Journal of Data Science 0 (0), 1–15
??? 2026

DOI: 10.6339/26-JDS1224
Data Science in Action

Relative Growth Modeling of Anthropometric Outcomes

John M. Williamson[1], Hung-Mo Lin[2,*], and Molin Wang[3]

[1]*McKing Consulting Corporation, Atlanta, GA 30341, University, USA*
[2]*Department of Anesthesiology and Yale Center for Analytical Sciences, Yale University, New Haven, CT 06520-8051, USA*
[3]*Department of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, and Channing Division of Network Medicine, Harvard Medical School and Brigham and Women's Hospital, Boston, MA 02115, USA*

Abstract

Traditionally z-scores specified from the WHO population growth curves have been used to describe a child's growth in relation to his age- and sex-matched population distribution. We propose a new regression approach that offers a straightforward interpretation of the relative growth in terms of the original anthropometric variable. We create a hybrid data set consisting of the observations from the study of interest and counterpart pseudo-population observations imputed from the WHO population growth curves matched to each study participant. We then fit linear and quantile regression models to the hybrid data incorporating demographic variables (usually age and biological sex) corresponding to the growth curves of demographically-similar individuals, a study versus population indicator, and its interactions with demographic variables. We further control for confounding variables from the study by adding their interactions with the study indicator variable. The interaction terms between the study indicator and the demographic variables age and biologic sex can be interpreted as relative growth parameters that depict the differences in means (or quantiles) between the study participants and their pseudo-population counterparts of the original anthropometric variables, rather than the associated z-scores. We use anthropometric growth data from a prospective birth cohort study conducted in Uganda for illustration.

Keywords *linear regression; population life tables; quantile regression*

1 Introduction

Anthropometry is a major component of nutritional status assessment in children (Borghi et al., 2006). Children's anthropometric data reflect their physical development and environment. Longitudinal growth studies with repeated measurements over time (e.g., weight or length) detail how infants and young children grow (de Onis et al., 2007). The World Health Organization (WHO) has developed an international growth standard statistical distribution that describes the growth of children ages 0 to 5 years living in environments that would not deter growth (Borghi et al., 2006; WHO Multicentre Growth Reference Study Group, 2006). The WHO Multicentre Growth Reference Study collected growth data from 8,440 healthy breastfed infants and young children from various cultural locations and ethnic groups (WHO Multicentre Growth Reference Study Group, 2006; de Onis et al., 2004). The mothers of the selected children were

---

*Corresponding author. Email: hung-mo.lin@yale.edu.

Received June 2, 2025; Accepted February 22, 2026

non-smoking and participated in healthy practices such as breastfeeding. The WHO growth charts for boys and girls are standards and describe how children should grow under ideal circumstances (de Onis et al., 1997). They are roughly normally distributed and are specified for biological sex and age (for example, a curve for body weight for girls aged 10 months) (Borghi et al., 2006). To evaluate or describe children's growth in relation to their age- and sex-matched population, physicians or researchers often use the corresponding z-score from this distribution. A negative z-score suggests that the child's development is below average and an extremely low z-score indicates that the child may be failing to reach their growth potential. For example, an infant's z-score for length below -2 (i.e., two standard deviations below the mean) is defined as "stunted" if they are shorter than others of the same age (Krebs et al., 2011).

The WHO growth curves accompanying z-scores are frequently used as an outcome that serves as a surrogate for growth development in randomized control trials from epidemiological or nutritional studies (Humphrey et al., 2019; Luby et al., 2019; Null et al., 2018), as well as for surveillance. However, the associated covariate effects are not easily interpretable in the analysis of observational growth studies. For example, z-scores are not as useful for describing adioposity change in children as the change in BMI itself (Cole et al., 2005). In addition, transforming and interpreting these effects back to the scale of the original anthropometric variable is not straightforward.

Our motivating example arises from the Uganda Birth Cohort Study (UBCS), which was a prospective birth cohort study (Bater et al., 2020) implemented within the context of the Uganda Community Connector Program (UCCP). The UCCP was a five-year cluster-randomized study of an integrated agriculture-nutrition intervention designed in part to improve the nutritional status of women and children in rural Uganda (Madzorera et al., 2021; Bater et al., 2020). The UBCS sample was drawn from women-children pairs in the UCCP intervention and control clusters (Madzorera et al., 2021). Women recruited for the birth cohort were selected in the second and third trimesters of pregnancy in 16 rural subcounties of Uganda from 2014 through 2016 (Madzorera et al., 2021). They and their infants were followed until the infants reached one year of age and anthropometric measurements were made on the infants every three months (Madzorera et al., 2021). (Namirembe et al., 2022) described infant's growth using the UBCS data with group-based trajectory modeling. Here we develop a method that enables a straightforward interpretation of a child's relative growth using the original anthropomorphic variable.

We propose to impute pseudo-population observations from the WHO growth curves and incorporate them into linear and quantile regression models to allow for comparison with ideal growth for analysis of studies of this type. These observations are not the pseudo-observations used in jackknife methodology (Miller, 1974). Relative growth parameters comparing population and study parameters (e.g., means, quantiles) of relevant subgroups can then be estimated and compared. This proposed method is based on relative survival methods, which compare the survival experience of a cohort with that of the background population (Dickman et al., 2004; Andersen et al., 1985; Esteve et al., 1990). In relative survival analysis, the mortality of the underlying population (e.g., a nation) is depicted with census life tables. Usual factors accounted for in the source population are biological sex, age, calendar time (for changing trends in mortality) and sometimes race (Dickman et al., 2004). Imputation of pseudo-population observations has been used in a quantile regression approach for the analysis of relative survival data (Williamson et al., 2023).

We introduce our approach for analyzing anthropometric growth data in Section 2. Parameter estimate standard errors and confidence intervals are estimated with the bootstrap approach. In Section 3 we present the analyses of the UBCS growth data using body weight and length measures as outcomes and analyses using the Z-score adjusted values as outcomes. We then compare these analyses with those using our proposed approach. We conclude with a short discussion on the merits of our method.

2 Methods

Let Y_{it} denote the anthropometric value (e.g. weight, length, BMI) for child i ($i = 1, \ldots, N$, where N is the sample size) at occasion t. The attained growth standards were constructed from applying the Box-Cox-power-exponential transformation to normalize the data (Flegal and Cole, 2013; Cole and Green, 1992). This “LMS” method allows for the development of smoothed curves and simultaneous calculation of z scores. The LMS parameters are the power in the Box-Cox transformation (L), the median (M), and the generalized coefficient of variation (S) (WHO Multicentre Growth Reference Study Group, 2006). We use the respective LMS parameters for each child to specify their accompanying normally distributed growth curve for the anthropometric variable of interest. We then multiply impute (see step 3 at the end of this section) an age- and sex-matched pseudo-population observation for each subject at each study visit. Multiple imputation is used to fully represent the wide range of possible values from the underlying normal distribution of each infant in the study. Furthermore, it helps in the estimation of the uncertainty (e.g., standard errors, 95% confidence intervals) of the parameter estimates. The pseudo-population value is denoted by Y^*_{it}. We plot the probability density functions of the normal distributions for the variable infant body weight of two hypothetical infants of different age and biological sex with their associated observed study values in Figure 1 for illustration.

We propose fitting a regression model to a hybrid data set that contains both study observations (e.g., those from the UBCS) and the imputed pseudo-population observations. The type of model can be flexible. For illustration, we considered linear and quantile regression models. The proposed regression model is comprised of four components that reflect: 1) the effects of the demographic characteristics of interest (e.g. age, biological sex) on the growth variable, 2) the effect comparing study versus population values, 3) the relative growth effects, 4) and potential covariates specific to the study. For the t^{th} occasion, the demographic covariates are contained in the $p \times 1$ vector $\boldsymbol{X}_{it}$ with the accompanying $p \times 1$ parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ describing the effects of demographic factors on ideal growth and the ‘relative growth’ of the study members in relation to their pseudo-population counterparts, respectively. Other covariates used to control for potential confounding are contained in the $q \times 1$ vector $\boldsymbol{W}_{it}$ with the accompanying $q \times 1$ parameter vector $\boldsymbol{\phi}$. The associated parameter vector $\boldsymbol{\phi}$ is interpreted as covariate effects specific to the study and not relevant to ideal growth. Specifically, for each occasion, the hybrid data set consists of one line for each study observation and a corresponding line for the accompanying imputed pseudo-population observation, with a binary study indicator $\delta = 1$ for the study observation and 0 for the pseudo-population observation. The regression model is thus:

$$E(Y_{it}, Y^*_{it} | \delta, \boldsymbol{X}_{it}, \boldsymbol{W}_{it}) = \alpha_0 + \boldsymbol{X}'_{it}\boldsymbol{\beta} + \mathrm{I}(\delta = 1)\left[\beta_0 + \boldsymbol{X}'_{it}\boldsymbol{\gamma} + \boldsymbol{W}'_{it}\boldsymbol{\phi}\right]. \tag{1}$$

Equivalently, the model can be re-expressed as

$$E(Y_{it} \mid \delta, \boldsymbol{X}_{it}, \boldsymbol{W}_{it}) = \alpha_0 + \boldsymbol{X}'_{it}\boldsymbol{\beta} + \beta_0 + \boldsymbol{X}'_{it}\boldsymbol{\gamma} + \boldsymbol{W}'_{it}\boldsymbol{\phi},$$

for the ith study observation at occasion t, and

$$E(Y^*_{it} | \boldsymbol{X}_{it}) = \alpha_0 + \boldsymbol{X}'_{it}\boldsymbol{\beta},$$

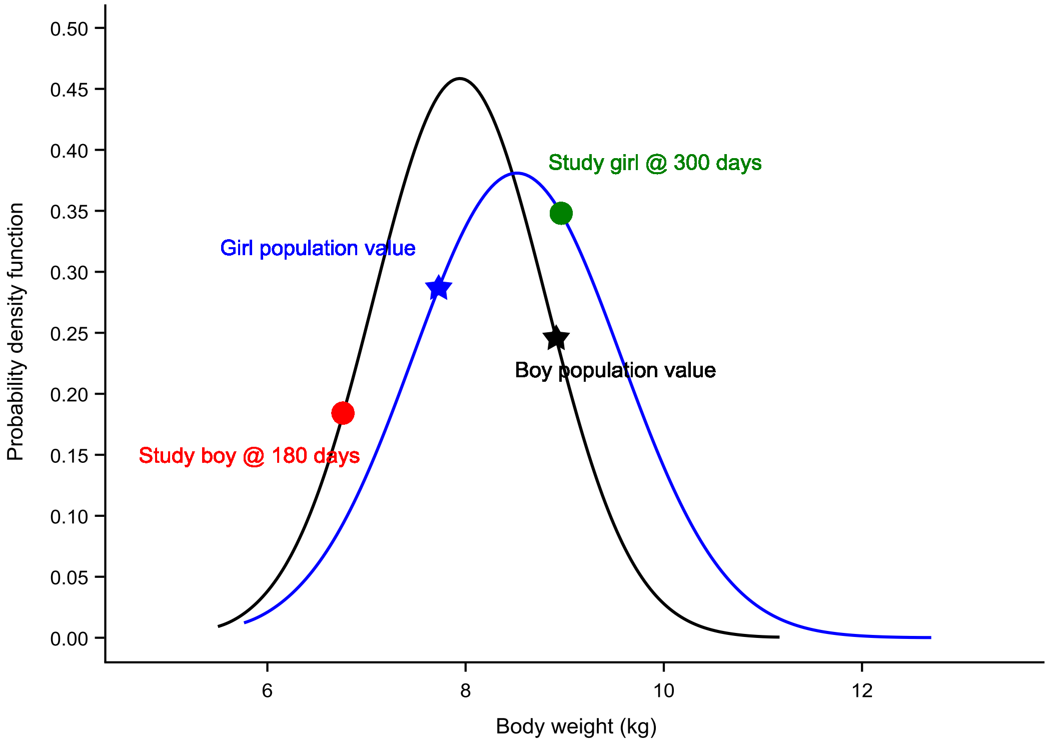

Figure 1: Randomly-imputed pseudo population values of body weight for a hypothetical 180-day old boy and a hypothetical 300-day old girl in the study.

for the ith pseudo-population observation at occasion t. For instance, if biological sex is included in the interaction term then there is a different biological sex effect in the study group than what is expected in the population (e.g., boys are not as tall compared to girls in the study given expectations from the population growth curves). We recommend that the covariates in $\boldsymbol{W}_{it}$ be centered at their respective means to allow for a more meaningful interpretation of β_0.

The anthropometric variable of interest can also be modeled with quantile regression (Koenker and Bassett, 1978; Koenker and Geling, 2001), which models the relationship between the covariates and the conditional quantiles of the response. The main advantage of quantile regression over least-squares regression is that it allows covariates to be introduced into the model differently at each quantile level of the response (Chen, 2005). It is particularly advantageous when the rate of change in the conditional quantile, expressed by the beta parameters, depends on the quantile. The quantile function for $F(y)$ (assumed to be continuous) is defined as follows: $Q(\tau) = F^{-1}(\tau)$ for $0 < \tau < 1$, where τ is referred to as the quantile level (e.g., $Q(0.1)$ denotes the 10th percentile value). Quantile regression extends the concept of an univariate quantile to a conditional quantile given one or more covariates (Koenker and Bassett, 1978).

Let the parameters $\tilde{\alpha}_0$, $\tilde{\boldsymbol{\beta}}(\tau)$, $\tilde{\beta}_0$, $\tilde{\boldsymbol{\gamma}}(\tau)$, and $\tilde{\boldsymbol{\phi}}(\tau)$ correspond to the covariate effects on the τth quantile of Y_{it} (potentially varying with τ), with a framework similar to equation (1). The linear quantile regression model for Y_{it} and Y_{it}^* is

$$Q_\tau(Y_{it}, Y_{it}^*|\delta, \boldsymbol{X}_{it}, \boldsymbol{W}_{it}) = \tilde{\alpha}_0(\tau) + \boldsymbol{X}_{it}'\tilde{\boldsymbol{\beta}}(\tau) + \mathrm{I}(\delta = 1)\left[\tilde{\beta}_0(\tau) + \boldsymbol{X}_{it}'\tilde{\boldsymbol{\gamma}}(\tau) + \boldsymbol{W}_{it}'\tilde{\boldsymbol{\phi}}(\tau)\right], \qquad (2)$$

for $\tau \in (0, 1)$. Linear programming methods such as the simplex, smoothing, and interior point algorithms are commonly used to fit quantile regression models (Chen, 2005).

We need to account for the introduction of the pseudo-population observations into the data as they are imputed based on known parameters (e.g., LMS) and should not be treated as random variables when estimating the standard errors and their associated confidence intervals for the parameters in models (1) and (2). We propose using resampling techniques such as the bootstrap (Efron, 1981) or jackknife methods to accomplish this as the default standard errors from the usual software output (e.g., SAS) will be invalid. Specifically, steps for fitting the proposed approach involve the following:

1. Calculate the corresponding age- and sex-specific LMS parameters from the WHO growth curves for each subject at each study visit.
2. Randomly generate Q bootstrap samples of study subjects with replacement.
3. Impute a normally-distributed pseudo-population value for each subject at each occasion specified from the corresponding WHO growth curves (step 1) for the Q bootstrap samples.
4. Concatenate the study observations with the pseudo-population observations resulting in a total sample size now of $2N$ for each time point t (N observations from the study and N counterpart pseudo-population observations) for each bootstrap sample. The final data set is created by concatenating these data sets across the time points.
5. Create a study indicator variable for each of the observations to indicate whether it is affiliated with the study or pseudo population.
6. Fit the proposed linear or quantile regression model to the Q bootstrap samples to obtain the corresponding Q sets of parameter estimates that are then used to calculate the parameter estimate confidence intervals.
7. Calculate summary statistics based on the Q bootstrap samples regression results (e.g., the mean, median, 2.5^{th} percentile, 97.5^{th} percentile, and standard deviation of the $\hat{\beta}_q$ estimates). The median $\hat{\beta}$ estimate across the bootstrap samples may be used as the estimate for β and the 95% confidence intervals can be calculated using the 2.5^{th} and 97.5^{th} percentiles, or the average $\hat{\beta} \pm 1.96 \times$ the standard error estimated from the parameter estimate standard deviations across the bootstrap samples.

3 Uganda Birth Cohort Study

We utilized data from the UBCS to illustrate the proposed relative growth approach. First, we used two common approaches, one that focuses on the UBCS cohort only, and the other using the sex- and age-normalized anthropometric measure (i.e., z-score) as the outcome (Bater et al., 2020). Anthropometric measurements for infants were made at birth (+3 weeks) and 3-, 6-, 9- and 12-month postpartum scheduled visits. Following (Madzorera et al., 2021), we restricted our analyses to anthropometric measurements obtained from three to 12 months and did not include birth measurements. We focus on this age span to illustrate the proposed approach as infant's growth is approximately linear, although the method is easily extended to nonlinear modeling (see Discussion). Twins and self-reported HIV-positive mothers were excluded (Madzorera et al., 2021). We analyzed 3,456 infants who had the 3-month follow-up interview and complete data for the variables considered in the model. At 3 months, 1,712 (49.5%) were boys and 95% were breastfed. The average mother's age was 27.0 (SD 6.2) years with an average height of 158.7 (SD 6.1) centimeters. Eighty-one percent of the mothers were married and 52% of the fathers had 7 or more years of education. The numbers of infants were 3,391; 2,683; and 1,809 at the 6-, 9- and 12-month follow-up visits, respectively. Growth outcomes of WAZ, LAZ and WLZ were

Table 1: Analysis of Uganda growth data using a linear regression model. The outcomes are infant's weight (kilograms) and length (centimeters).

Variable	Weight		Length	
	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI
Intercept	4.966	(4.868, 5.064)	54.066	(53.694, 54.437)
Boy vs Girl	0.387	(0.334, 0.440)	0.911	(0.691, 1.131)
Infant's Age in Months	0.312	(0.304, 0.319)	1.402	(1.374, 1.431)
North vs Southwest Region	−0.221	(−0.278, −0.164)	−0.310	(−0.493, −0.127)
Mother's Age[a] 20–29 Years	0.033	(−0.064, 0.131)	0.087	(−0.255, 0.429)
Mother's Age[a] 30+ Years	−0.048	(−0.151, 0.056)	0.145	(−0.213, 0.503)
Mother's Height in Centimeters	0.023	(0.019, 0.028)	0.091	(0.077, 0.106)
Paternal Education 7+ Years	0.056	(0.000, 0.111)	0.434	(0.257, 0.611)
Married vs not	−0.009	(−0.082, 0.063)	0.063	(−0.174, 0.299)
Currently Breastfeeding	0.116	(0.002, 0.230)	1.696	(1.321, 2.072)
Boy by Infant's Age in Months	0.011	(−0.000, 0.022)	0.014	(−0.027, 0.055)

[a] Mother's age in years versus the reference category of < 20.

calculated using the WHO growth reference standards (WHO Multicentre Growth Reference Study Group, 2006).

We first fit a linear regression model to infants' body weight (in kilograms) and length (in centimeters) not taking into account the population growth charts for comparison. Weight and length measures were made in triplicate, and mean measures were used in this analysis (Madzorera et al., 2021). The UCBS was composed of eight subcounties randomly selected from the UCCP intervention arm and matched to eight non-intervention subcounties (Madzorera et al., 2021). We adjusted for clustering at the infant level to control for repeated measurements on the same individual. Covariates included in the model were region (North versus Southwest), biological sex (boys $= 1$ versus girls $= 0$), infant's age in months, mother's age (three categories), mother's height in centimeters, paternal education (seven or more years versus six or less), married versus not, and currently breastfeeding. To allow a more meaningful interpretation of the model intercept, we centered the covariates (excluding age and sex) around their respective averages. The intercepts now represent the average weight and length of the newborn girls in the study.

The results of the analyses (parameter estimates and 95% confidence intervals) are presented in Table 1. We focus discussion on the body weight outcome. Infants from the Southwest region and those currently being breastfed were significantly heavier than their counterparts, as were infants born to taller mothers and more educated fathers. Marital status and mother's height were not statistically associated with infant's weight. As expected, boys were heavier at birth and there is some evidence that they gained weight quicker than females ($\approx$ an additional 11 grams per month with a 95% confidence interval $(-0.000, 0.022)$). Results for the infant length model were similar, except that there was no evidence that the body length growth velocity differs between boys and girls.

We then fit a linear regression model to infants' body weight and length using their accompanying z-scores from the WHO growth reference curves as the response, with the same covariates from the above model. The results of the analyses (parameter estimates and 95%

Table 2: Analysis of Uganda growth data using usual linear regression model with z-score adjusted infant's weight (kilograms) and length (centimeters) as outcomes.

Variable	Weight		Length	
	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI
Intercept	−0.075	(−0.194, 0.045)	−0.750	(−0.918, −0.581)
Boy vs Girl	−0.217	(−0.310, −0.123)	−0.407	(0.554, −0.260)
Infant's Age in Months	−0.043	(−0.051, −0.035)	−0.064	(−0.077, −0.052)
North vs Southwest Region	−0.265	(−0.331, −0.199)	−0.147	(−0.229, −0.065)
Mother's Age[a] 20–29 Years	0.051	(−0.066, 0.168)	0.052	(−0.102, 0.206)
Mother's Age[a] 30+ Years	−0.049	(−0.174, −0.075)	0.084	(−0.077, 0.245)
Mother's Height in Centimeters	0.027	(0.022, 0.032)	0.041	(0.034, 0.048)
Paternal Education 7+ Years	0.081	(0.017, 0.144)	0.200	(0.121, 0.279)
Married vs Not	0.002	(−0.083, 0.086)	0.030	(−0.076, 0.136)
Currently Breastfeeding	0.113	(−0.024, 0.251)	0.724	(0.557, 0.892)
Boy by Infant's Age in Months	−0.002	−(0.014, 0.010)	−0.005	(−0.023, 0.013)

[a] Mother's age in years versus the reference category of < 20.

confidence intervals) are presented in Table 2. Parental demographic effects on their child's growth had similar qualitative conclusions as those in Table 1; however, mother's age became statistically significant and breastfeeding became non-significant. The parameter estimates for boy versus girl, infant's age in months, and boy by age interaction term were all negative for both the weight and length outcomes. The biologic sex and age parameter estimates should be interpreted carefully because the z-scores were normalized based on the corresponding sex and age of the subject. For illustration, we focus on body weight and the comparison of boys and girls.

The model intercept represents the z-score for newborn girls predicted at the means of the parental demographic variables, which was 0.075 standard deviations below the average newborn girl under ideal growth. The predicted z-score for the study boys at birth was 0.292 ($= 0.075 + 0.217$) standard deviations below the average new born boy under ideal growth. Hence, the coefficient of the boy versus girl parameter, although not a direct estimate of their relative growth, is indicative of the difference in the boys and girls' growth in relation to their respective ideal growth. The negative coefficient of the boy versus girl parameter suggests that the boys were more underweight than expected compared to the girls. Similarly, the statistically significant negative parameter estimate for age in months and statistically non-significant boy by age interaction suggest that the study subjects gained weight slower than expected under ideal growth, regardless of biologic sex.

Although the z-score analysis was able to provide some insight into the relative growth of the study subjects, it can be difficult to interpret the corresponding parameter estimates in relation to the z-scores. Ideally, relative growth would be expressed in terms of the same unit as that used for the anthropometric measure, allowing easy and straightforward interpretation. For example, a statement such as the relative weight gain for boys versus girls was 30 grams less per month during the 1st year of age is easy to comprehend because it directly uses the original scale to depict the difference in growth between the observed and ideal values for boys versus girls.

Table 3: Analysis of Uganda data using the proposed relative growth linear regression model. The outcomes are infant's weight (kilograms) and length (centimeters). Median β estimates and percentile-based 95% confidence intervals were calculated using 1000 replicate bootstrap data sets.

Model	Weight		Length	
	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI
Main Effects				
Intercept	4.951	(4.903, 4.999)	55.392	(55.261, 55.533)
Study vs Population	0.025	(−0.076, 0.128)	−1.318	(−1.692,−0.954)
Boy vs Girl	0.509	(0.451, 0.565)	1.671	(1.480, 1.838)
Infant's Age in Months	0.361	(0.354, 0.368)	1.609	(1.590, 1.627)
Boy * Infant's Age in Months	0.014	(0.006, 0.022)	0.015	(−0.010, 0.040)
Study * North vs Southwest Region	−0.222	(−0.276, −0.165)	−0.313	(−0.504, −0.114)
Study * Mother's Age[a] 20–29 Years	0.036	(−0.067, 0.128)	0.081	(−0.271, 0.434)
Study * Mother's Age[a] 30+ Years	−0.045	(−0.158, 0.053)	0.141	(−0.236, 0.515)
Study * Mother's Height in Centimeters	0.023	(0.019, 0.028)	0.093	(0.078, 0.106)
Study * Paternal Education 7+ Years	0.055	(−0.006, 0.106)	0.430	(0.241, 0.610)
Study * Married vs Not	−0.010	(−0.081, 0.061)	0.078	(−0.167, 0.311)
Study * Currently Breastfeeding	0.115	(0.007, 0.221)	1.688	(1.309, 2.080)
Relative Growth[b]				
Study * Boy	−0.176	(−0.235, −0.113)	−0.803	(−0.987, −0.615)
Study * Infant's Age in Months	−0.051	(−0.059, −0.043)	−0.207	(−0.232, −0.182)

[a] Mother's age in years versus the reference category of < 20.

[b] Relative growth parameters for study versus population comparison.

The proposed relative growth approach aims to complement traditional z-score methods. The results of the analyses (median parameter estimates and nonparametric 95% bootstrap confidence intervals based on the 2.5^{th} and 97.5^{th} percentiles of 1000 replicate data sets) are presented in Table 3. Infants from the Southwest region, with taller mothers, and currently being breastfed were significantly heavier than their counterparts. We estimated two relative growth parameters: biological sex and infant's age in months. Figure 2 depicts the predicted body weight gain (at the mean levels of the parental covariates) from 3 to 12 months after birth for boys and girls in the study and their counterparts under ideal growth.

The relative growth comparison of infant boys to girls (interaction parameter) resulted in a value of -0.176 kilograms with an accompanying 95% bootstrap confidence interval of $(-0.235, -0.113)$ kilograms, indicating significantly less relative weight for infant boys compared to infant girls. For illustration, we present how the body weight relative growth comparison between boys and girls at 3 months of age was derived. From the fitted model presented in Table 3, the predicted body weight for infant boys at 3 months for all other covariates at their mean levels was 4.951 (intercept) +0.025 (study indicator) + 0.509 (boy vs girl) − 0.176 (study by boy) + 0.361 (infant's age in months) + 0.014 (boy by infant's age in months) − 0.051 (study by infant's age in months)) × 3 = 6.281 kilograms, and for infant girls at 3 months was $4.951+0.025+(0.361-0.051)\times 3 = 5.906$ kilograms. The population counterpart for infant boys (girls) was $4.951+0.509+(0.361+0.014)\times 3 = 6.585$ (6.034) kilograms. Thus, the relative body

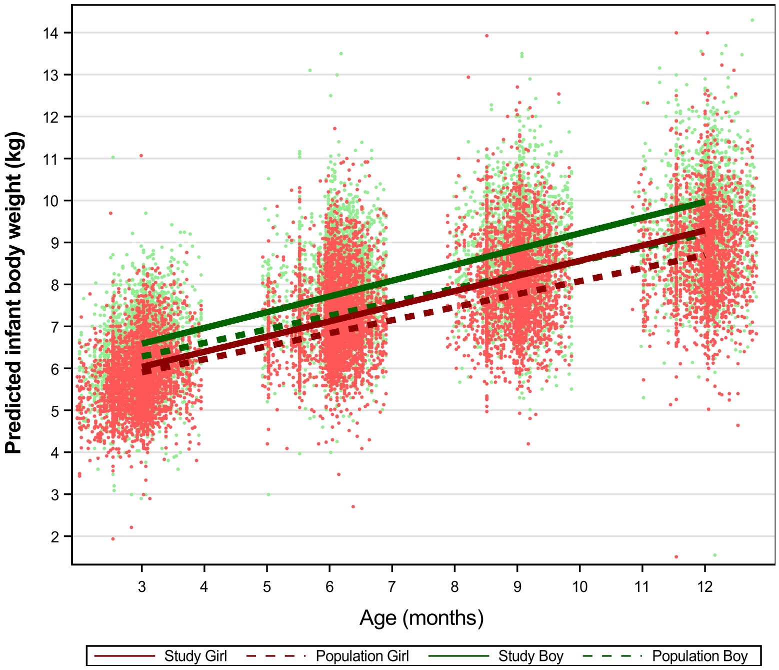

Figure 2: Predicted infant body weight during the first year for boys and girls (study and under ideal growth) predicted from the proposed relative growth linear regression model, with the accompanying observed values.

weight between infant boys and girls at 3 months is $(6.281 - 6.585) - (5.906 - 6.034) = -0.176$ kilograms. For monthly change in infant body weight, the rate for boys was $0.361 + 0.014 - 0.051 = 0.324$ kg/month, and $0.361 - 0.051 = 0.310$ kg/month for girls. The rates for boys and girls under ideal growth was 0.375 (0.361) kg/month. The model we fit assumed that the relative growth rates were the same for infant boys and girls. To investigate whether there was a different relative growth rate between infant boys and girls, we included a 3-way interaction between the variables study, sex, and age in months, which was not statistically significant ($\hat{\beta} = -0.007$ with 95% CI $(-0.022, 0.010)$). The relative growth for the outcome infant body length can be interpreted in a similar fashion.

In Table 3 the coefficient for "Study vs. Population" represents the average difference in weight or length between the study cohort and the WHO reference population, which differed significantly for length (-1.32 cm) but not weight (0.025 kg). The terms for age, sex, and their interaction describe the WHO growth curves for children aged 3 months to 1 year, separately by sex. Interpretation of the main effects of the other demographic variables (other than age and sex) on growth is captured through their respective interaction terms with the study versus population indicator. For example, the estimated coefficient for the interaction between "Study vs. Population" and "North vs. Southwest Region" is -0.222. This indicates that, with all other variables constant, the infants in the study in the North region had, on average, 0.222 kilograms lower weight (95% CI: 0.165, 0.276) than the study infants in the Southwest region. The average weight of infants in the two regions are assumed equal in the population because the growth curves are not specified separately by region.

We also fit four quantile regression models (quantiles $\tau = 0.025$, $\tau = 0.05$, $\tau = 0.10$, and

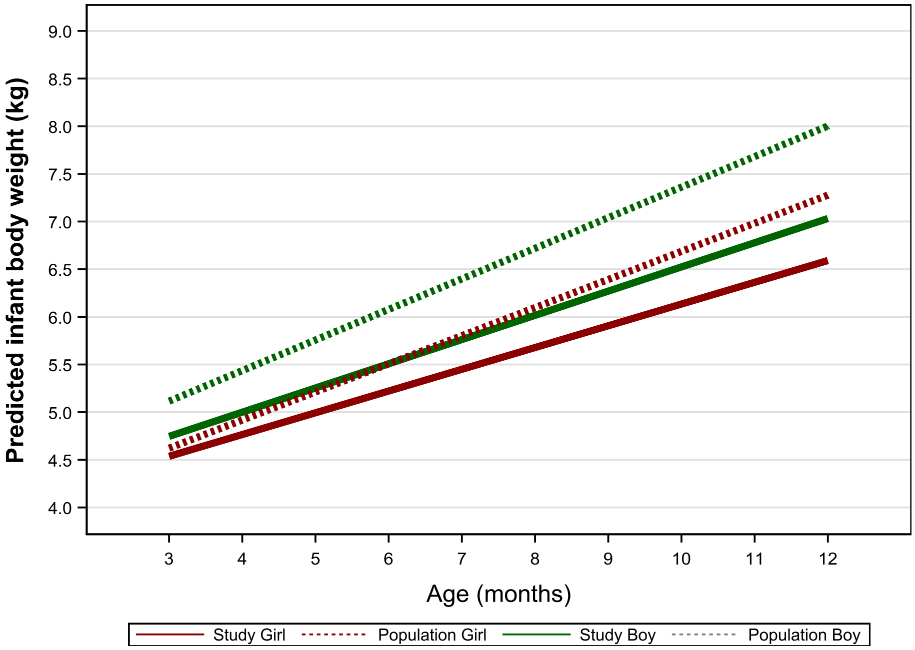

Figure 3: Predicted infant body weight during the first year for boys and girls (study and under ideal growth) predicted from the proposed relative growth quantile regression model for $\tau = 0.025$.

$\tau = 0.20$) to the outcome infant's body weight. We focus on the these quantiles to mimic settings where wasting is of concern. The results of the analyses (median parameter estimates and 95% bootstrap confidence intervals based on 1000 replicate data sets) are presented in Figure 3 (for $\tau = 0.025$ only) and Table 4 (for the other three τ's). For the $\tau = 0.05$ quantile regression model, the relative growth parameter estimate for biological sex (interaction term) is -0.289 kilograms with an accompanying 95% bootstrap confidence interval of $(-0.396, -0.172)$ indicating that boys were more likely to be severely underweight than females compared to what was expected from the WHO growth curves. Under ideal growth, the 5th percentile body weight for boys was 0.589 kilograms heavier than the corresponding value for girls. The respective value observed in the study was 0.300 $(0.589 - 0.289)$ kilograms, after controlling for the other covariates in Table 3.

Under ideal growth, the 5th percentile of monthly weight gain was 0.317 kilograms during the first year of life. The relative growth weight gain in months (interaction term) was 0.061 kilograms resulting from the monthly weight gain of 0.271 kilograms in the study cohort. Estimates of the relative growth parameters from the other three quantile regression models ($\tau = 0.025$, $\tau = 0.10$, and $\tau = 0.20$) were very similar. Results for a median regression model (not presented) were very similar to those from the linear regression model as infant's body weight appeared symmetric at each follow-up visit. Figure 3 depicts predicted infant body weight for wasted infants in the study compared with ideal growth during their first year ($\tau = 0.025$ quantile regression model). All analyses were conducted via SAS IML (SAS/IML Software, 2012).

We assessed the quantile regression models as follows. We first calculated residuals, the difference between the observed and predicted values based on the parameter estimates in Table 4, for each of the quantile regression models. The model goodness of fit was evalu-

Table 4: Analysis of Uganda relative growth data (infant's weight in kilograms) using quantile regression models (quantiles $\tau = 0.05$, $\tau = 0.10$ and $\tau = 0.20$). Median β estimates and percentile-based 95% confidence intervals were calculated using 1000 replicate bootstrap data sets.

Model	$\tau = 0.05$		$\tau = 0.10$		$\tau = 0.20$	
	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI
Main Effects						
Intercept	3.897	(3.821, 3.975)	4.090	(4.026, 4.156)	4.332	(4.277, 4.393)
Study vs Population	0.173	(0.026, 0.322)	0.155	(0.022, 0.272)	0.110	(−0.003, 0.251)
Boy vs Girl	0.443	(0.343, 0.546)	0.472	(0.388, 0.554)	0.509	(0.439, 0.577)
Infant's Age in Months	0.305	(0.295, 0.316)	0.318	(0.309, 0.327)	0.335	(0.326, 0.343)
Boy * Infant's Age in Months	0.023	(0.009, 0.036)	0.020	(0.009, 0.032)	0.016	(0.007, 0.026)
Study * North vs Southwest Region	−0.127	(−0.241, −0.016)	−0.121	(−0.189, −0.051)	−0.122	(−0.191, −0.059)
Study * Mother's Age[a] 20 − 29 Years	−0.012	(−0.146, 0.119)	−0.013	(−0.120, 0.104)	0.048	(−0.076, 0.149)
Study * Mother's Age[a] 30+ Years	−0.156	(−0.327, 0.002)	−0.110	(−0.238, 0.018)	−0.041	(−0.186, 0.066)
Study * Mother's Height in Centimeters	0.015	(0.007, 0.023)	0.016	(0.010, 0.021)	0.018	(0.013, 0.023)
Study * Paternal Education 7+ Years	0.141	(0.033, 0.240)	0.079	(0.009, 0.146)	0.072	(0.012, 0.130)
Study * Married vs Not	−0.005	(−0.126, 0.128)	0.018	(−0.072, 0.090)	0.031	(−0.058, 0.111)
Study * Currently Breastfeeding	0.391	(0.130, 0.655)	0.222	(0.068, 0.486)	0.153	(−0.006, 0.314)
Relative Growth[b]						
Study * Boy	−0.285	(−0.394, −0.165)	−0.265	(−0.343, −0.182)	−0.270	(−0.341, −0.191)
Study * Infant's Age in Months	−0.062	(−0.077, −0.049)	−0.059	(−0.071, −0.047)	−0.059	(−0.068, −0.049)

[a] Mother's age in years versus the reference category of < 20.
[b] Relative growth parameters for study versus population comparison.

ated using the pseudo R-squared statistic (Koenker and Machado, 1999), resulting in values of $R_1^2 = 0.43, 0.49, 0.52, 0.54$ for the $\tau = 0.025, 0.05, 0.10, 0.20$ models.

Furthermore, the percentage of negative residuals were 2.1%, 4.3%, 8.9%, and 18.4%, which were close to the targeted values of $\tau = 0.025, 0.05, 0.10, 0.20$ models indicating reasonable model fit (Koenker and Bassett, 1978). For illustration, we assessed the linearity assumption between the predictor infant's age in months and the conditional quantiles of infant's weight. We plotted x_{it} against $r_{it} + \hat{\beta}_{age} x_{it}$ (a component-plus-residual plot). See Figure 4 for the $\tau = 0.05$ and 0.1 plots that support the linearity assumption. If the linearity assumption does not hold, a spline model may be utilized to depict the relationship. For brevity, we do not present model fit of the linear regression models. See (Kutner et al., 2005) for assessment of linear model fit.

4 Discussion

The proposed relative growth approach models the child's actual anthropometric value instead of the accompanying z-value derived from the WHO population growth curves. The relative growth is expressed on the original scale of the anthropometric variable. A linear regression model may be used if the average effect of the covariates is of interest. Alternatively, quantile regression may be useful in applications when extremes are important (e.g., underweight infants in resource-poor settings). We utilized a linear infant's age in months variable in our modeling due to the short age span (3–12 months) considered in the UBCS example, although nonlinear (e.g., quadratic, cubic) terms or fractional polynomials in equation (1) may be alternatives if follow-up duration is longer (Royston and Altman, 1994; Royston et al., 1999).

We propose an imputation scheme for the anthropometric variable under ideal growth rather

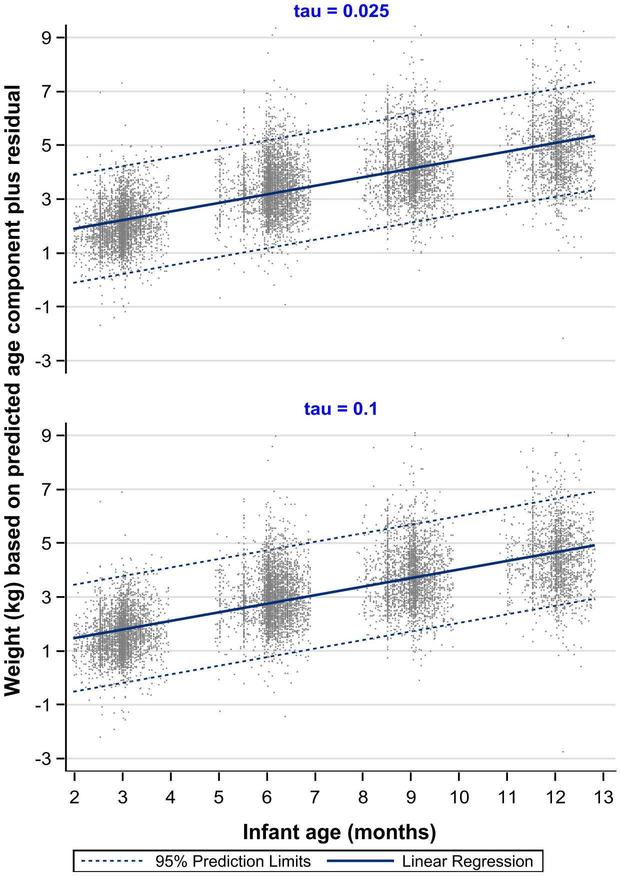

Figure 4: Age component plus residual plots for $\tau = 0.05$ and 0.10. Each dot represents the sum of the predicted age effect and the residual. The blue line shows the fitted regression line of these values against infant age.

than specifying a summary value for it (e.g., the mean value for linear regression or the actual quantile for quantile regression) to depict the random variation in the normally-distributed population growth curves. Consequently, the bootstrap method is used for parameter estimate confidence interval estimation. For comparison, we repeated the analyses in Table 4 replacing the imputed pseudo-populations values with the mean value of the respective individual's WHO growth curve instead of multiply imputing from the curve. The resulting parameter estimates were similar but not exact, and the parameter estimate standard errors were narrower.

Our approach is to mimic a matched-pair framework by randomly generating response values for the population according to the WHO growth charts. We do not use univariate tests for matched data (e.g., paired t-test) here because they treat the population values as random variables. In addition, we are interested in modeling the relative growth over time (e.g., sex comparison) while adjusting for confounding in the study. Incorporating confounders into the proposed model may be necessary when researching the relative growth of a subpopulation where study individuals may have comorbidities that should be taken into account. An alternative approach for analyzing such relative growth data is to directly model the difference in the study and pseudo-population observations directly with linear regression. However, one would not be able to estimate parameter effects on the study observations directly ($\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}$ in equation (1)).

Finally, the proposed approach has the potential for application in diverse contexts where a comparison with sex- and age-matched populations is relevant. For example, this approach may be extended to cognitive outcome research,, where neurocognitive test scores for normal individuals are categorized by age and sex, similar to life tables from census data. Specifically, for distinctive subgroups (e.g., diabetic elderly individuals or patients undergoing deep hypothermic circulatory arrest anesthesia) who could undergo more pronounced cognitive decline compared to demographically similar individuals in the general population, the proposed method can be employed to investigate relative cognitive change by sex and age within these subgroups.

Supplementary Material

The data set, WHO growth curves, and the SAS IML programs can be downloaded from https://github.com/Hungmolin/Relative-growth-modeling-of-anthropometric-outcomes.git. A complete list of the files is described in README.txt.

Acknowledgments

We gratefully thank Dr. Lynne Ausman of the School of Nutrition Science and Policy of Tufts University and Dr. Grace Namirembe of the Department of Economics of the Uganda Martyrs University for providing the illustrative data set. This paper was prepared by John Williamson in his personal capacity. The opinions expressed in this article are the author's own and do not reflect McKing Consulting Corporation.

Funding

HML was partially supported by the National Health Institute under CTSA Grant UL1 TR001863.

References

Andersen P, Borch-Johnsen K, Deckert T, Green A, Hougaard P, ..., Kreiner S (1985). A Cox regression model for the relative mortality and its application to diabetes mellitus survival data. *Biometrics*, 41: 921–932. https://doi.org/10.2307/2530964

Bater J, Lauer JM, Ghosh S, Webb P, Agaba E, ..., Duggan B (2020). Predictors of low birth weight and preterm birth in rural Uganda: Findings from a birth cohort study. *PLoS ONE*, 15: e0235626. https://doi.org/10.1371/journal.pone.0235626

Borghi E, de Onis M, Garza C, Van den Broeck J, Frongillo E, ..., Martines J (2006). Construction of the world health organization child growth standards: Selection of methods for attained growth curves. *Statistics in Medicine*, 25: 247–265. https://doi.org/10.1002/sim.2227

Chen C (2005). An introduction to quantile regression and the quantreg procedure. SAS Global Forum, Paper 213–230.

Cole T, Faith M, Pietrobelli A, Heo M (2005). What is the best measure of adiposity change in growing children: Bmi, bmi %, bmi z-score or bmi centile? *European Journal of Clinical Nutrition*, 59: 419–425. https://doi.org/10.1038/sj.ejcn.1602090

Cole T, Green P (1992). Smoothing reference centile curves: The lms method and penalized likelihood. *Statistics in Medicine*, 11: 1305–1319. https://doi.org/10.1002/sim.4780111005

De Onis M, Garza C, Habicht J (1997). Time for a new growth reference. *Pediatrics*, 100: E8.

De Onis M, Garza C, Victora C, Bhan M, Norum K, eds (2004). Who multicentre growth reference study (mgrs): Rationale, planning and implementation. *Food and Nutrition Bulletin*, 25: S1–S89.

De Onis M, Onyango A, Borghi E, Siyam A, Nishida C, Siekmann J (2007). Development of a who growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization*, 85: 660–667. https://doi.org/10.2471/BLT.07.043497

Dickman P, Sloggett A, Hills M, Hakulinen T (2004). Regression models for relative survival. *Statistics in Medicine*, 23: 51–64. https://doi.org/10.1002/sim.1597

Efron B (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68: 589–599. https://doi.org/10.1093/biomet/68.3.589

Esteve J, Benhamou E, Croasdale M, Raymond L (1990). Relative survival and the estimation of net survival: Elements for further discussion. *Statistics in Medicine*, 9: 529–538. https://doi.org/10.1002/sim.4780090506

Flegal K, Cole T (2013). Construction of lms parameters for the centers for disease control and prevention 2000 growth charts. *Hyattsville, MD: National Health Statistics Reports*, 63: 51–64.

Humphrey J, Mbuya M, Ntozini R, Moulton L, Stolzfus R, ..., Prendergast A (2019). Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: A cluster randomized trial. *The Lancet Global Health*, 7: e132–147. https://doi.org/10.1016/S2214-109X(18)30374-7

Koenker R, Bassett G (1978). Regression quantiles. *Econometrica*, 46: 33–50. https://doi.org/10.2307/1913643

Koenker R, Geling O (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of the American Statistical Association*, 96: 458–468. https://doi.org/10.1198/016214501753168172

Koenker R, Machado J (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94: 1295–1310. https://doi.org/10.1080/01621459.1999.10473882

Krebs N, Mazariegos M, Tshefu A, Bose C, Sami N, ..., Hambidge K (2011). Meat consumption is associated with less stunting among toddlers in fours diverse low-income settings. *Food and Nutrition Bulletin*, 32: 185–191. https://doi.org/10.1177/156482651103200301

Kutner M, Nachtsheim C, Neter J, Li W (2005). *Applied Linear Statistical Models*. McGraw-Hill Education, New York, NY. Chapter 3.

Luby S, Rahman M, Arnold B, Unicomb L, Ashraf S, ..., Colford J Jr (2019). Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: A cluster randomised controlled trial. *The Lancet Global Health*, 6: e302–315. https://doi.org/10.1016/S2214-109X(17)30490-4

Madzorera I, Ghosh S, Wang M, Fawzi W, Isanaka S, ..., Duggan C (2021). Prenatal dietary diversity may influence underweight in infants in a Ugandan birth-cohort. *Maternal & Child Nutrition*, 17: e13127. https://doi.org/10.1111/mcn.13127

Miller R (1974). The jackknife – a review. *Biometrika*, 61: 1–15.

Namirembe G, Ghosh S, Ausman L, Shrestha R, Zaharia S, ..., Webb P (2022). Child stunting starts in utero: Growth trajectories and determinants in Uganda infants. *Maternal & Child Nutrition*, 18: e13359. https://doi.org/10.1111/mcn.13359

Null C, Stewart C, Pickering A, Dentz H, Arnol B, ..., Colford J Jr (2018). Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: A cluster-randomised controlled trial. *The Lancet Global Health*, 6: e316–329. https://doi.org/10.1016/S2214-109X(18)30005-6

Royston P, Altman D (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Applied Statistics*, 43: 429–467. https://doi.org/10.2307/2986270

Royston P, Ambler G, Sauerbrei W (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology*, 28: 964–974. https://doi.org/10.1093/ije/28.5.964

SAS/IML Software (2012). *Changes and Enhancements Through Release 12.1*. SAS Institute, Inc., Cary, NC.

WHO Multicentre Growth Reference Study Group (2006). Who child growth standards based on length/height, weight and age. *Acta Paediatric Supplement*, 450: 76–85.

Williamson J, Lin HM, Lyles R (2023). A censored quantile regression approach for relative survival analysis: Relative survival quantile regression. *Biometrical Journal*, 65: e2200127. https://doi.org/10.1002/bimj.202200127