1.Study Title: Linear Mixed Effects Quantile Regression (LMQR) Model

2. Working Group and Investigators :

This analysis will be conducted under Epidemiology/Biostatistics Working Group with the following investigators:

Mi-Ok Kim	miok.kim@cchmc.org	513-636-1895
Mai Zhou	mai@ms.uky.edu	859-257-6912
Wendy Leisenring	wleisenr@fhcrc.org	206-667-4374
Yutaka Yasui	yutaka.yasui@ualberta.ca	780-492-4220
Kevin Oeffinger	oeffingk@mskcc.org	212-639-8469

3. Background and Rationale

Linear mixed effects model is arguably the most popular analytic tool when regression is concerned about a continuous response and a group of covariates in dependent data. The dominant paradigm in the mixed effects model literature has been a Gaussian structure in which the errors are homogenous and the covariates exert a pure location shift effect on the response variable, uniformly across the conditional distribution. Under the uniform covariate effects on the conditional distribution, the conditional mean structure is of the primary interest. In some applications, however, we need to consider a broader class of covariate effects or the covariate effects on other statistical quantities such as the conditional 90^{th} percentile of the response. We propose a linear mixed effects quantile regression (LMQR) model in such applications.

We illustrate the assumptions and limitations of the Gaussian structure using obesity study in survivors of acute lymphoblastic leukemia (ALL). Obesity has been identified as a potential late effect of cancer therapy for childhood cancer in survivors of acute lymphoblastic leukemia (ALL) and the role of cranial radiation therapy (CRT) has been questioned as a potential risk factor. Both problems have been investigated either by comparing means of age-, race- and/or gender-adjusted body mass index (BMI) or by comparing proportions of obese or over-weight. Such analyses tend to be snap shots and are reasonable summaries of the potential late effect only under rather strict assumptions of normality and/or homogeneity. If the variance of BMI is affect by the late effect of cancer therapy differentially by CRT dosage or the BMI distribution is skewed, the least squares analysis under the Gaussian paradigm can be biased. An extreme example where these analyses are simplistic or can be even misleading is when cancer therapy has a polarizing late effect, making the ALL survivors either obese or underweighted. The least squares mean analysis would not find any significant difference and analysis on the proportions would only find a half of the truth. A more practical example where the least squares analysis is limiting is as follows: the late effect of cancer therapy varies across the BMI distribution, for example, making BMI of the ALL survivors bigger increasingly at higher percentiles. The least squares analysis will ignore such increasing magnitude of

the late cancer therapy effect at the higher conditional percentiles and provide a one number summary. Quantile regression enables one to investigate the late cancer therapy effect at various conditional percentiles directly, without assuming the effect to be same as on the mean, and complements the least square analysis by providing a more complete picture of the late cancer therapy effect across the conditional distribution.

Proposed by Koenker and Bassett (1978), quantile regression extends the idea of ordering and quantiles from a univariate analysis to regression analysis, using an asymmetric L1 type loss function, and provides a framework for modeling statistical quantities of interest other than the conditional mean. The methodology is well developed for linear models in independent data. Koenker (2004) extended the methodology to a longitudinal data and proposed a quantile regression model with a random subject effect. In this study we propose a more comprehensive approach to the methodology in dependent data that can accommodate general random effects, permitting random coefficients quantile regression model analysis.

In this methodological study of LMQR model we propose to use the Childhood Cancer Survivor Study (CCSS) as a motivating example and to do ancillary analyses of its data to illustrate its applications. Specifically, we would like to use the baseline data that was used in Oeffinger KC. (2003). Oeffinger KC. (2003) reported that after adjusting for age at diagnosis and cranial radiation dose, ALL survivors did not have a significantly higher odds for being overweight compared with their nearest-age living siblings, but their odds for being obese were significantly increased for the group received cranial radiation dose $\geq$ 20 Gy. This suggests that the late effect of cancer therapy with cranial radiation dose $\geq$ 20 Gy affects the BMI differentially across its distribution. An ancillary analysis of this data with LMQR model can complement Oeffinger KC. (2003)'s analysis by supplementing its findings with information about the late effect of cancer therapy on other parts of the conditional distribution.

We also propose to use longitudinal BMI measurements of the CCSS as another motivating example and application of the methodology. LMQR model analysis can investigate whether change in BMI over time is differentially affected by cancer therapy at higher percentiles than at the median or lower percentiles. This is particularly interesting if the analysis of the baseline data finds an increasing magnitude of the late effect on BMI at higher percentiles. Whether such increasingly pronounced late effect at higher percentiles will be also manifested over time can be answered by LMQR analysis. The height of ALL survivors is an additional independent interest as impaired linear growth is a well-recognized complication in this population. However, many patients achieve a final height between the 5th and the 95th percentile and the true incidence of linear growth impairment has not been investigated in this direction. We propose to use this as an additional motivating example.

4. Specific Aims

(1) to develop a comprehensive approach to the quantile regression methodology in dependent data that can accommodate a general class of random effects other than subject effects.

(2) to apply the methodology to the Childhood Cancer Survivor Study (CCSS) to investigate

(a) whether BMI in adult survivors of childhood ALL is differentially affected by cancer therapy across its distribution, when compared with BMI distribution of their nearest-aged siblings.

(b) whether change in BMI over time in adult survivors of childhood ALL is differentially affected by cancer therapy across its distribution

(3) apply the methodology to the Childhood Cancer Survivor Study (CCSS) to investigate similar questions as (a) and (b) with respect to height.

5. Analysis Framework

For SA1 and SA2 we use age-, race- and gender-adjusted BMI as response variable and regress the so-called BMI z score on covariates such as age at cancer diagnosis, types of cancer treatment, and/or CRT dosage to investigate how the response is related to the covariates differently at every 10^{th} percentile. In other words we fit 9 quantile regression models at every 10^{th} centile. We adjust BMI using 2000 Census data and the covariates will be properly categorized, if appropriate and necessary.

For SA3 and SA4, we define final height as stature at 18 years of age or older and regress the final height on covariates such as age, gender, race, age at cancer diagnosis types of cancer treatment, and/or CRT dosage for the relationships at every 10^{th} percentile. Alternatively we can similarly adjust final height for age, race and gender using 2000 Census data and regress the adjusted final height on the covariates.

SA1 is concerned about the baseline CCSS data (available as of November 2000) that was analyzed in Oeffinger, KC. (2003).

(a) Outcomes of interest : weight and height measurements of ALL survivors and their nearest-age living siblings

(b) Subject population to be included: Out of ALL survivors (n=2,447) who were alive and 18 years of age or older at time of completion of questionnaire, those for whom complete treatment data, height and weights are available should be included as cases. Controls are a cohort of the nearest-age living sibling for whom complete treatment data, height and weights are available.

(c) Explanatory variables for cases

(i) age, gender, race, age at cancer diagnosis, whether adopted or not,

(ii) treatment, types of chemotherapy, CRT dosage, Cumulative CRT dosages, whether they were diagnosed with another cancer

(d) Explanatory variables for controls

(i) age, gender, race, whether adopted or not, whether diagnosed with a cancer or not.

SA2 is concerned about follow up data (follow-up, follow-up 2, follow-up 3) of the CCSS that are available.

(a) Outcomes of interest: weight and height measures of ALL survivors

(b) Subject population to be included: out of those ALL survivors who are included in the analysis of SA1, all for whom height and weights are available

(c) Explanatory variables: whether they were diagnosed with another cancer

SA3 and SA 4 do not require independent data collection.

6. Special Consideration:

We request two identifiers be included in the data, a subject identifier to identify longitudinal measurements from the same ALL survivors and a sibling identifier to match ALL survivors and their siblings.

7. Reference

Koenker, R and Bassett, G., (1978), Regression quantiles, *Econometrica* 46 33-50

Koenker, R. (2004) "Quantile regression for longitudinal data". *Journal of Multivariate Analysis*, 91, 74–89

Oeffinger KC, Mertens AC, Sklar CA, Yasui Y, Fears T, Stovall M, Vik TA, Inskip PD, Robison LL; Childhood Cancer Survivor Study. "Obesity in adult survivors of childhood acute lymphoblastic leukemia: a report from the Childhood Cancer Survivor Study", *J Clin Oncol.* 2003 Apr 1;21(7):1359-65