

**Supplementary Material “On rare-variant analysis in
population-based designs: Decomposing the likelihood to two
informative components” by**

Sungho Won^{1,2*}, Youngdoe Kim¹, Christoph Lange^{3,4,5,6,7}

¹Department of Applied Statistics, Chung-Ang University, Seoul, Korea

²The Research Center for Data Science, Chung-Ang University, Seoul, Korea

³German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

⁴Institute for Genomic Mathematics, University of Bonn, Germany

⁵Department of Biostatistics, Harvard School of Public Health, Boston MA U.S.A.

⁶Harvard Medical School, Boston MA U.S.A.

⁷Center for Genomic Medicine, Brigham and Women's Hospital Boston MA U.S.A.

*Corresponding Author: Sungho Won,

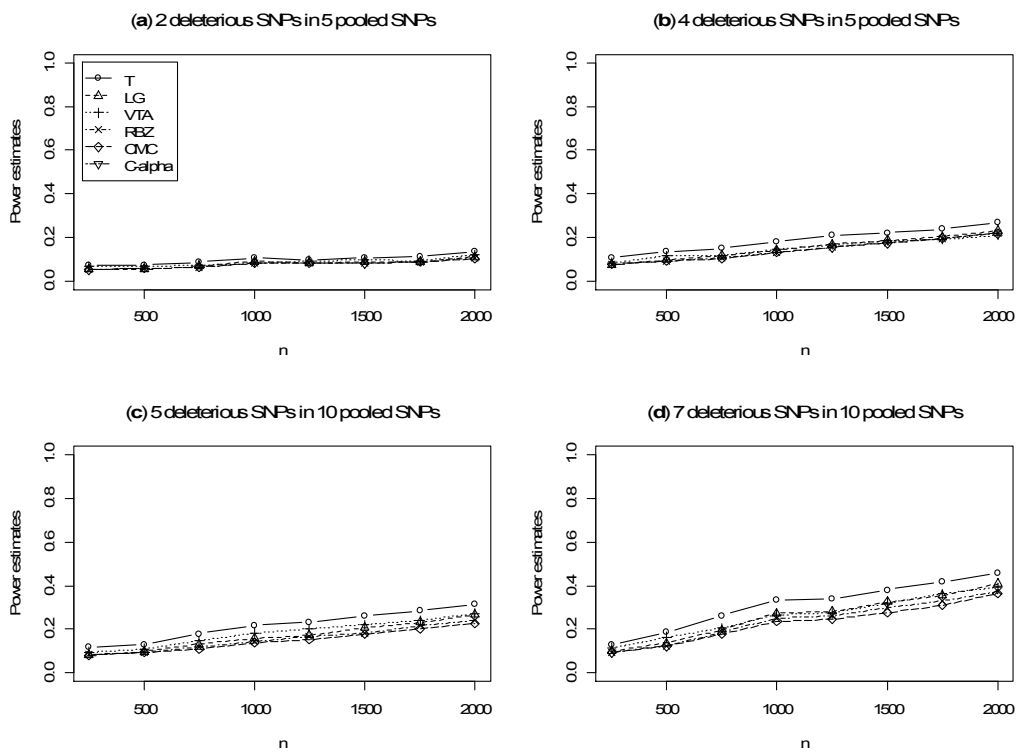
Dept. of Applied Statistics, Chung-Ang Univ

221 Heukseok-dong Dongjak-gu Seoul 156-756 Korea

(Email) swon@cau.ac.kr,

(Tel) +82-2-820-5217

Supplementary Figure 1. Empirical power estimates at the 0.05 significance level under additive disease mode of inheritance. Empirical power estimates for the proposed method at the 0.05 significance level were calculated from 2,000 replicates under the additive disease mode of inheritance. P-values were obtained with 1,000 permuted samples. 5 or 10 rare variants which minor allele frequencies were between 0.001 and 0.02 were collapsed. The numbers of both cases and controls were n and the disease prevalence was assumed to be 0.15. The proposed method was denoted by T, and it was compared with other approaches; combined multivariate and collapsing (CMC) method, the variable threshold approach (VTA), and rank-based Z-test (RBZ), logistic regression where the disease status is regressed on the weighted sum of genotypes (LR), and variance component approaches such as the C-alpha and the SKAT-O tests.



Supplementary Figure 2. Empirical power estimates at the 0.05 significance level under the dominant disease mode of inheritance. Empirical power estimates for the proposed method at the 0.05 significance level were calculated from 2,000 replicates under the dominant disease mode of inheritance. P-values were obtained with 1,000 permuted samples. 5 or 10 rare variants of which minor allele frequencies were between 0.001 and 0.02 were collapsed. n indicates the numbers of both cases and controls, and the disease prevalence was assumed to be 0.15. The proposed method was denoted by T, and it was compared with other approaches; combined multivariate and collapsing (CMC) method, the variable threshold approach (VTA), and rank-based Z-test (RBZ), logistic regression where the disease status is regressed on the weighted sum of genotypes (LR), and variance component approaches such as the C-alpha and the SKAT-O tests.

