Supplementary Information for Genome-Wide Gene-Environment Interactions on Quantitative Traits using Family Data

Colleen M. Sitlani[1], Josée Dupuis[2], Kenneth M. Rice[3], Fangui Sun[2],

Achilleas N. Pitsillides[2], L. Adrienne Cupples[2] and Bruce M. Psaty[4]

[1]Department of Medicine, University of Washington, Seattle, WA

[2]Department of Biostatistics, Boston University School of Public Health, Boston, MA

[3]Department of Biostatistics, University of Washington, Seattle, WA

[4]Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA; Group Health Research Institute, Group Health Cooperative, Seattle, WA
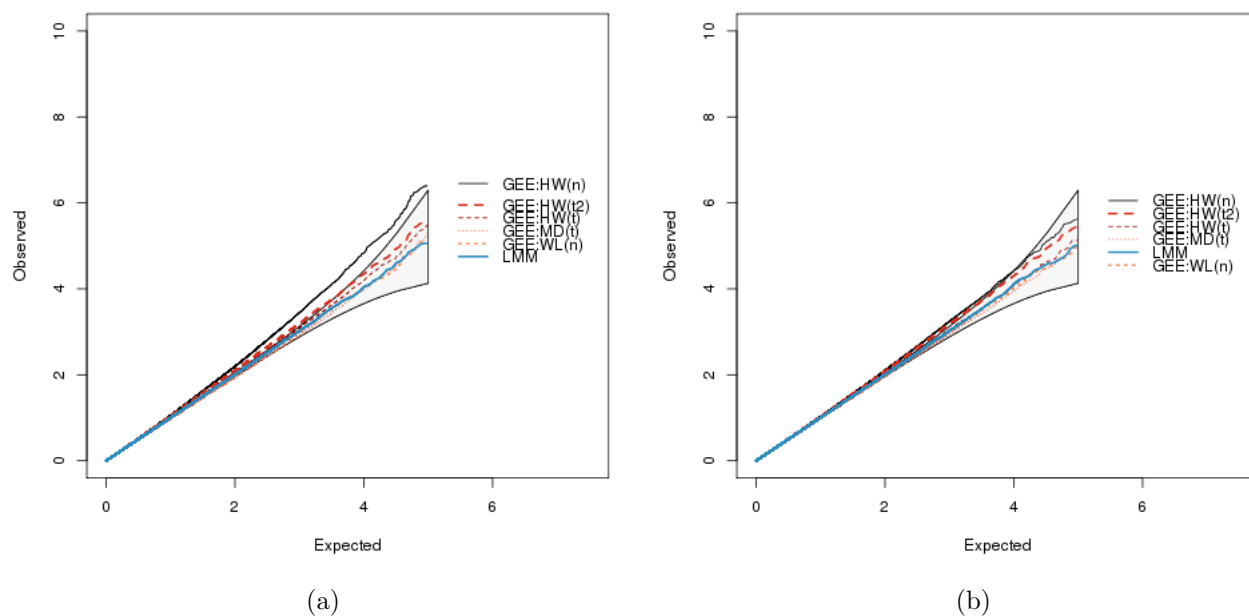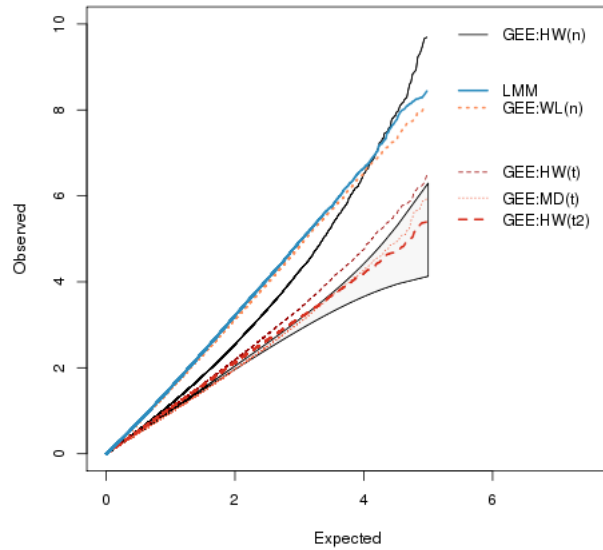
Supplementary Figures
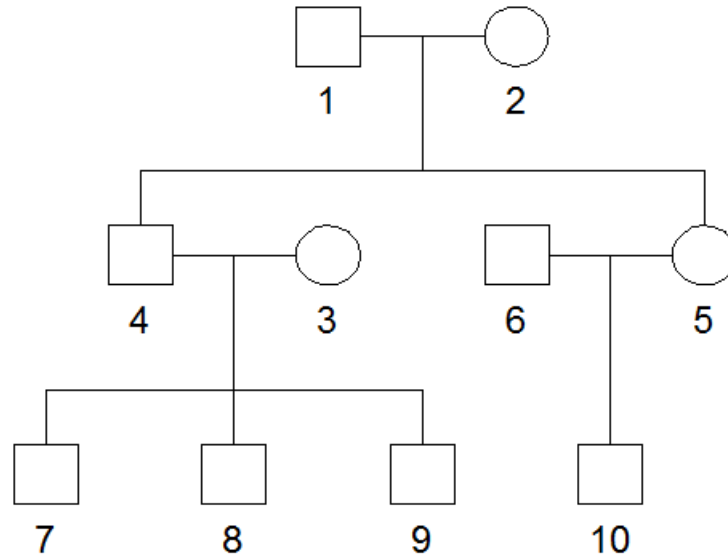
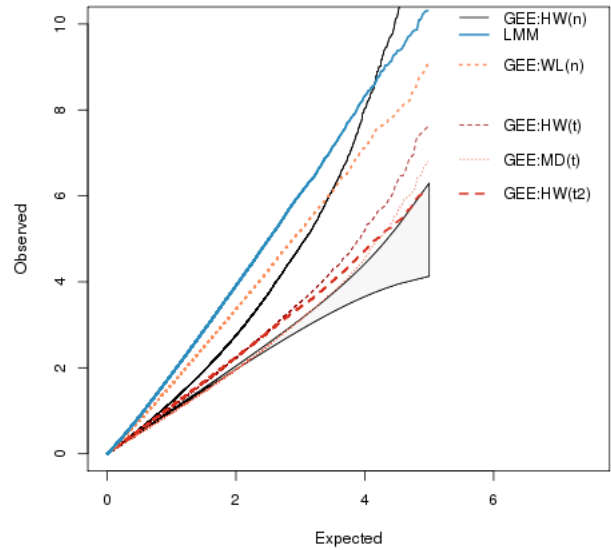Supplementary Figure 1: Sample pedigree for three-generational family generated in simulations.

(a)

(b)

(c)

(d)
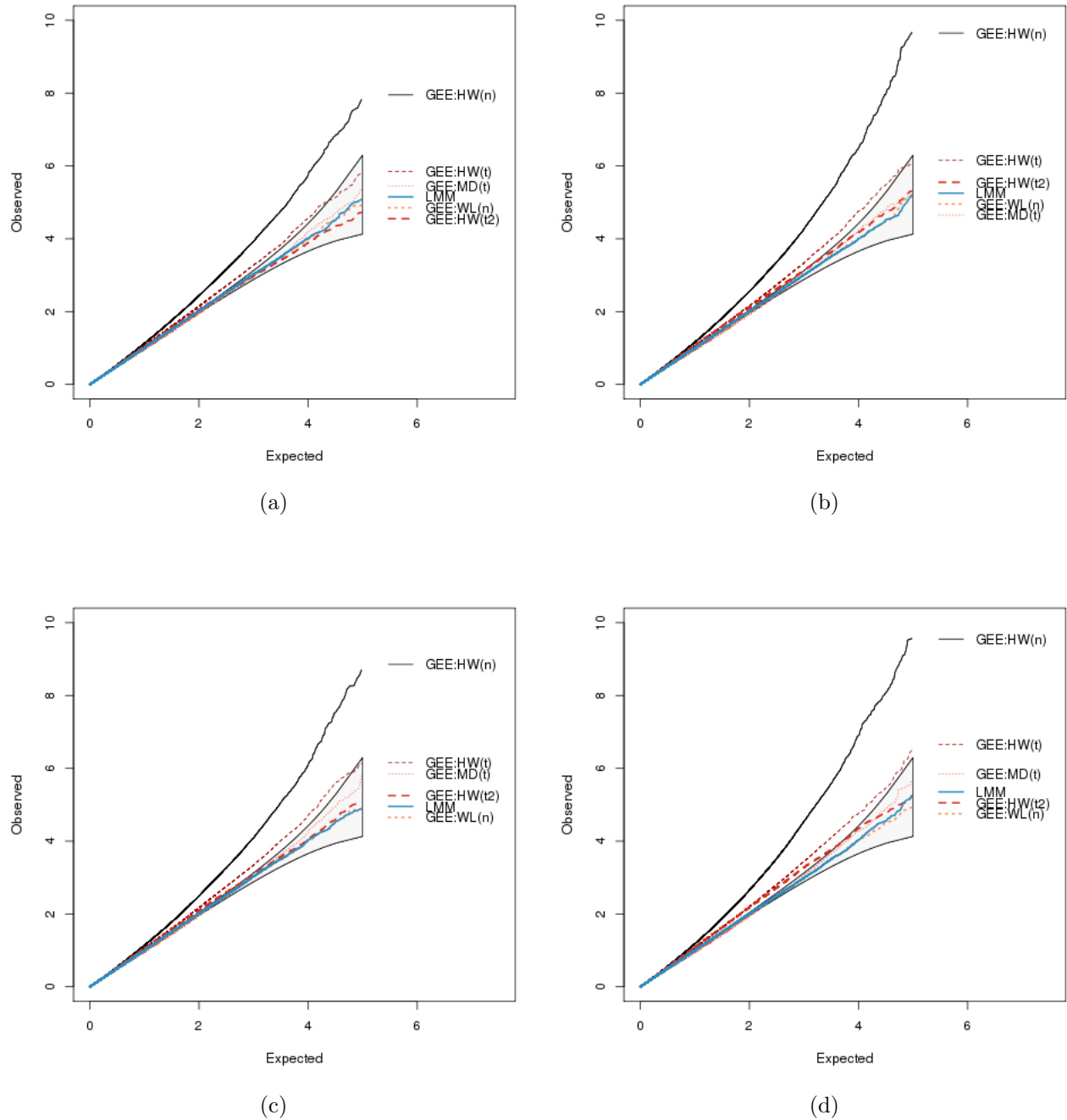
Supplementary Figure 2: Quantile-quantile plots of -log10(p-values), with each plot derived from one million simulations. Simulated data in the top row are from 200 nuclear families, each of size 5, while data in the bottom row are from 100 three-generational families, each of size 10. Subfigures (a) and (c) assume a single cross-sectional measurement with MAF=0.10 and P(exposure)=10%, while subfigures (b) and (d) assume four longitudinal measurements with MAF=0.05 and P(exposure)=5%. GEE models use either sandwich (HW), Mancl & DeRouen (MD), or Wang & Long (WL) standard error estimates, with reference distribution being normal (n), t with Satterthwaite estimates of degrees of freedom (t), or approximate estimates of degrees of freedom (t2).

3

(a)   (b)

Supplementary Figure 3: Quantile-quantile plots of -log10(p-values), with each plot derived from one million simulations. Simulated data are single cross-sectional measurements from 100 three-generational families, each of size 10. Subfigure (a) assumes MAF=0.10 and P(exposure)=40%, while subfigure (b) assumes MAF=0.40 and P(exposure)=40%. GEE models use either sandwich (HW), Mancl & DeRouen (MD), or Wang & Long (WL) standard error estimates, with reference distribution being normal (n), t with Satterthwaite estimates of degrees of freedom (t), or approximate estimates of degrees of freedom (t2).
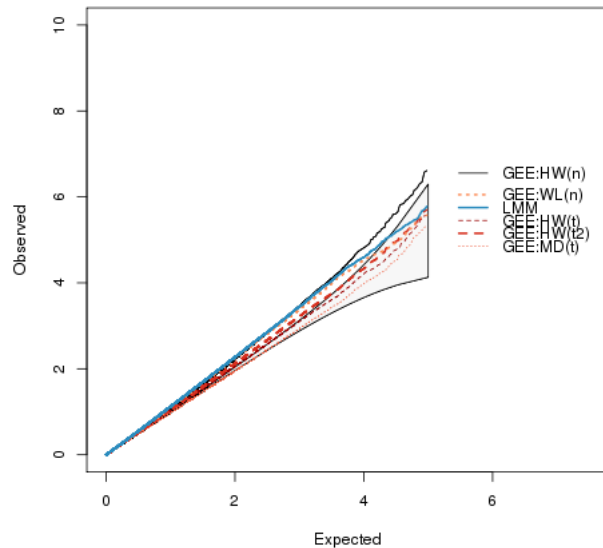
Supplementary Figure 4: Quantile-quantile plots of -log10(p-values), with each plot derived from one million simulations. Simulated data are from 100 three-generational families, each of size 10. Models are mis-specified because outcome variance is twice as high among exposed participants as it is among unexposed participants. Subfigure (a) assumes a single cross-sectional measurement with MAF=0.10 and P(exposure)=10%; subfigure (b) assumes four longitudinal measurements with MAF=0.05 and P(exposure)=5%; and subfigure (c) assumes a single cross-sectional measurement with MAF=0.10 and P(exposure)=40%. GEE models use either sandwich (HW), Mancl & DeRouen (MD), or Wang & Long (WL) standard error estimates, with reference distribution being normal (n), t with Satterthwaite estimates of degrees of freedom (t), or approximate estimates of degrees of freedom (t2).