, respectively. In CLTMs, in contrast, the explanatory variables influence the mean and variance of the transformed response non-linearly (compare equation (4)). Nevertheless, the effects of the explanatory variables are interpretable in CLTMs, which is a main advantage over the more complex model class of CTMs. Moreover, we were primarily interested in predicting BWs accurately, and this is accompanied by correct and precise prediction intervals.

Acknowledgement

Financial support by Deutsche Forschungsgemeinschaft (grant DFG HO 3242-4/1) is gratefully acknowledged.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. McCormick MC. The contribution of low birth weight to infant mortality and childhood morbidity. *N Engl J Med* 1985; **312**: 82–90.
2. Sappenfield WM, Buehler JW, Binkin NJ, et al. Differences in neonatal and postneonatal mortality by race, birth weight, and gestational age. *Public Health Rep* 1987; **102**: 182–192.
3. Boulet SL, Alexander GR, Salihu HM, et al. Macrosomic births in the United States: determinants, outcomes, and proposed grades of risk. *Am J Obstet Gynecol* 2003; **188**: 1372–1378.
4. Ecker JL, Greenberg JA, Norwitz ER, et al. Birth weight as a predictor of brachial plexus injury. *Obstet Gynecol* 1997; **89**: 643–647.
5. Bernstein IM, Horbar JD, Badger GJ, et al. Morbidity and mortality among very-low-birth-weight neonates with intrauterine growth restriction. *Am J Obstet Gynecol* 2000; **182**: 198–206.
6. McIntire DD, Bloom SL, Casey BM, et al. Birth weight in relation to morbidity and mortality among newborn infants. *N Engl J Med* 1999; **340**: 1234–1238.
7. Scioscia M, Vimercati A, Ceci O, et al. Estimation of birth weight by two-dimensional ultrasonography: a critical appraisal of its accuracy. *Obstet Gynecol* 2008; **111**: 57–65.
8. Siemer J, Egger N, Hart N, et al. Fetal weight estimation by ultrasound: comparison of 11 different formulae and examiners with differing skill levels. *Eur J Ultrasound* 2008; **29**: 159–164.
9. Siggelkow W, Schmidt M, Skala C, et al. A new algorithm for improving fetal weight estimation from ultrasound data at term. *Arch Gynecol Obstet* 2011; **283**: 469–474.
10. Hoopmann M, Abele H, Wagner N, et al. Performance of 36 different weight estimation formulae in fetuses with macrosomia. *Fetal Diagnos Ther* 2010; **27**: 204–213.
11. Dammer U, Goecke TW, Voigt F, et al. Sonographic weight estimation in fetuses with breech presentation. *Arch Gynecol Obstet* 2013; **287**: 851–858.
12. Faschingbauer F, Yazdi B, Goecke TW, et al. A new formula for optimized weight estimation in extreme fetal macrosomia ($\geq$ 4500 g). *Eur J Ultrasound* 2012; **33**: 480–488.
13. Hothorn T, Kneib T and Bühlmann P. Conditional transformation models. *J R Statist Soc B* 2014; **76**: 3–27.
14. Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. *Ultrasound Obst Gyn* 2005; **25**: 80–89.
15. Hadlock FP, Harrist RB, Sharman RS, et al. Estimation of fetal weight with the use of head, body, and femur measurements. A prospective study. *Am J Obstet Gynecol* 1985; **151**: 333–337.
16. Sabbagha RE, Minogue J, Tamura RK, et al. Estimation of birth weight by use of ultrasonographic formulas targeted to large-, appropriate-, and small-for-gestational-age fetuses. *Am J Obstet Gynecol* 1989; **160**: 854–862.
17. Schild RL, Maringa M, Siemer J, et al. Weight estimation by three-dimensional ultrasound imaging in the small fetus. *Ultrasound Obst Gyn* 2008; **32**: 168–175.
18. Hart NC, Hilbert A, Meurer B, et al. Macrosomia: a new formula for optimized fetal weight estimation. *Ultrasound Obst Gyn* 2010; **35**: 42–47.
19. Choi J, Lawson AB, Cai B, et al. A Bayesian latent model with spatio-temporally varying coefficients in low birth weight incidence data. *Stat Meth Med Res* 2012; **21**: 445–456.
20. Mayr A, Hothorn T and Fenske N. Prediction intervals for future BMI values of individual children – a non-parametric approach by quantile boosting. *BMC Med Res Meth* 2012; **12**.
21. Montgomery DC, Peck EA and Vining GG. *Introduction to linear regression analysis*. Vol. 821, Hoboken, New Jersey: Wiley, 2012.
22. Koenker R and Portnoy S. Quantile smoothing splines. *Biometrika* 1994; **81**: 673–680.
23. Koenker R. *Quantile regression*. *Economic society monographs*. New York: Cambridge University Press, 2005.
24. Meinshausen N. Quantile regression forests. *J Mach Learn Res* 2006; **7**: 983–999.
25. Dette H and Volgushev S. Non-crossing non-parametric estimates of quantile curves. *J R Statist Soc B* 2008; **70**: 609–627.
26. Cheng SC, Wei LJ and Ying Z. Analysis of transformation models with censored data. *Biometrika* 1995; **82**: 835–845.
27. Gneiting T and Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc* 2007; **102**: 359–378.
28. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria, http://www.R-project.org/ (2012).
29. Hothorn T. *ctm: Conditional transformation models, 2012. R package version 0.0-3,* https://r-forge.r-project.org/projects/ctm.
30. Koenker R. *quantreg: Quantile Regression, 2012. R package version 4.94*, http://CRAN.R-project.org/package=quantreg.
31. Fenske N, Kneib T and Hothorn T. Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *J Am Stat Assoc* 2011; **106**: 494–510.
32. Hothorn T, Bühlmann P, Kneib T, et al. *mboost: Model-Based Boosting; 2013. R package version 2.2-2*, http://CRAN.R-project.org/package=mboost.
33. Cole TJ. Fitting smoothed centile curves to reference data. *J R Statist Soc A* 1988; **151**: 385–418.
34. Wei Y, Pere A, Koenker R, et al. Quantile regression methods for reference growth charts. *Statist Med* 2006; **25**: 1369–1382.
35. Hothorn T, Kneib T and Bühlmann P. Conditional transformation models by example. In: Muggeo VMR, Capursi V, Boscaino G, Lovison G (eds) *Proceedings of the 28th international workshop on statistical modelling*. Palermo, Italy: Universitá Degli Studi Di Palermo, 2013, pp.15–26.

Appendix 1. Predicting birth weight with conditionally linear transformation models

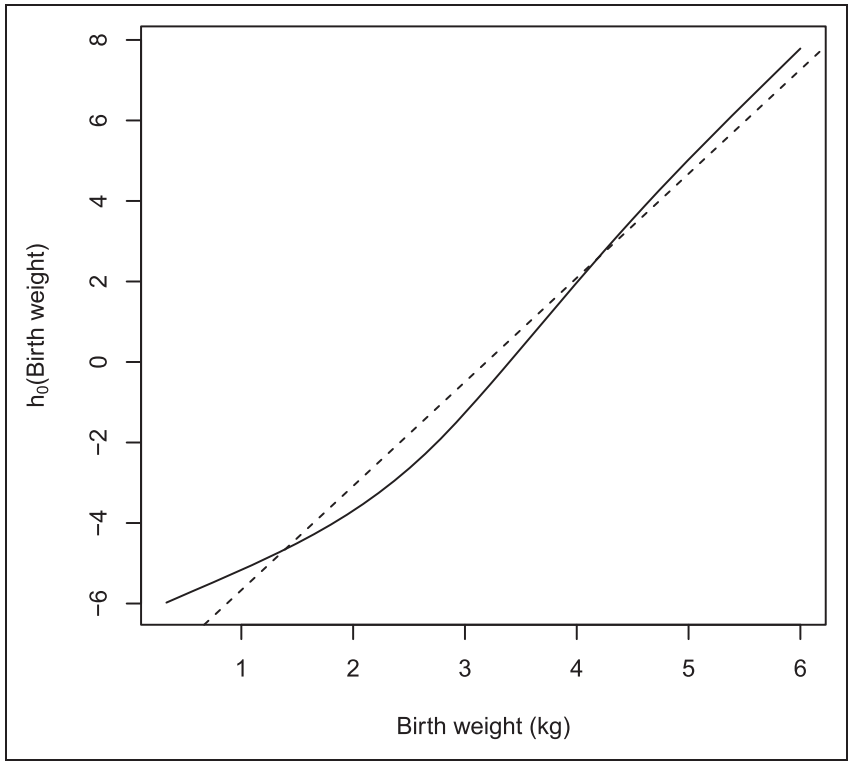

Figure 6. Estimated unconditional monotone transformation function resulting from model CLTM 4. The dashed line symbolizes the linear relationship between the birth weights and their monotone transformation.

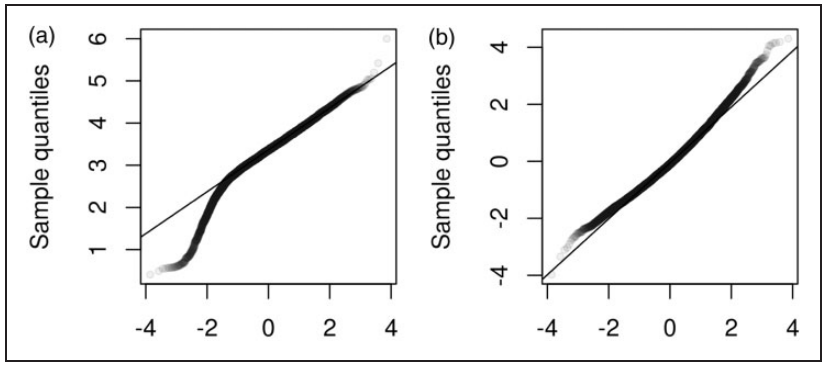

Figure 7. Normal Q-Q plot of (a) original and (b) transformed birth weights resulting from model CLTM 4.

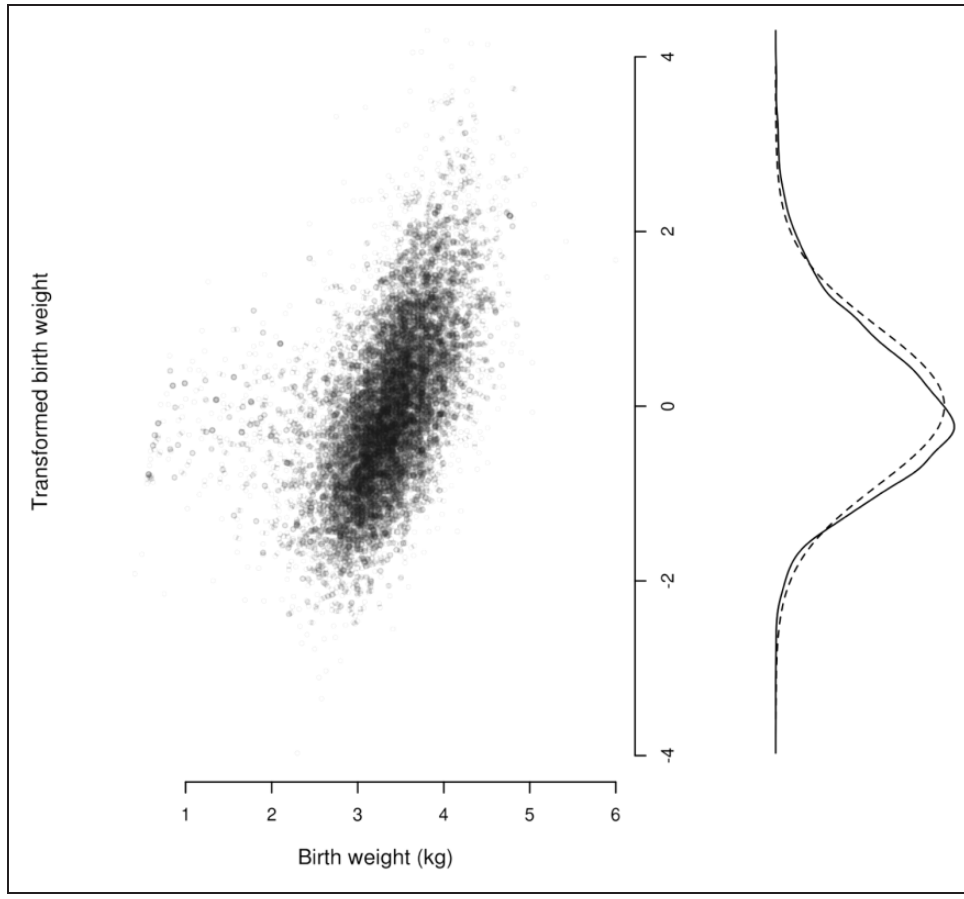

Figure 8. Scatterplot of the original birth weights vs. the transformed birth weights resulting from model CLTM 4 accompanied by a kernel density estimation of the transformed birth weights (solid line) and the corresponding normal density (dashed line).

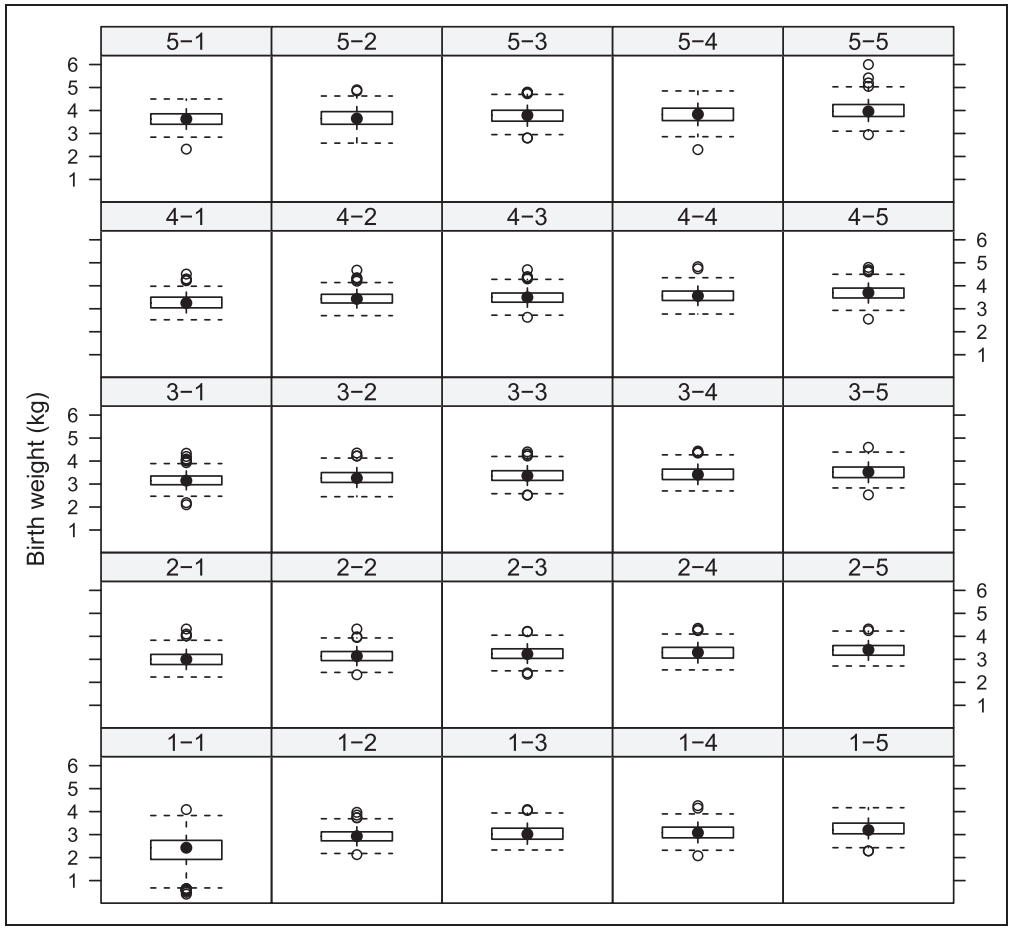

Figure 9. Boxplots for the birth weights in the 25 categories of abdominal circumference and femur length (AC–FL).

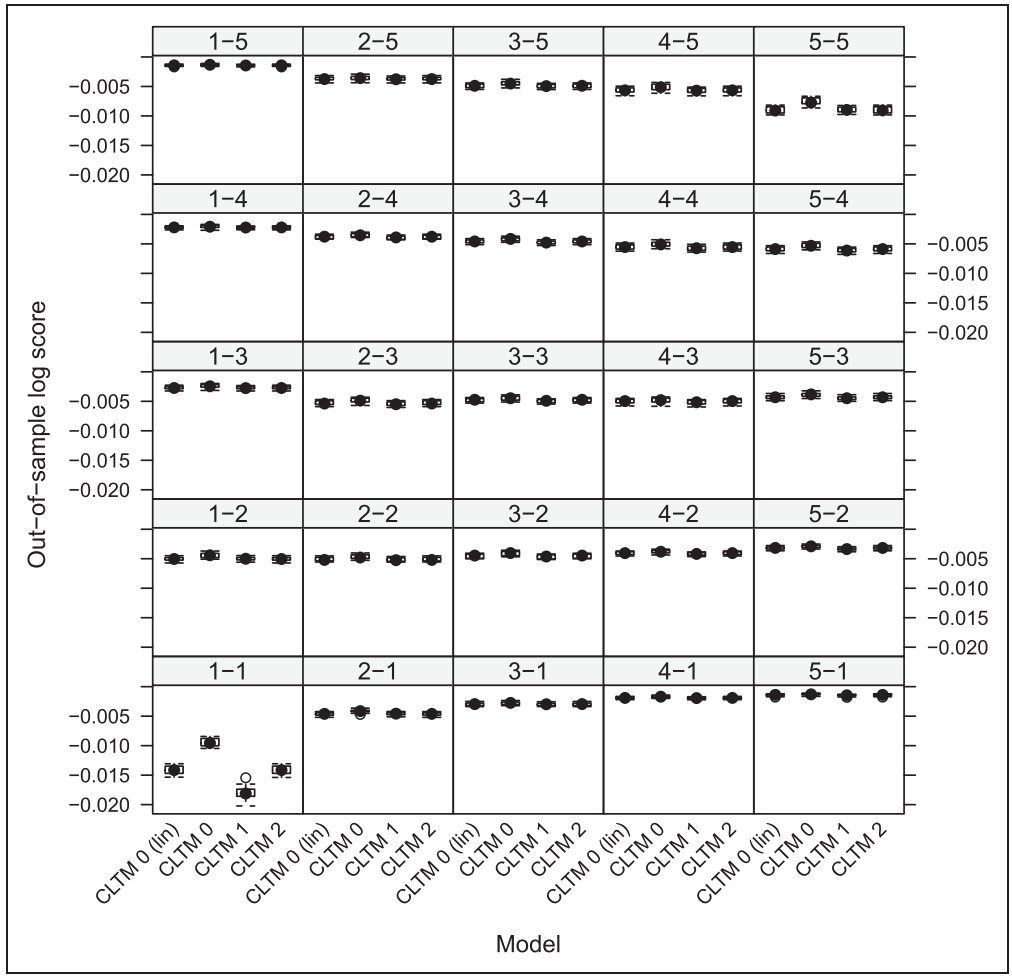

Figure 10. Boxplots of the out-of-sample log scores based on 25 evaluation data sets. The log scores were determined for the 25 categories for abdominal circumference and femur length (AC–FL) separately. Model estimation was carried out for CLTM 0 (linear), CLTM 0, CLTM 1, and CLTM 2.

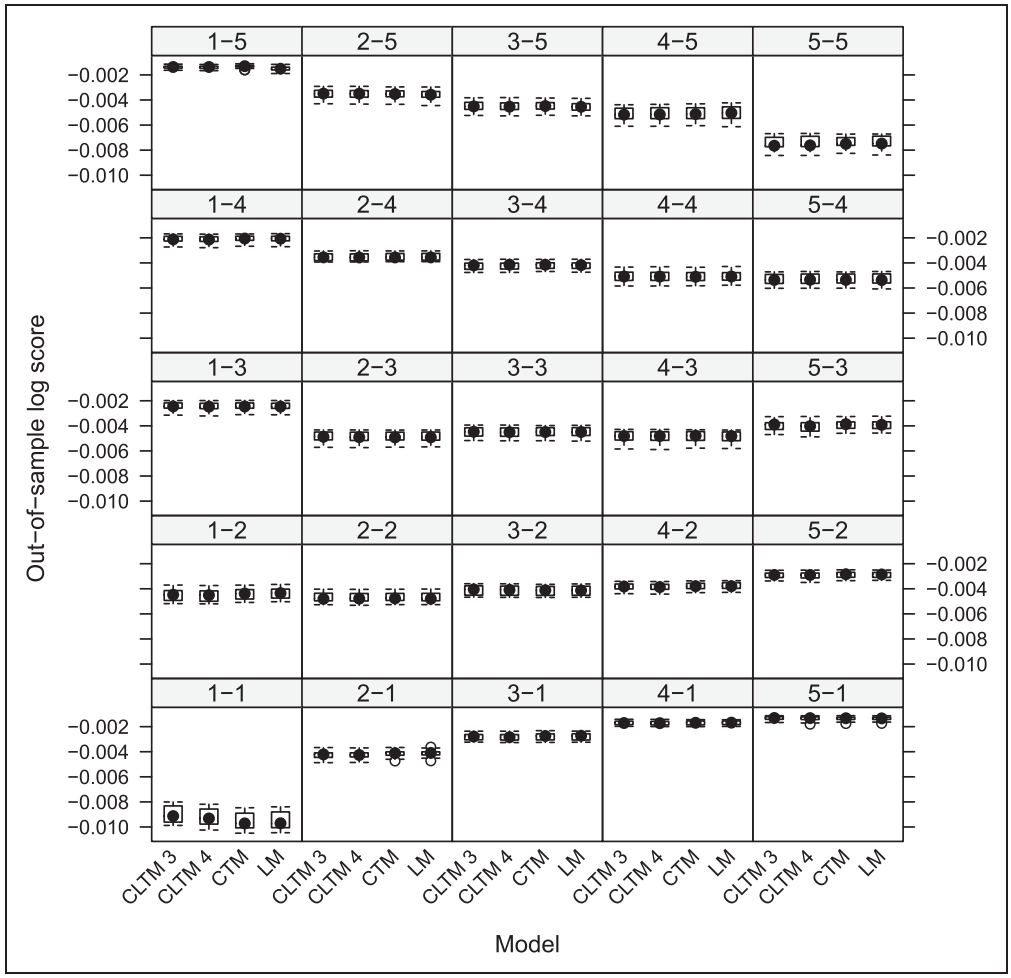

Figure 11. Boxplots of the out-of-sample log scores based on 25 evaluation data sets. The log scores were determined for the 25 categories for abdominal circumference and femur length (AC–FL) separately. Model estimation was carried out for CLTM 3, CLTM 4, CTM, and LM.

Table 3. Conditional coverage for the prediction intervals of fetuses in the 25 categories defined by abdominal circumference (AC) and femur length (FL).

AC	FL	CLTM 0 (linear)	CLTM 0	CLTM 1	CLTM 2	CLTM 3
1	1	0.826	0.826	0.587	0.826	0.783
2	1	0.784	0.784	0.838	0.784	0.784
3	1	0.905	0.857	0.905	0.905	0.857
4	1	0.933	0.800	0.933	0.933	0.800
5	1	1.000	0.818	1.000	1.000	0.909
1	2	0.944	0.833	0.944	0.944	0.833
2	2	0.952	0.881	0.976	0.952	0.881
3	2	0.884	0.721	0.907	0.884	0.744
4	2	0.903	0.871	0.903	0.903	0.903
5	2	0.885	0.885	0.962	0.885	0.885
1	3	0.957	0.870	0.957	0.957	0.870
2	3	0.973	0.892	0.973	0.973	0.892
3	3	0.923	0.897	0.974	0.923	0.897
4	3	0.947	0.842	1.000	0.947	0.842
5	3	0.857	0.714	0.857	0.857	0.714
1	4	0.800	0.800	0.867	0.800	0.733
2	4	0.938	0.875	0.938	0.938	0.875
3	4	0.875	0.825	0.925	0.875	0.825
4	4	0.841	0.841	0.886	0.841	0.909
5	4	0.857	0.833	0.881	0.857	0.833
1	5	0.800	0.600	0.900	0.800	0.600
2	5	0.880	0.840	0.880	0.880	0.840
3	5	0.970	0.879	0.970	0.970	0.879
4	5	0.935	0.913	0.957	0.935	0.913
5	5	0.817	0.817	0.850	0.817	0.867

Note: Estimation based on the conditionally linear transformation models CLTM 0 (linear) and CLTM 0–CLTM 3.

Table 4. Conditional coverage for the prediction intervals of fetuses in the 25 categories defined by abdominal circumference (AC) and femur length (FL).

AC	FL	CLTM 4	CTM	LM	LQR	AQR
1	1	0.826	0.870	0.815	0.772	0.739
2	1	0.784	0.757	0.865	0.757	0.757
3	1	0.857	0.857	1.000	0.857	0.857
4	1	0.800	0.867	0.933	0.867	0.867
5	1	0.909	0.909	0.909	0.818	0.818
1	2	0.833	0.861	0.778	0.833	0.806
2	2	0.881	0.881	0.833	0.881	0.881
3	2	0.744	0.721	0.930	0.721	0.721
4	2	0.871	0.871	0.742	0.871	0.903
5	2	0.885	0.846	0.846	0.846	0.846

(continued)

Table 4. Continued

AC	FL	CLTM 4	CTM	LM	LQR	AQR
1	3	0.870	0.913	0.739	0.870	0.957
2	3	0.892	0.892	0.919	0.892	0.892
3	3	0.897	0.897	0.769	0.872	0.897
4	3	0.842	0.842	0.895	0.842	0.842
5	3	0.714	0.743	0.857	0.714	0.714
1	4	0.800	0.800	0.800	0.800	0.800
2	4	0.875	0.906	0.750	0.906	0.906
3	4	0.825	0.825	0.750	0.825	0.825
4	4	0.864	0.841	0.864	0.841	0.795
5	4	0.833	0.857	0.690	0.786	0.810
1	5	0.600	0.600	1.000	0.600	0.600
2	5	0.840	0.840	0.880	0.840	0.760
3	5	0.879	0.848	0.727	0.788	0.818
4	5	0.913	0.913	0.717	0.935	0.870
5	5	0.850	0.783	0.767	0.733	0.750

Note: Estimation based on the regression models CLTM 4, CTM, LM, LQR, and AQR. CLTM: conditionally linear transformation models; CTM: conditional transformation model; LM: linear regression model; LQR: linear quantile regression model; AQR: additive quantile regression model.