

## Quantile Analysis for human quantitative traits in GWAS data

Ying Wei\*

Columbia University, New York, USA [ying.wei@columbia.edu](mailto:ying.wei@columbia.edu)

Mengling Liu

New York University, New York, USA [Mengling.Liu@nyumc.org](mailto:Mengling.Liu@nyumc.org)

Iuliana Lonita-Laza

Columbia University, New York, USA [ii2135@columbia.edu](mailto:ii2135@columbia.edu)

To understand the genetic basis of human quantitative traits has been a long-standing research topic in the field of genetics, and is of great importance for public health. Quantitative traits are often associated with the susceptibility to many diseases, including autism, cancer, diabetes and numerous others. Hence, it is critical in order to address the high prevalence of many diseases. The main strategy in genetic association studies is to identify genetic variants that influence the mean of a quantitative trait of interest. However, it is often the case that the most vulnerable or high-risk group to certain diseases are the subjects who have either high or low values for their quantitative traits. Recent studies suggest that genotypes often influence the entire distribution of quantitative traits, and their impact could differ at various quantiles. We propose quantile-based analysis to understand the genetic basis of quantitative traits.

The main tool for identifying genetic variants associated with complex traits has been the GWAS, where a large number of genetic variants, typically 500K to 1 million variants, are measured across the genome for each individual in the study. Most GWAS are collected using case-control sampling schemes. Hence these data are not representative samples to the general population. Naively estimated regression quantiles could be substantially biased from the true association in general. Existing methods recovering the population associations are likelihood based, and only estimate the genetic effect on the mean of the trait. Hence, they cannot be applied directly to obtain quantile estimates. In order to make consistent and efficient estimation on conditional quantiles, herein we construct a new family of estimating equations, and also develop related statistical tools for inference.

**Key Words:** quantile regression; estimating equations; secondary analysis; GWAS; case-control studies.