**Supplementary Information for “Longitudinal Characterization of Impulsivity Phenotypes Boosts Signal for Genomic Correlates and Heritability”**

Wei Q. Deng1,2\*, Kyla Belisario1,2, Marcus R. Munafò3,4, James MacKillop1,2\*

1. Peter Boris Centre for Addictions Research, St. Joseph’s Healthcare Hamilton, Hamilton, Ontario, Canada
2. Department of Psychiatry and Behavioural Neurosciences, McMaster University, Canada
3. School of Psychological Science, University of Bristol, Bristol, UK
4. MRC Integrative Epidemiology Unit, University of Bristol, Bristol. UK

\*Correspondence to:

**Wei Q. Deng, PhD**

Peter Boris Centre for Addictions Research

St. Joseph’s Healthcare Hamilton/McMaster University

Hamilton, Ontario L8P 3R2, Canada

Email: dengwq@mcmaster.ca

**James MacKillop, PhD, CPsych, FCAHS**

Peter Boris Centre for Addictions Research

St. Joseph’s Healthcare Hamilton/McMaster University

D171, 100 West 5th Street

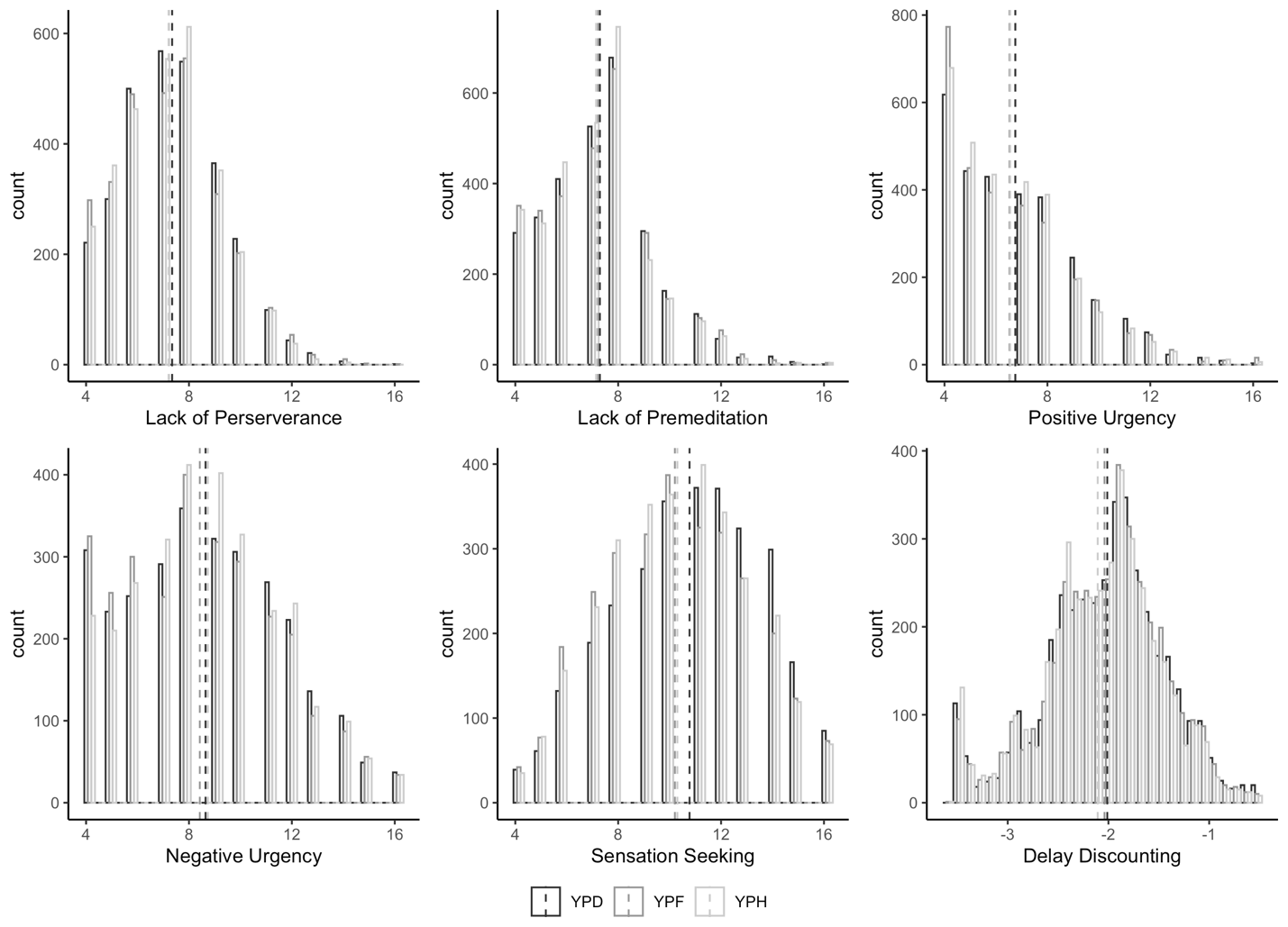
Hamilton, Ontario L8P 3R2, Canada

Email: jmackill@mcmaster.ca

Telephone: (+1) 905-522-1155, x39492

Fax: (+1) 905-575-608

**Supplementary Figures**

**Suppl. Figure 1. Distribution of the cross-sectional impulsivity measures over three waves.**

The histograms illustrate the distribution of cross-sectional impulsivity measures stratified by wave (YPD, YPF, and YPH). The x-axis represents values of each impulsivity measure, range from 4-16 for UPPS-P subscales and -4 to 0 for DD. The mean value was highlighted by a dotted line with color corresponding to each wave.

**Suppl. Figure 2. Scatterplots of the first two genetic principal components based on the combined ALSPAC samples and cosmopolitan samples from the 1000 Genomes Project.**

A screenshot of a graph showing different colored dots

Description automatically generatedPanel A shows a scatterplot of the first two principal components (PCs) for the combined study samples and 1000 Genomes Project samples, with different colors to indicate the subcontinental populations of origin. Panel B shows a scatterplot of the first two PCs for the combined samples, with different colors to indicate the continental origins of the samples. The study samples are shown in gray dots and only those overlapped with the European subset of the 1000 Genome Project samples were retained for the analysis.

**Suppl. Figure 3. A consort diagram depicting the sample and genotype data inclusion and exclusion criteria.**

A consort diagram that shows the inclusion and exclusion of sample and SNP data at each data processing stage.

**Suppl. Figure 4. Alluvial plot of missing data over time.**

A graph of different types of data

Description automatically generated with medium confidenceThe alluvial plots visualize the patterns and changes in missing data across A) delay discounting and B) UPPS-P subscales over three time points. The width of the streams represents the proportion of missing data, illustrating how these proportions evolve. Orange indicates data available and gray indicates missing data. The flow shows under which data combination strategy/wave that incompleteness issues was the most prevalent and how they shift through time (YPD at year 24, YPF at year 26, YPH at year 28).

**Suppl. Figure 5. Change in impulsivity over time.**

A screenshot of a graph

Description automatically generated

The plot shows the trajectory of impulsivity in ALSPAC (in Panel A) and PATH CANN (in Panel B) from overlapping period in time. The lines depict the progression of mean impulsivity measures across different waves. The x-axis shows the waves in chronological order, while the y-axis represents the impulsivity values for each wave. The vertical error bars representing the 95% confidence interval around the mean to convey variability.

**Suppl. Figure 6. Heatmap of phenotypic correlation of longitudinal impulsivity intercept in ALSPAC and PATH CANN.**

A diagram of a graph

Description automatically generated with medium confidenceHeatmaps display the correlation matrix between longitudinal impulsivity measures in ALSPAC (in Panel A) and PATH CANN (in Panel B). Each cell represents the correlation coefficient between the variables on the corresponding x and y axes, ranging from -1 (strong negative correlation, shown in blue) to +1 (strong positive correlation, shown in red). Cells with a correlation close to 0 are colored in neutral (white), indicating no correlation. The color intensity increases with the strength of the relationship.

**Suppl. Figure 7. Manhattan plot of genome-wide association results for cross-sectional and longitudinal Delay Discounting intercept.**

A graph of data on a white background

Description automatically generated with medium confidence

Manhattan plots for the cross-sectional and longitudinal interceptGWASs were depicted in four panels. The y-axis shows the -log10(p-value), with a solid horizontal line indicating the genome-wide significance threshold and the thin dashed line for the –log10(p-value) threshold at 5. The orange dots represent SNPs reaching genome-wide significance (*p* < 5×10-8) or SNPs that are within 2Mbp from the 28 genes in Table S2 and also with *p* < 0.005.

**Suppl. Figure 8. Manhattan plot of genome-wide association results for cross-sectional and longitudinal Negative Urgency intercept.**

A graph of data on a white background

Description automatically generated with medium confidence

Manhattan plots for the cross-sectional and longitudinal interceptGWASs were depicted in four panels. The y-axis shows the -log10(p-value), with a solid horizontal line indicating the genome-wide significance threshold and the thin dashed line for the –log10(p-value) threshold at 5. The orange dots represent SNPs reaching genome-wide significance (*p* < 5×10-8) or SNPs that are within 2Mbp from the 28 genes in Table S2 and also with *p* < 0.005.

**Suppl. Figure 9. Manhattan plot of genome-wide association results for cross-sectional and longitudinal Positive Urgency intercept.**

A graph of a number of data

Description automatically generated with medium confidence

Manhattan plots for the cross-sectional and longitudinal interceptGWASs were depicted in four panels. The y-axis shows the -log10(p-value), with a solid horizontal line indicating the genome-wide significance threshold and the thin dashed line for the –log10(p-value) threshold at 5. The orange dots represent SNPs reaching genome-wide significance (*p* < 5×10-8) or SNPs that are within 2Mbp from the 28 genes in Table S2 and also with *p* < 0.005.

**Suppl. Figure 10. Manhattan plot of genome-wide association results for cross-sectional and longitudinal Sensation Seeking intercept.**

A graph of a graph

Description automatically generated with medium confidence

Manhattan plots for the cross-sectional and longitudinal interceptGWASs were depicted in four panels. The y-axis shows the -log10(p-value), with a solid horizontal line indicating the genome-wide significance threshold and the thin dashed line for the –log10(p-value) threshold at 5. The orange dots represent SNPs reaching genome-wide significance (*p* < 5×10-8) or SNPs that are within 2Mbp from the 28 genes in Table S2 and also with *p* < 0.005.

**Suppl. Figure 11. Manhattan plot of genome-wide association results for cross-sectional and longitudinal Lack of Premeditation intercept.**

A graph of different sizes and colors

Description automatically generated with medium confidence

Manhattan plots for the cross-sectional and longitudinal interceptGWASs were depicted in four panels. The y-axis shows the -log10(p-value), with a solid horizontal line indicating the genome-wide significance threshold and the thin dashed line for the –log10(p-value) threshold at 5. The orange dots represent SNPs reaching genome-wide significance (*p* < 5×10-8) or SNPs that are within 2Mbp from the 28 genes in Table S2 and also with *p* < 0.005.

**Suppl. Figure 12. Manhattan plot of genome-wide association results for cross-sectional and longitudinal Lack of Perseverance intercept.**

A graph of data on a white background

Description automatically generated with medium confidence

Manhattan plots for the cross-sectional and longitudinal interceptGWASs were depicted in four panels. The y-axis shows the -log10(*p*-value), with a solid horizontal line indicating the genome-wide significance threshold and the thin dashed line for the –log10(p-value) threshold at 5. The orange dots represent SNPs reaching genome-wide significance (*p* < 5×10-8) or SNPs that are within 2Mbp from the 28 genes in Table S2 and also with *p* < 0.005.

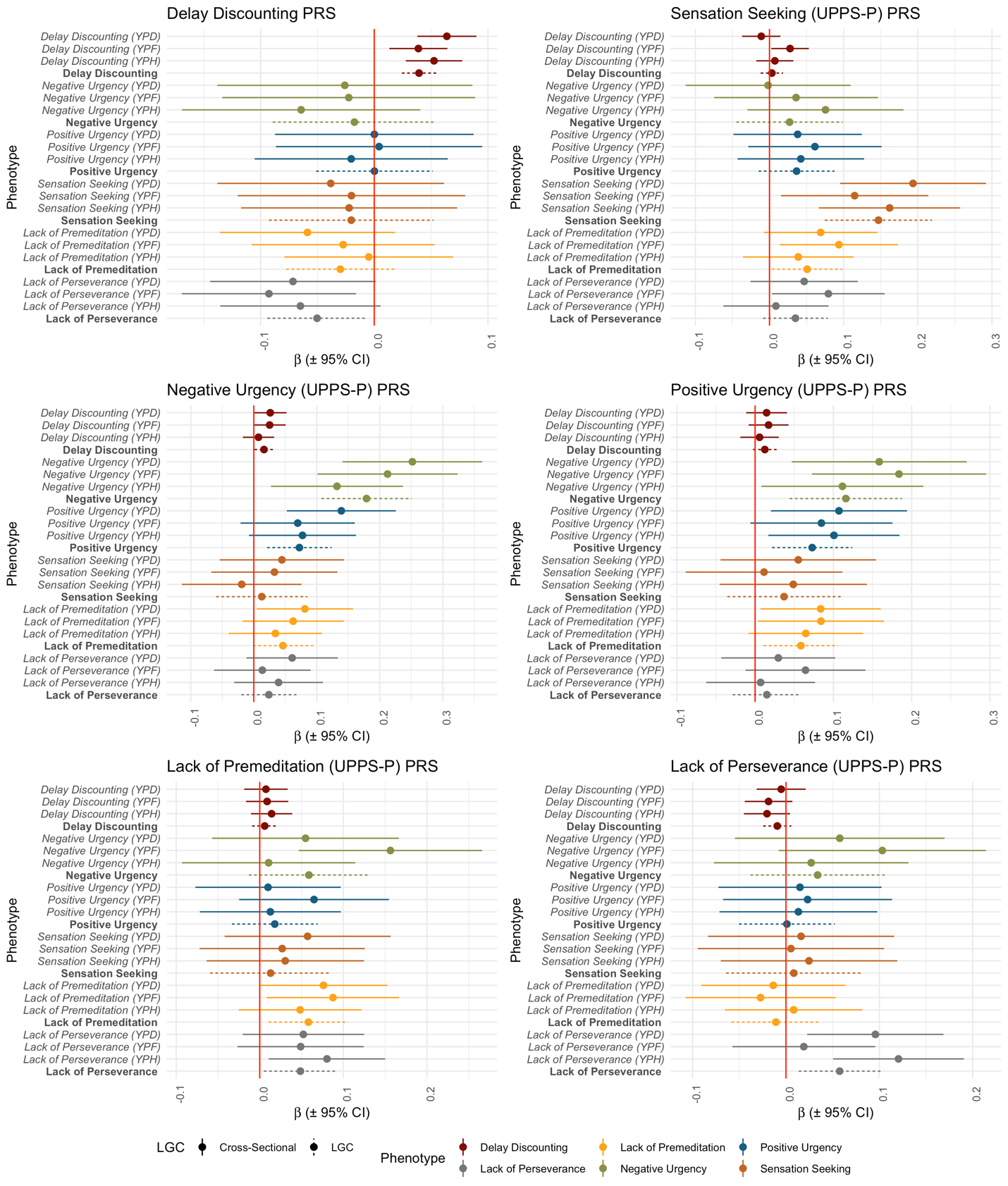
**Suppl. Figure 13. Summary of PRS association with cross-sectional and longitudinal impulsivity intercept phenotypes.**

A collage of graphs

Description automatically generated

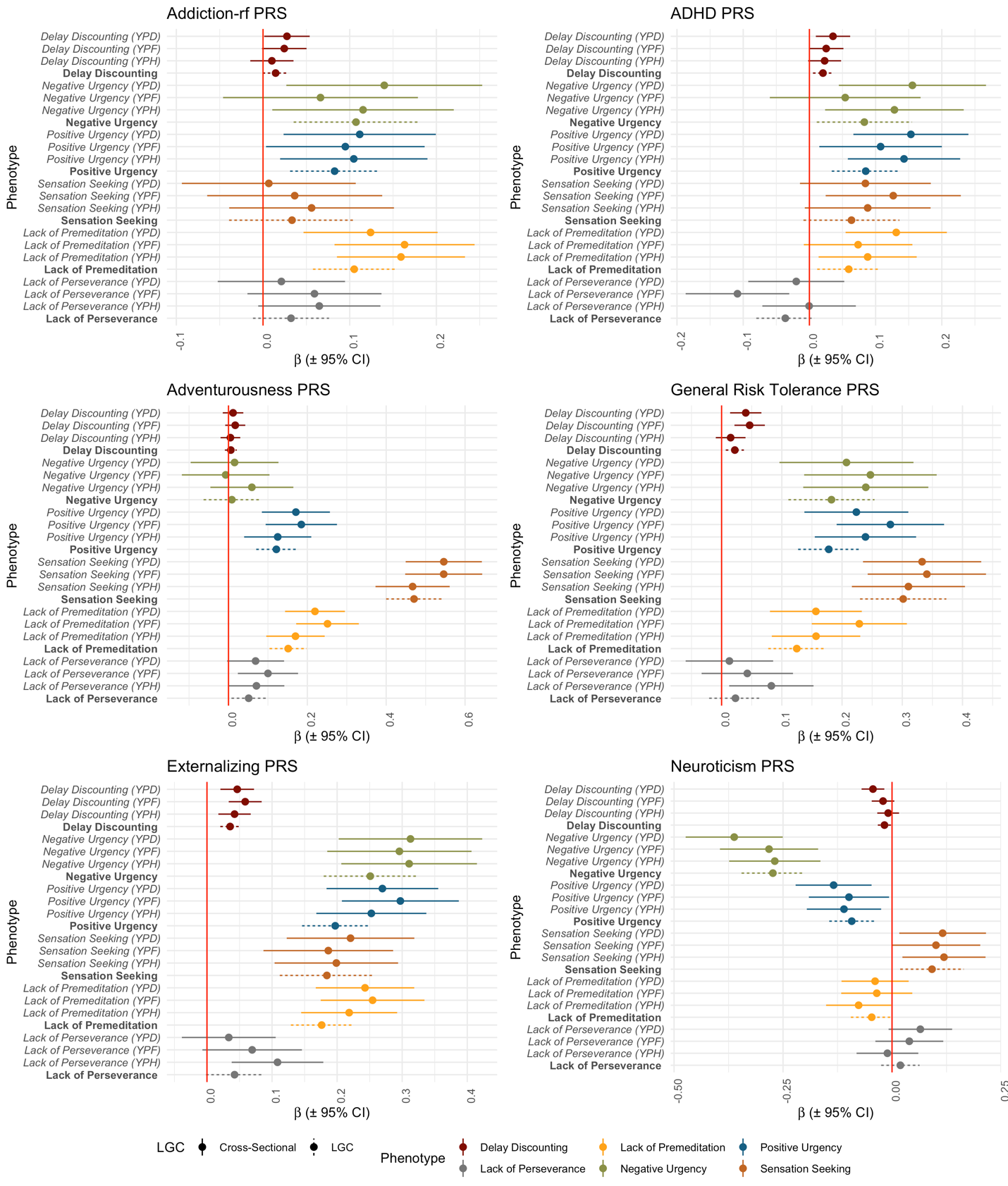
Each plot visualizes the association between 17 PRSs and cross-sectional (Panels A-C), ~~or~~ longitudinal intercept phenotypes (Panel D), or mean phenotype (Panel E). The x-axis categorizes phenotypes grouped into each impulsivity measure, while the y-axis represents the -log10 of the FDR adjusted *p*-value to indicate the strength of each association, with higher values denoting more significant associations. Points above the horizontal line of y = 3 reached a significance of *p*-value < 0.001. The number of total significant associations after FDR correction is shown in the brackets for each wave or the LGC intercept phenotypes.

**Suppl. Figure 14. Summary of association effects between impulsivity PRSs and impulsivity phenotypes.**



Each plot visualizes the association effects between an impulsivity PRS and all impulsivity measures in ALSPAC. The x-axis indicates the association regression coefficient, while the y-axis represents the cross-sectional and longitudinal impulsivity intercept phenotypes. Each dot represents the association regression coefficient and the lines represent the 95% confidence interval (CI). The solid lines are used to indicate 95% CI for cross-sectional impulsivity and the dashed line is reserved for the longitudinal impulsivity.

**Suppl. Figure 15. Summary of association effects between self-regulation PRSs and impulsivity phenotypes.**



Each plot visualizes the association effects between a self-regulation PRS and all impulsivity measures in ALSPAC. The x-axis indicates the association regression coefficient, while the y-axis represents the cross-sectional and longitudinal impulsivity intercept phenotypes. Each dot represents the association regression coefficient and the lines represent the 95% confidence interval (CI). The solid lines are used to indicate 95% CI for cross-sectional impulsivity and the dashed line is reserved for the longitudinal impulsivity.

**Suppl. Figure 16. Summary of association effects between substance use PRSs and impulsivity phenotypes.**

A graph of various types of alcohol

Description automatically generated with medium confidence

Each plot visualizes the association effects between a substance use PRS and all impulsivity measures in ALSPAC. The x-axis indicates the association regression coefficient, while the y-axis represents the cross-sectional and longitudinal impulsivity intercept phenotypes. Each dot represents the association regression coefficient and the lines represent the 95% confidence interval (CI). The solid lines are used to indicate 95% CI for cross-sectional impulsivity and the dashed line is reserved for the longitudinal impulsivity.

**Suppl. Figure 17. A summary of within-trait PRS associations with cross-sectional trajectory and longitudinal impulsivity intercept in ALSPAC and PATH CANN.**

The x-axis shows the time of collection for impulsivity phenotypes and the y-axis shows the adjusted R2 estimates from cross-sectional measurements and the LGC phenotype (in red). These plots illustrated the strength of association between PRSs of impulsivity, the corresponding LGC phenotype and cross-sectional phenotype measured at multiple time points. Each point estimate is marked by a dot with a vertical bar representing the standard error around the estimate. The fitted line was obtained using Generalized Additive Models (GAM) with smoothness determined by 3 degrees of freedom. The shaded area around the fitted line represents the 95% confidence interval, estimated using bootstrap sampling techniques. This visualization helps in understanding the dynamic nature of the cross-sectional trait as a function of attainable genetic predispositions over time.

A collage of graphs

Description automatically generated

**Suppl. Figure 18. Distribution of the estimated heritability across simulation runs.**



The x-axis shows the phenotype for which the heritability estimates were generated and the number of phenotypes corresponded to the number of waves (labelled by Y1, Y2, etc.) plus the longitudinal phenotype (labelled as LGC) and the mean phenotype (labelled as mean). The y-axis shows the estimated heritability with the red line indicating the theoretical heritability and the black line the upper bound of a heritability estimate under a cross-sectional design. The 3-by-4 grid corresponds to the number of scenarios considered by varying the number of time points (3, 5, 10) and the sample size (*n*=2500, 5000, 10000, 20000).