1 Background

Longitudinal studies are ubiquitous in biostatistics context. For example, in a randomized clinical trials (RCT) where patients are randomly allocated into different treatment arms and are followed over time to collect some outcome of interest. Repeated measurements will then be produced from this follow-up mechanism. One of the important features of longitudinal data is that the repeated measurements from the same subject are more "similar" to each other compared to those measures from different subjects, i.e. within subject measures tend to be intercorrelated. This feature requires special statistical techniques to handle the correlation thus valid scientific inference can be drawn from the data. As discussed in (Diggle et al., 2002), there are mainly three methods we can use to analyze longitudinal data: marginal model, transition model and random effects model. Estimates of the regression coefficients from different models have different interpretations and the choice of a model depends on study objectives, the source of correlation as well as model's capacity. In this work we will focus on applying random effects model to longitudinal data. A model that contains both random effects and fixed effects is called mixed effects model. The mixed effects model methodology, first introduced by R.A. Fisher (Fisher, 1919), is a statistical tool that is used across a wide variety of disciplines including biostatistical contexts. Mixed effects models are especially popular in researches involving repeated measurements or observations from multilevel (or hierarchical) structure where the correlation between observations is not negligible as discussed above.

Linear mixed model (LMM) is a widely used application of mixed effects methods. In brief, a LMM assumes the expected value of the outcome is a linear function of a set of covariates and observations from the same subject share a same unobserved latent variable, i.e. random effect, to account for the correlation among those observations. When conditional random effects, observations from the same subject are treated as independent. In addition, it also assumes the unobserved random error follows normal distribution.

Our concern for the widely used LMM is that in many circumstance the normality assumption of the error term can not always be satisfied, even after trying various ways of transformation. A commonly encountered situation is when there exits outliers or skewness in the outcome, where LMM is not appropriate to use. In other cases, the conditional mean may not be the primary interest and researchers may also be interested the covariate effects on the lower/upper quantiles of the outcome. *For example.* Instead of trying to fix limitations of LMM, quantile regression is a single method that can meet all above needs at once and it's becoming more and more popular in the statistical community in recent years. There are several advantages of quantile regression over the ubiquitous mean regression (a.n.a. linear regression) model. To list a few, quantile regression provides a much more comprehensive and focused insight into the association between the variables by studying the conditional quantiles functions of the outcome, which may not be observed by looking only at conditional mean of the outcome (Koenker, 2005). In quantile regression, the regression coefficients (β) are functions of the quantile (τ), and their estimates vary according to different quantiles. Thus quartile regression provides a way to studying the heterogeneity of the outcome that is associated with the covariates (Koenker, 2005). Moreover, as mentioned above, quantile regression is more robust against outliers in the outcome compared with the mean regression, which is an immediate extension from the property of quantiles.

For a linear quantile regression model, (Koenker and Bassett Jr, 1978) introduced a method in estimating the conditional quantiles. As an introductory material, Koenker (Koenker and Hallock, 2001) briefly covers the fundamentals of quantile regression, parameter estimation techniques, inference, asymptotic theory, etc., his book provides more comprehensive and deeper introduction on quantile regression related topics (Koenker, 2005). (Yu and Moyeed, 2001) introduced the idea of Bayesian quantile regression by modeling the error term using asymmetric Laplace distribution (ALD) followed the idea proposed in (Koenker and Bassett Jr, 1978). (Kozumi and Kobayashi, 2011) developed a Gibbs sampling algorithm for Bayesian quantile regression models, in which they used a location-scale mixture representation of the ALD. Many works have been done to extend the quantile regression method to accommodate longitudinal data. (Jung, 1996) developed a quansi-likehood method for median regression model for longitudinal data. (Geraci and Bottai, 2007) proposed to fit the quantile regression for longitudinal data based on ALD and the estimation is made by using Monte

Carlo EM algorithm. Later on, (Liu and Bottai, 2009) followed the idea of (Geraci and Bottai, 2007) and extended the model from random intercept to including random slope as well. The study of longitudinal data using quantile regression is becoming popular in recent year. (Fu and Wang, 2012) proposed a working correlation model for quantile regression for longitudinal data, a induced smoothing method was used to make the inference of the estimators. Fully Bayesian techniques and Gibbs sampling algorithm became possible when the error term is decomposed as the mixture of normal and exponential random variables for the quantile linear mixed model (Kotz et al., 2001) , see (Luo et al., 2012) and (Kozumi and Kobayashi, 2011) for applications. The fully Bayesian method is appealing to us because it is easy to implement and to make inference, the uncertainty of the unknowns is taken into account, and it is flexible in the distribution of random effects. The detailed background about Bayesian quantile linear mixed model will be introduced in Section **??**.

Time-to-event data is another common and interesting data type that attracts researchers' great interest in biostatistical field.

Another concern that frequently arises from longitudinal study is that the subjects may be lost at follow up due to events. Such cases are called informative drop-outs (ID) when the event occurrences are associated with the longitudinal outcome. Simply ignore the missing mechanism can lead to biased estimates of the parameters in the longitudinal model. The informative drop-out problem has attracted much attention from the statisticians and a wide array of methods have been proposed to handle this issue (Diggle and Kenward, 1994) (Lipsitz et al., 1997) (Touloumi et al., 2003) (Yuan and Yin, 2010). For more information in this research area, there are several good review papers of modeling longitudinal data with non-ignorable drop-outs, including (Diggle et al., 2007) and (Hogan et al., 2004). *[To-do: put some review of previous work and their limitations and propose our model then state the advantage.]*

[Add background for topic two here.]

To sum up, this research project focuses on developing and applying new statistical methods in analyzing longitudinal data and will cover the following topics. First, a fully Bayesian framework in modeling conditional quantiles of the longitudinal outcome, incorporating with time-to-event process to account for the informative drop-outs issue. In this part, the linear quantile mixed model is used so that the within subject correlation is accounted and the results also will be more robust to potential outlying observations. Since we used quantile regress instead of mean regression, a more comprehensive insight about the association between the outcome and covariates will be learned. In this model we can also directly gain a sense of the association between the longitudinal and time-to-event processes. Second, *[to-do]*.

References

Peter Diggle and Michael G Kenward. Informative drop-out in longitudinal data analysis. *Applied statistics*, pages 49–93, 1994.

Peter Diggle, Patrick Heagerty, Kung-Yee Liang, and Scott Zeger. *Analysis of longitudinal data*. Oxford University Press, 2002.

Peter Diggle, Daniel Farewell, and Robin Henderson. Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(5): 499–550, 2007.

Ronald A Fisher. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the royal society of Edinburgh*, 52(02):399–433, 1919.

Liya Fu and You-Gan Wang. Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56(8):2526–2538, 2012.

Marco Geraci and Matteo Bottai. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–154, 2007.

Joseph W Hogan, Jason Roy, and Christina Korkontzelou. Handling drop-out in longitudinal studies. *Statistics in medicine*, 23(9):1455–1497, 2004.

Sin-Ho Jung. Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, 91(433):251–257, 1996.

Roger Koenker. *Quantile regression*. Cambridge university press, 2005.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Roger Koenker and Kevin Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.

Samuel Kotz, Tomasz Kozubowski, and Krzysztof Podgorski. *The Laplace Distribution and Generalizations: A Revisit With Applications to Communications, Exonomics, Engineering, and Finance*. Number 183. Springer, 2001.

Hideo Kozumi and Genya Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation*, 81(11):1565–1578, 2011.

Stuart R Lipsitz, Garrett M Fitzmaurice, Geert Molenberghs, and Lue Ping Zhao. Quantile regression methods for longitudinal data with drop-outs: Application to cd4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46 (4):463–476, 1997.

Yuan Liu and Matteo Bottai. Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics*, 5(1), 2009.

Youxi Luo, Heng Lian, and Maozai Tian. Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, 82(11):1635–1649, 2012.

G Touloumi, AG Babiker, MG Kenward, SJ Pocock, and JH Darbyshire. A comparison of two methods for the estimation of precision with incomplete longitudinal data, jointly modelled with a time-to-event outcome. *Statistics in medicine*, 22(20):3161–3175, 2003.

Keming Yu and Rana A Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4): 437–447, 2001.

Ying Yuan and Guosheng Yin. Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, 66(1):105–114, 2010.