

# Review of: "Benchmarking statistical methods for analyzing parent-child dyads in genetic association studies"

Miriam Gjerdevik<sup>1</sup>

<sup>1</sup> The Norwegian Institute of Public Health (NIPH)

**Potential competing interests:** The author(s) declared that no potential competing interests exist.

I appreciate the opportunity to review the paper by Debashree Ray and co-authors. The authors present an overview of statistical methods for analyzing standard autosomal effects in genetic association studies, with particular focus on benchmarking the performance of parent-child dyads. They compare different study designs and methods in extensive simulation experiments as well as in genotyped and imputed data from the multi-ethnic Gene, Environment Association Studies consortium (GENEVA) on nonsyndromic cleft lip with or without cleft palate. Type I errors and statistical power are assessed using a genome-wide significance level of  $5 \times 10^{-8}$ .

It is unknown to what extent methods for analyzing dyads protect against population stratification, and the current investigations of their type I errors in heterogeneous populations are interesting. Possible population stratification should be controlled for, and the authors' effort to meta-analyze results from different ancestry groups is appropriate.

The paper is concise and mostly well written, and the comparison of study designs and statistical methods in multi-ancestry data is useful. I have, however, a few suggestions for improvement.

In the effort of benchmarking the performance of the dyad design, I am missing a comparison with the standard case-control design, which can be heavily influenced by population stratification. In particular, I would like to see how their type I error rates compare to those of the dyad, triad, and hybrid designs in genetically heterogeneous samples.

An advantage of family-based study designs is the ability to analyze genetic effects beyond the standard autosomal effects, such as maternal and parent-of-origin effects. I believe an extension of the current paper to also include these effects would be valuable. However, if the authors still feel this is beyond the scope of their paper, the restriction to standard genetic effects should be made clear in the abstract (and perhaps even in the title).

The TDT implicitly assumes multiplicative effects of alleles, while several alternatives are available for the gTDT (e.g., additive, dominant, and recessive effects). For the gTDT, which alternative (pre-specified inheritance model) was used for the power calculations? Although one would not know in advance the true mode of inheritance, these assumptions are likely to affect the statistical power in Figures 3 and 4. In the current paper, Haplin is employed using a 2 or 4 degree-of-freedom test, estimating both single- and double-dose effects separately. Of note, Haplin also accommodates a 1 degree-of-freedom test for multiplicative effects, and it is reasonable to assume the statistical power is even higher under the multiplicative inheritance model. I am not expecting the authors to run through all possible alternatives for all tests, but an overview of the underlying assumptions used for the power calculations would yield a more straightforward comparison and ease the interpretation of results.

I would not expect case-mother/control-mother dyads to achieve better power than case-parent/control-parent triads in the 30:70 case-to-control ratio scenarios in Figure 4. The rare disease assumption does not seem to be met in the simulation experiments involving independent controls, which may influence the results.

The authors aim to review existing statistical methods for analyzing genome-wide data on dyads. However, other software such as EMIM<sup>[1]</sup> and UNPHASED<sup>[2]</sup> have not been mentioned and should be referenced.

Family-based and population-based study designs have their own set of advantages and disadvantages. Although strengths and limitations have been adequately addressed elsewhere, a short overview would be appreciated. For example, independent control families may identify deviations from Mendelian transmission or false positive results. Independent controls are also necessary to estimate the main effect of an environmental exposure. The latter is important for the Environmental influences on Child Health Outcomes (ECHO) study (mentioned as a motivating factor for the current paper), which aims to detect both environmental and genetic exposures associated with child development and disease. Note that Haplin also enables the estimation of gene-environment interactions from a wide variety of family- and population-based study designs.<sup>[3] [4]</sup>

Regarding the time consumption, I would like to mention that Haplin has a built-in parallel implementation, which efficiently speeds up genome-wide scans.

Unfortunately, I cannot find Table 1, the summary of your recommendations, neither in the main document nor in the supplementary material. I believe such a table will serve as a useful supplement to the *Recommendations* section.

In a few instances, the authors introduce unnecessary notation which are not re-used (e.g.,  $N$  individuals).

As a minor suggestion, I would avoid redundant notation to improve readability.

I hope my suggestions are helpful, and I look forward to reading the final version of the paper.

Sincerely,

Miriam Gjerdevik

## References

1. ^Richard Howey, Heather J Cordell. (2012). *PREMIM and EMIM: tools for estimation of maternal, imprinting and interaction effects using multinomial modelling*. *BMC Bioinformatics*, vol. 13 (1). doi:10.1186/1471-2105-13-149.
2. ^Frank Dudbridge. (2008). *Likelihood-Based Association Analysis for Nuclear Families and Unrelated Subjects with Missing Genotype Data*. *Hum Hered*, vol. 66 (2), 87-98. doi:10.1159/000119108.
3. ^Øivind Skare, Astanand Jugessur, Rolv Terje Lie, Allen James Wilcox, et al. (2012). *Application of a Novel Hybrid Study Design to Explore Gene-Environment Interactions in Orofacial Clefts*. doi:10.1111/j.1469-1809.2012.00707.x.
4. ^Miriam Gjerdevik, Øystein A. Haaland, Julia Romanowska, Rolv T. Lie, et al. (2017). *Parent-of-origin-environment interactions in case-parent triads with or without independent controls*. *Annals of Human Genetics*, vol. 82 (2), 60-73. doi:10.1111/ahg.12224.