**A Comprehensive Assessment of Neurocognitive and Psychological Functioning in Adults with Early-Treated Phenylketonuria (PKU)**

**Supplementary Materials**

**Supplementary Methods – Calculation of Index of Dietary Control (IDC)**

As described previously (Christ et al., 2021), the index of dietary control (IDC) represents the mean of all half-year median phe levels. IDC values were computed for early childhood (0-5 yrs), middle childhood (6-11 yrs), adolescence (12-17 yrs), adulthood (18+ yrs), and lifetime (0-present). To explain further: A participant’s median phe level is calculated separately for each year of life. The early childhood IDC represents the average of the five median phe levels associated with the first five years of life. The other IDC values are calculated in similar fashion.

In case of small gaps in records (e.g., a person is missing phe levels for their 4th year of life), values were extrapolated using linear regression and phe levels of adjacent years for purposes of IDC calculation. In rare cases of larger gaps, the participant was excluded from analysis of the relevant developmental epoch. For example, if a participant was missing phe levels for 7-9 yrs of age, then they would not be included in analysis of middle childhood (6-11 yrs) or lifetime IDC values.

**Supplementary Methods – Detailed Task Descriptions:**

**Wechsler Abbreviated Scale of Intelligence (WASI-2).** Overall intellectual ability was estimated with the four-subtest version of the WASI-2 (Wechsler, 1999).

**NIH Toolbox.**Ten subtests from the NIH Toolbox were included: (a) Dimensional Change Card Sort - a measure of executive function where participants shift between matching images based on shape or color; (b) Flanker Inhibitory Control and Attention - a measure of attention where participants indicate the direction of a central stimulus while ignoring distractors; (c) List Sorting Working Memory - a measure of working memory where participants recall a list of items in size order from smallest to largest; (d) Pattern Comparison - a measure of processing speed where participants indicate is two pictures are the same or different; (e) Oral Symbol Digit - a measure of processing speed where participants match numbers with symbols following a key; (f) 9-hole pegboard - a measure of fine motor skills where participants place and remove pegs into a pegboard as quickly as possible with one hand at a time; (g) Grip Strength - a strength measure where participants grip a dynamometer as tightly as possible with each hand; (h) Standing Balance - a measure of balance where participants hold a series of poses on solid and soft surfaces; (i) Walk Endurance Test - a measure of endurance where participants walk as far as possible in two minutes; and (j) Walk Gait Speed Test – a measure of locomotion where participants walk four meters to assess their typical walking speed.

In addition to generating individual subtest scores, the Toolbox also yields a Fluid Cognition Composite Score based on the Dimensional Change Card Sort, Flanker Inhibitory Control and Attention, Picture Sequence Memory, List Sorting Working Memory, and Pattern Comparison subtests. The NIH Toolbox was normed with data from a sample of over 4,800 individuals representative of the United States population, and as discussed previously, the validity and reliability of the measure has been established in the general population as well as several clinical populations (Heaton et al., 2014; Shields et al., 2020; Tulsky et al., 2017).

**Select Subtests from the Wechsler Adult Intelligence Scale (WAIS -IV).**Three subtests from the WAIS-IV were included: (a) Coding - a measure of visuomotor speed involving the transcription of digit-symbol pairs; (b) Symbol Search - a measure of visual scanning and processing speed requiring participants to respond as quickly as possible whether a target stimulus matches an array of sample stimuli; (c) Visual Puzzles - a measure of spatial reasoning skills in which participants view a completed puzzle and must select three pieces that together would reconstruct the puzzle (Wechsler, 2008).

**Connors’ Continuous Performance Test (CPT-3).**CPT-3 is a computerized measure of sustained attention, vigilance, and impulsivity. Participants are presented a continuous series of letter stimuli. They are instructed to press a button in response to every letter stimuli and must inhibit their response when an ‘X’ is presented. Participants complete 360 trials over 14 minutes. The CPT-3 was normed on a sample of 1,400 individuals representative of the United States population and has been shown to have good reliability and validity (Conners, 2014). Specific scores from the CPT-3 used include Hit Reaction Time as a metric of processing speed, and Omissions and Commissions as measures of executive function.

**Behavior Rating Inventory of Executive Function - Adult Version (BRIEF-A)**. The BRIEF-A is a standardized self-report questionnaire designed to assess executive function within the context of an individual’s day-to-day environment. It consists of 75 items assessing behavioral manifestations of executive problems in daily life. Scores on individual items contribute to the calculation of nine non-overlapping clinical scales reflecting different aspects of EF, including inhibitory control, self-monitoring ability, planning and organizational skill, emotional control, and working memory. In addition, two broader indices (BRI and MI) as well as a composite score reflecting overall executive ability (Global Executive Composite [GEC]) can be computed. The BRIEF-A has shown good reliability and validity in adult samples (Roth et al., 2005).

**Connors’ Adult ADHD Rating Scales (CAARS).** The CAARS is a standardized self-report measure designed to assess attention within the context of an individual’s day-to-day environment. It consists of 66 items assessing behavioral manifestations of attention problems in daily life. Scores on individual items contribute to the calculation of nine clinical subscales, including symptoms of inattention, hyperactivity and impulsivity, problems with self-concept, and an overall ADHD index. An inconsistency scale is also included to capture random response styles. The CAARS has shown good reliability and validity (Conners et al., 1999).

**Social Responsiveness Scale - II (SRS - II)**. The SRS-II is a standardized report measure designed to assess impairments in social communication associated with autism spectrum disorders. It consists of 65 items that assess an individual’s engagement in reciprocal social interactions, understanding of emotional and social cues, and motivation to engage with others. The self-report version of the SRS-II was utilized in the present study. Scores on individual items contribute to a total score along with five subscales including social awareness, social cognition, social communication, social motivation and autistic mannerisms. The SRS-II has been shown to have good reliability and validity (Constantino & Gruber, 2012).

**Beck Anxiety Inventory (BAI).** The BAI is a self-report questionnaire designed to assess the prevalence and impact of symptoms of anxiety. It consists of 21 items that assess the presence of symptoms common among individuals experiencing anxiety. A total score derived from the individual items can then be categorized via score cut-offs as reflecting mild, moderate, or severe anxiety. The BAI has been shown to have high reliability and good validity among adult samples (Beck et al., 1988).

**Beck Depression Inventory II (BDI-2)**. The BDI-2 is a self-report questionnaire designed to assess depressive symptomatology. It consists of 21 items assessing the presence of symptoms common among those experiencing a depressive episode. A total score derived from the individual items can then be categorized via score cut-offs as reflecting minimal, mild, moderate, or severe depression. The BDI-2 has been shown to have high reliability and validity in adult samples (Beck et al., 1996).

**FIGURE S1**

Text

Description automatically generatedChart

Description automatically generated

*Figure S1.* Correlations between cognitive composite scores and blood phe levels during different developmental epochs.

\*p<.05 FDR corrected; \*\*p<.005 FDR corrected.

**FIGURE S2**

Chart, bar chart

Description automatically generatedText

Description automatically generated

*Figure S2.* Correlations between self-report measures and blood phe levels during different developmental epochs.

\*p<.05 FDR corrected; \*\*p<.005 FDR corrected.

**FIGURE S3**

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

*Figure S3.* Scatter plot graphs illustrating the relationship between cognitive composite scores and most recent blood phe levels. Lower composite scores are indicative of worse functioning. The dashed reference line in each scatterplot represents the average score/performance for the non-PKU group. Stated *p* values are corrected for FDR.

**FIGURE S4**

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

*Figure S4*. Scatter plot graphs illustrating the relationship between report measure scores and most recent blood phe levels. Higher t scores are indicative of worse functioning. The dashed reference line in each scatterplot represents the average score/performance for the non-PKU group. Stated *p* values are corrected for FDR.

**FIGURE S5**



*Figure S5*. Bar graph illustrating the mean values of all CAARS subscales. Asterisks indicate significant group differences.

\*p<.05; \*\*p<.005