Sparse multiple factor analysis to integrate genetic data, neuroimaging features, and attention-deficit/hyperactivity disorder domains

Vilor-Tejedor N, Alemany S, Cáceres A, Bustamante M, Mortamais M, Pujol J, Sunyer J, González JR.


Abstract

OBJECTIVES:
We proposed the application of a multivariate cross-sectional framework based on a combination of a variable selection method and a multiple factor analysis (MFA) in order to identify complex meaningful biological signals related to attention-deficit/hyperactivity disorder (ADHD) symptoms and hyperactivity/inattention domains.

METHODS:
The study included 135 children from the general population with genomic and neuroimaging data. ADHD symptoms were assessed using a questionnaire based on ADHD-DSM-IV criteria. In all analyses, the raw sum scores of the hyperactivity and inattention domains and total ADHD were used. The analytical framework comprised two steps. First, zero-inflated negative binomial linear model via penalized maximum likelihood (LASSO-ZINB) was performed. Second, the most predictive features obtained with LASSO-ZINB were used as input for the MFA.

RESULTS:
We observed significant relationships between ADHD symptoms and hyperactivity and inattention domains with white matter, gray matter regions, and cerebellum, as well as with loci within chromosome 1.

CONCLUSIONS:
Multivariate methods can be used to advance the neurobiological characterization of complex diseases, improving the statistical power with respect to univariate methods, allowing the identification of meaningful biological signals in Imaging Genetic studies.