Predicting Attention Problems from Brain Connectivity Using High-Dimensional Poisson Regression

IMPS Conference 2025

Kelly A. Duffy, Nathaniel E. Helwig

ADHD is a Common and Complex Disorder with Numerous Identified Risk Factors and Correlates

- ADHD is very common, affecting roughly 5-10% of school-aged children (Scahill & Schwab-Stone, 2000)
- Many factors associated with or predictive of ADHD have been identified:
  - **Genetic**: heritability has been estimated to be as high as 88% (Larsson et al., 2014)
  - **Environmental**: parental substance abuse history (Clark et al., 1997; Knopik et al., 2005), socioeconomic status (Russell et al., 2016)
- Many differences have been identified in the brains of individuals with ADHD relative to controls, both structurally (Hoogman et al., 2017) and functionally, including at task (Cubillo et al., 2012) and at rest (Posner et al., 2014)
  - Many rs-fMRI studies have focused on changes in *functional connectivity*, the correlated activity among different regions of the brain (e.g., Mills et al. (2018))

Despite Large Number of Findings, There Has Been Limited Success in Neuroimaging Predictive Models

- Early success in small studies (e.g., Zhu et al. (2008)) has not replicated
  - Sample size and accuracy are actually negatively correlated (Pulini et al., 2019)

- Many studies have focused on *classification* of known diagnoses
  - Average classification accuracy in ADHD-200 study was 56% (The ADHD-200 Consortium, 2012), average sensitivity much lower

- Lack of consistent findings in recent meta-analyses using case-control approach (Cortese et al., 2021; Gao et al., 2019; Hoogman et al., 2019; Sutcubasi et al., 2020) raises the question of whether this approach obscures significant heterogeneity in individuals (Cortese et al., 2021)

Studies Using Large, Representative Samples and Dimensional Measures in Exploratory Models May Improve Results

- Approaches that focus more on assessing ADHD along a *continuum* may be more fruitful and generalizable
- Likewise, a large, representative sample – ideally non-clinical in recruitment – may give us a more realistic indication of what brain changes are related to ADHD in overall population
- Analyzing many potential brain-related variables in this setting may be more informative
- The Adolescent Brain Cognitive Development (ABCD) Study provides us a unique opportunity to address all of these concerns
  - 11,878 subjects at baseline (when participants are ~9-10 years old), aim is to continue data collection until age ~20
  - Population-based and demographically diverse sample, intentionally sampled to be reflective of US children
  - Non-binary (i.e., dimensional) measures of ADHD are collected
  - Many measures of brain-related activity (and FC in particular) are collected

Present Study Aims to Leverage ABCD to Identify What Features are Most Predictive of ADHD Symptomatology

There are a number of factors to take into consideration for modeling:

- Large number of predictors (87 total: 78 within- and between-network resting state functional connectivity predictors and a number of demographic and family history variables)
- A Poisson response (a count of symptoms on the Attention Problems subscale of the CBCL)
- Potentially nonlinear effects

We can address all of these using a (groupwise) penalized nonparametric Poisson regression model, implemented in the `grpnet` (Helwig, 2023) package in R

Penalized Regression Can Produce More Stable and Sparse Coefficient Estimates

Most common penalty is the LASSO penalty (Tibshirani, 1996)

- Either used alone or in combination with the ridge penalty (Hoerl & Kennard, 1970) as part of an elastic net penalty (Zou & Hastie, 2005)
- LASSO pushes some coefficients completely to zero, helping to identify which predictors are most relevant

LASSO penalty takes the form

$$P_1(\boldsymbol{\beta}|\lambda) = \sum_{j=1}^{p} |\beta_j|$$

added to the OLS function

$$PLS_1(\beta_0, \boldsymbol{\beta}) = OLS(\beta_0, \boldsymbol{\beta}) + \lambda\, P_1(\boldsymbol{\beta}|\lambda)$$

Despite its Popularity, LASSO Penalty has Well-Documented Drawbacks

- LASSO has been shown to be variable-selection consistent under particular sets of assumptions (Zhang & Huang, 2008)
    - But this set of necessary assumptions is quite strict, and in general LASSO is *not* variable-selection consistent (Zhao & Yu, 2006)
- LASSO also has non-ignorable bias in the parameter estimates of the non-zero coefficients, particularly for larger effects (Xiao & Xu, 2015; Zhao & Yu, 2006) and a large number of false positive effects (Xiao & Xu, 2015)
- LASSO has particularly poor performance, due to a violation of many key assumptions, in variable selection in the presence of correlated variables (Zhao & Yu, 2006)
- But alternative penalties can be used in place of the lasso that may mitigate some of these issues and improve variable selection

SCAD and MCP Are Two Alternative Penalties

SCAD penalty (Fan & Li, 2001) takes the form

$$P_1(\boldsymbol{\beta}|\lambda,\gamma) = \sum_{j=1}^{p} \rho_1(|\beta_j|;\lambda,\gamma)$$

$$\rho_1(|\beta_j|;\lambda,\gamma) = \begin{cases} |\beta_j|\lambda & \text{if } |\beta_j| \leq \lambda \\ \frac{2\gamma\lambda|\beta_j| - \beta_j^2 - \lambda^2}{2(\gamma-1)} & \text{if } \lambda < |\beta_j| < \gamma\lambda \\ \frac{\lambda^2(\gamma+1)}{2} & \text{if } |\beta_j| \geq \gamma\lambda \end{cases}$$

MCP penalty (Zhang, 2010) takes the form

$$\rho_1(|\beta_j|;\lambda,\gamma) = \begin{cases} \lambda|\beta_j| - \frac{1}{2a}\beta_j^2 & \text{if } |\beta_j| \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2 & \text{if } |\beta_j| > \gamma\lambda \end{cases}$$

Both have been shown to have less biased estimates than the LASSO and better consistency

- Both can be added to an elastic net (as the L_1 penalty) and extended to a groupwise form (needed for `grpnet`)

Present Work Aims to Compare Performance of Different Penalties in Predicting ADHD

Data from $N = 7979$ subjects in ABCD study used

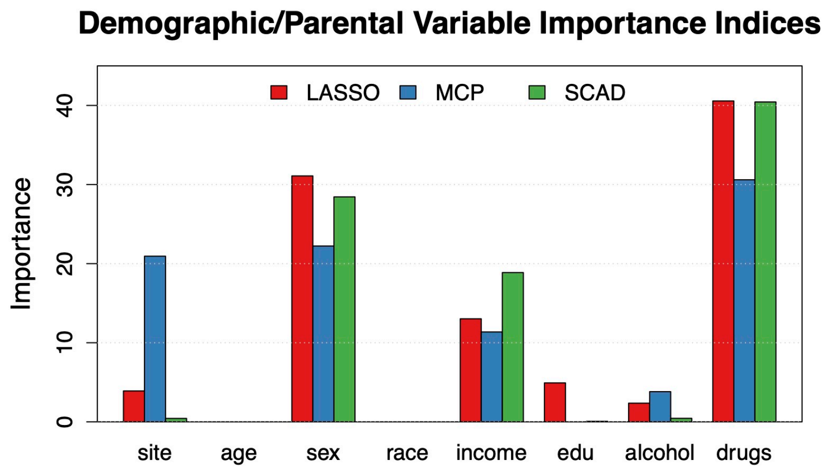
- Outcome: sum scores from Attention Problem subscale of Child Behavior Checklist (CBCL)
- Predictors: 78 within- and between-network rs-FC variables, demographic (HHI, parental education, race/ethnicity, age, sex assigned at birth, study site) and parental history (of problematic drug and alcohol use) covariates

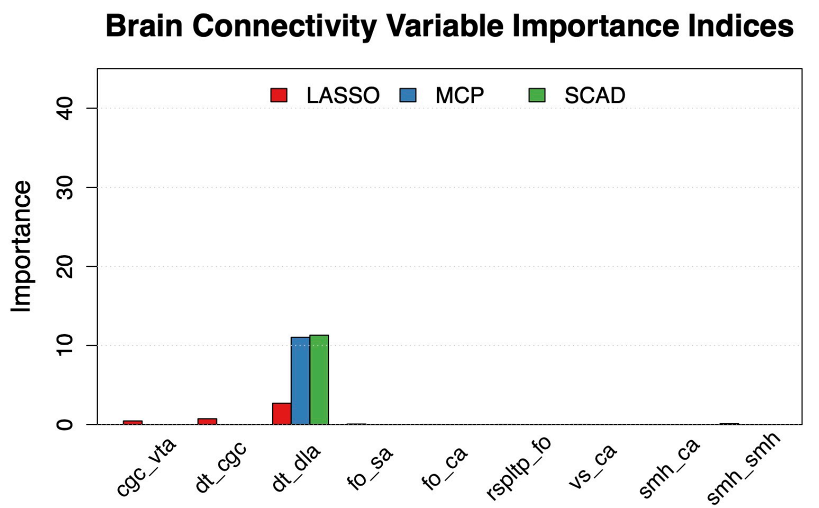
Models fit in `grpnet`

- Response modeled as Poisson with log link function
- Each feature expanded using a spectral smoothing spline basis to capture nonlinear effects
- Compared three L_1 penalties in the elastic net: LASSO, SCAD, and MCP
  - Tuning for each done using `cv.grpnet()` using 10-fold CV to tune hyperparameters (α, λ, and γ for SCAD/MCP)

Relative contribution of predictors assessed using variable importance indices

SCAD and MCP Penalties Produced Much Sparser Solutions Than LASSO





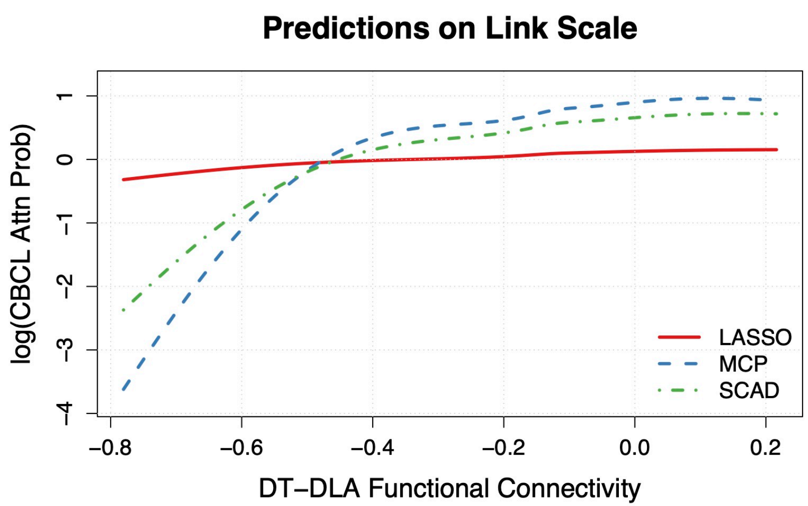
SCAD and MCP selected a single brain predictor, while LASSO selected 9

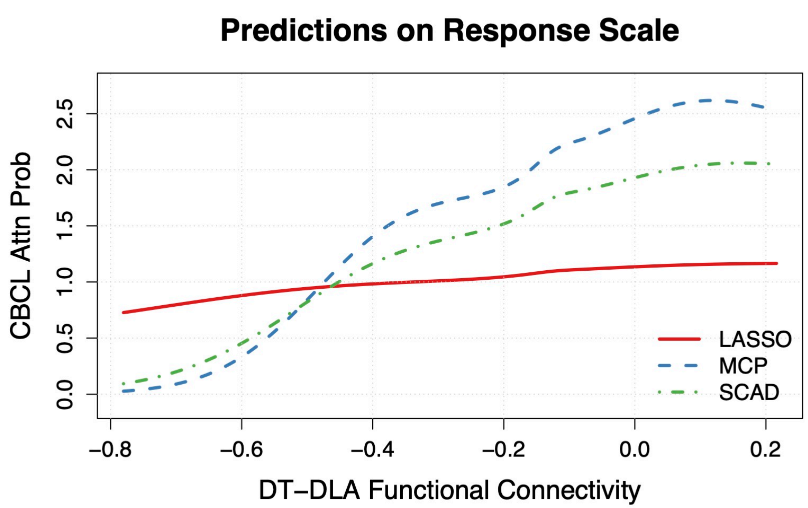
- All penalties selected a similar number of demographic variables, with MCP selecting the fewest (5 versus 6 for SCAD and LASSO)

Despite variability in sparseness, models explained a similar amount of variance (Poisson deviance)

	R^2
LASSO	8.516%
MCP	9.430%
SCAD	7.928%

Single Network Selected by SCAD/MCP Identified in Past Work as Best Predictor





- Results suggest that increased correlation between the default mode (DT) and dorsolateral attention (DLA) networks are related to increased attention problems
- Effect plots show nonlinearity in effect, suggesting nonparametric models may improve predictions
  - Also suggest LASSO may be resulting in overly linear effect
- A recent large multi-cohort analysis (Norman et al., 2023) also identified increased DT-DLA as having the largest effect size in ADHD

Conclusions

These results highlight the utility of using penalized regression in high-dimensional, exploratory settings to simplify interpretations

- In particular, results emphasize the benefit of alternative penalties like SCAD or MCP
- These results support past findings that LASSO tends to include many false positive effects (Xiao & Xu, 2015) and bias estimates of non-zero coefficients, particularly for larger effects (Zhao & Yu, 2006)

Selection of DT-DLA as single most important predictor suggests converging findings in a research area marred by inconsistency

- Emphasizes the benefit of using large, representative cohorts and consortia, as well as a potential benefit of using dimensional (rather than dichotomous) outcomes

Back Up Slides

Variable Importance Indices

To determine which predictors were most influential in each model, the variable importance indices were calculated for each predictor, given for the j-th variable as

$$\pi_j = (\eta_j{}^\top \eta_*)/(\eta_*^\top \eta_*)$$

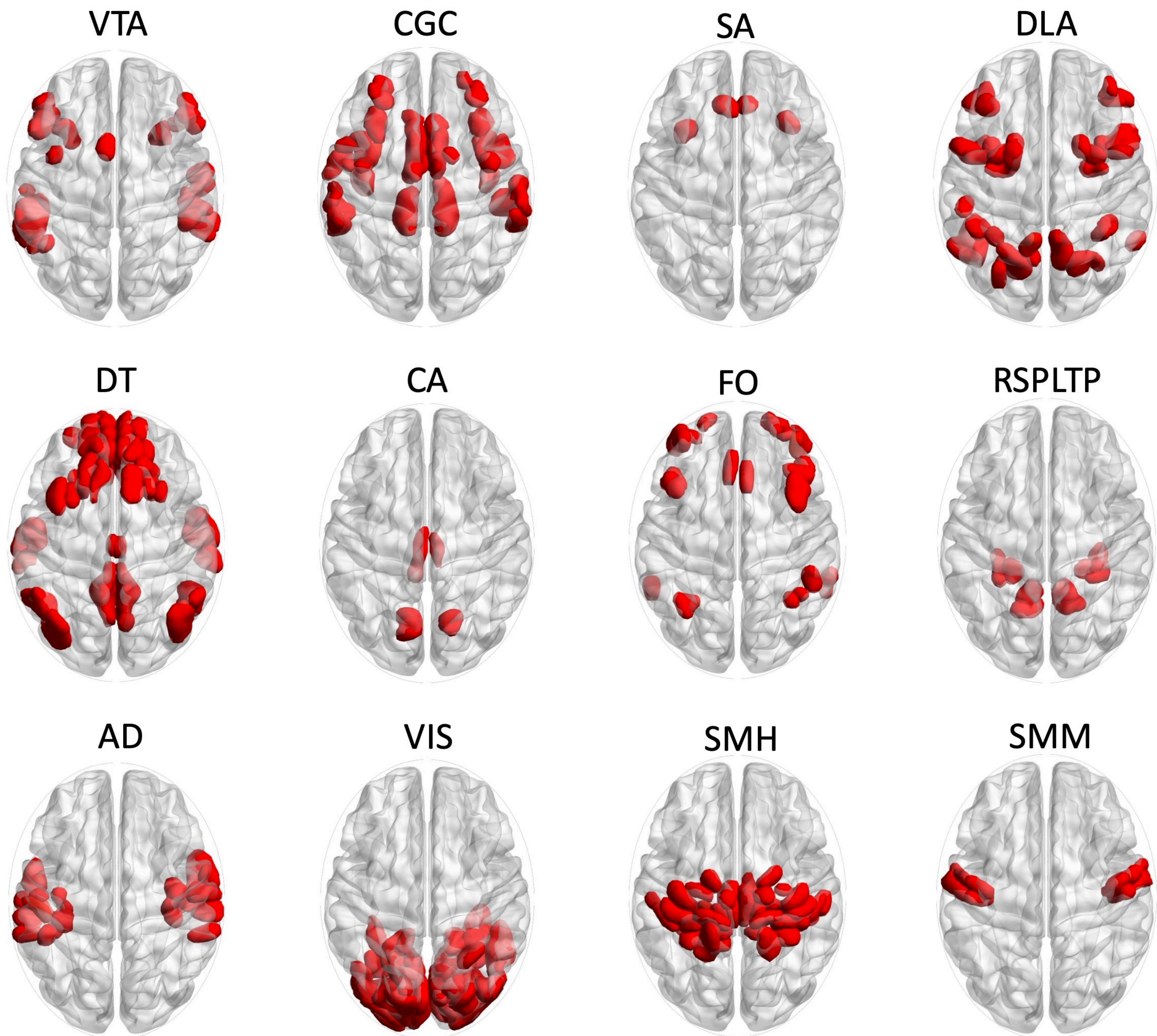
where η_j is the predicted effect for variable j evaluated at $\{x_i\}_{i=1}^n$, $\eta_* = \eta_1 + \eta_2 + \ldots + \eta_p$, and $\sum_{j=1}^p \pi_j = 1$ (Helwig et al., 2020)

Note that variable importance indices give the *approximate* percentage of the overall R^2 that is accounted for by each term

Variable importance indices sum to 100, so a value closer to 100 indicates greaeter importance

Distribution of CBCL Attention Problems (Outcome)

Brain Networks Under Study (Gordon Networks)



References

Clark, D. B., Moss, H. B., Kirisci, L., Mezzich, A. C., Miles, R., & Ott, P. (1997). Psychopathology in preadolescent sons of fathers with substance use disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(4), 495–502. https://doi.org/10.1097/00004583-199704000-00012

Cortese, S., Aoki, Y. Y., Itahashi, T., Castellanos, F. X., & Eickhoff, S. B. (2021). Systematic review and meta-analysis: Resting-state functional magnetic resonance imaging studies of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(1), 61–75. https://doi.org/10.1016/j.jaac.2020.08.014

Cubillo, A., Halari, R., Smith, A., Taylor, E., & Rubia, K. (2012). A review of fronto-striatal and fronto-cortical brain abnormalities in children and adults with attention deficit hyperactivity disorder (ADHD) and new evidence for dysfunction in adults with ADHD during motivation and attention. *Cortex*, *48*(2), 194–215. https://doi.org/10.1016/j.cortex.2011.04.007

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360. https://doi.org/10.1198/016214501753382273

Gao, Y., Shuai, D., Bu, X., Hu, X., Tang, S., Zhang, L., Li, H., Hu, X., Lu, L., Gong, Q., & Huang, X. (2019). Impairments of large-scale functional networks in attention-deficit/hyperactivity disorder: A meta-analysis of resting-state functional connectivity. *Psychological Medicine*, *49*(15), 2475–2485. https://doi.org/10.1017/S003329171900237X

Helwig, N. E. (2023). *Grpnet: Group elastic net regularized GLM*. https://CRAN.R-project.org/package=grpnet

Helwig, N. E., Atkinson, P., Delamont, S., Cernat, A., Sakshaug, J. W., & Williams, R. A. (2020). *Multiple and generalized nonparametric regression*. SAGE Publications. https://methods.sagepub.com/foundations/multiple-and-generalized-nonparametric-regression

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hoogman, M., Bralten, J., Hibar, D. P., Mennes, M., Zwiers, M. P., Schweren, L. S. J., Hulzen, K. J. E. van, Medland, S. E., Shumskaya, E., Jahanshad, N., & al., et. (2017). Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: A cross-sectional mega-analysis. *The Lancet Psychiatry*, *4*(4), 310–319. https://doi.org/10.1016/S2215-0366(17)30049-4

Hoogman, M., Muetzel, R., Guimaraes, J. P., Shumskaya, E., Mennes, M., Zwiers, M. P., Jahanshad, N., Sudre, G., Wolfers, T., Earl, E. A., & al., et. (2019). Brain imaging of the cortex in ADHD: A coordinated analysis of large-scale clinical and population-based samples. *American Journal of Psychiatry*, *176*(7), 531–542. https://doi.org/10.1176/appi.ajp.2019.18091033

Knopik, V. S., Sparrow, E. P., Madden, P. A. F., Bucholz, K. K., Hudziak, J. J., Reich, W., Slutske, W. S., Grant, J. D., McLaughlin, T. L., Todorov, A., & al., et. (2005). Contributions of parental alcoholism, prenatal substance exposure, and genetic transmission to child ADHD risk: A female twin study. *Psychological Medicine*, *35*(5), 625–635. https://doi.org/10.1017/S0033291704004155

Larsson, H., Chang, Z., D'Onofrio, B. M., & Lichtenstein, P. (2014). The heritability of clinically diagnosed attention deficit hyperactivity disorder across the lifespan. *Psychological Medicine*, *44*(10), 2223–2229. https://doi.org/10.1017/S0033291713002493

Mills, B. D., Miranda-Dominguez, O., Mills, K. L., Earl, E., Cordova, M., Painter, J., Karalunas, S. L., Nigg, J. T., & Fair, D. A. (2018). ADHD and attentional control: Impaired segregation of task positive and task negative brain networks. *Network Neuroscience*, *02*(02), 200–217. https://doi.org/10.1162/netn_a_00034

Norman, L. J., Sudre, G., Price, J., Shastri, G. G., & Shaw, P. (2023). Evidence from "big data" for the default-mode hypothesis of ADHD: A mega-analysis of multiple large samples. *Neuropsychopharmacology*, *48*(2), 281–289. https://doi.org/10.1038/s41386-