Submitter Name: Brad Verhulst
Submitted email: bverhuls@msu.edu

Developing Genome-Wide Structural Equation Modeling (GW-SEM) to analyze the genetic associations with substance use phenotypes

Brad Verhulst[1], Shaunna L. Clark[1], Joshua Pritikin[2,3], & Michael C. Neale[2,3]

[1]Department of Psychology, Michigan State University; [2]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University; [3]Department of Psychiatry, Virginia Commonwealth University

Background: Substance use (SU) behaviors are complex traits that exist within an interdependent network of cultural and biological factors. The extant SU phenotypic literature addresses this complexity through multivariate techniques like structural equation modeling (SEM), however, current genome-wide association study (GWAS) software do not reflect this complexity.

Rationale: Advancing our understanding of the genetic underpinnings of addictive behaviors requires the development of statistical methods and software packages that integrate phenotypically complex models with GWAS data. We developed GW-SEM, a GWAS software packages that combines SEM with GWAS data, to address these issues.

Results: GW-SEM has widespread applicability for a range of SU behaviors, as it extends phenotypic methods users are familiar with to a genomic scale allowing researcher to test hypotheses that were previously untestable in existing software packages. For example, GW-SEM users can simultaneously examine multiple substances or the sequential stages of addiction formation (e.g. from initiation to regular use to substance dependence), investigate the multi-factorial structure of SU traits to examine genetic pleiotropy, and even use growth modeling to determine the genetics underpinnings of the development of SU. We further demonstrate the utility of GW-SEM through an application to the structure of nicotine dependence.

Discussion: GW-SEM provides a user-friendly, fast, reliable method of examining the complex genomic interplay between myriad substance use phenotypes. With publicly available genomic data (e.g., UK Biobank and dbGaP) becoming increasingly accessible, it is essential that we develop the statistical tools that will make the best use of genomic data.