

# Review of “Characterization of direct and/or indirect genetic associations for multiple traits in longitudinal studies of disease progression”

Jiaxin Hu

June 8, 2021

This paper proposes a joint model to distinguish and estimate of direct and indirect effects of the single nucleotide polymorphisms (SNPs) to the time-to-event traits. Particularly, the indirect effects are reflected through the association of the SNPs to the intermediate quantitative traits (QTs) which are related to the diseases. The paper focuses on the application to the Diabetes Control and Complications Trial (DCCT). The real-data-based simulation and real data analysis indicates the ability of the proposed method to classify the SNPs effects and improve the estimation accuracy. Here are my comments:

1. In the Analysis of the simulated data, the paper compares the simulation results from four models. However, the four analysis schemes are all based on the proposed model. It would be more interesting to see the comparison with other methods that tackles the same questions in the literature.
2. The paper emphasizes the accuracy improvements by including the measurement error to the longitudinal trait and the dependencies between the time-to-event. However, mixed-effect models are usually computationally expensive and time consuming, especially with large datasets. There exists a trade-off between the computation and accuracy in practice. Therefore, it would be helpful to compare the computational complexity for the proposed model with and without the measurement errors and dependencies between time-to-event traits.
3. The paper mentions the right-censored event time  $T_{i(k)}$  and even indicator  $\delta_{i(k)}$  in Model Formulation. However, these two variables do not occur in the three main sub-models. More explanations about these two variables would be helpful.

## Minors:

1. The exposition of the model formulation should be improved. Particularly,
  - the paper should state the single-trait models firstly and then state the changes with dependent traits.
  - the paper should state the models in terms of matrices or tensors to avoid the heavy notations.
  - the paper should state only the main parameters such as  $\beta_{g(l)}$  and  $\gamma_{g(l)}$  in Equation 2, 3 and suppress the nuisance for better presentation.
  - the vector  $b_i$  in line 145 should be  $b_i = ((b_{i0(1)}, b_{i1(1)}), \dots, (b_{i0(L)}, b_{i1(L)}))'$ .

- the dimension of matrix  $D$  in line 147 should be clarified as  $D \in \mathbb{R}^{2L \times 2L}$  with  $D_{i,j} \in \mathbb{R}^{2 \times 2}, i, j \in [L]$ .
2. In line 148, why the covariance matrices  $D_{l,l}$  are unstructured? More explanations are needed.
  3. It would be helpful to improve the resolutions of the figures in the main text.