A lifetime with Phenylketonuria
Jahja, Rianne

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Cognitive profile and mental health in adult Phenylketonuria: A PKU-COBESO study

Rianne Jahja¹, Stephan C. J. Huijbregts², Leo M. J. de Sonneville², Jaap J. van der Meere³, Amanda M. Legemaat³, Annet M. Bosch⁴, Carla E. M. Hollak⁴, M. Estela Rubio-Gozalbo⁵, Martijn C. G. J. Brouwers⁶, Floris C. Hofstede⁶, Maaike C. de Vries⁸, Mirian C. H. Janssen⁸, Ans T. van der Ploeg⁹, Janneke G. Langendonk⁹, Francjan J. van Spronsen¹

¹ University of Groningen, University Medical Center Groningen, Beatrix Children’s Hospital, Groningen
² Leiden University, Department of Clinical Child and Adolescent Studies & Leiden Institute for Brain and Cognition, Leiden
³ University of Groningen, Department of Developmental and Clinical Neuropsychology, Groningen
⁴ Academic Medical Center, Amsterdam
⁵ University Hospital Maastricht and Laboratory Genetics Metabolic Diseases, Maastricht
⁶ University Hospital Maastricht, Department of Internal Medicine, Division of Endocrinology and Metabolic Diseases, Maastricht
⁷ Wilhelmina Children’s Hospital, University Medical Center Utrecht, Utrecht
⁸ University Medical Center St. Radboud Nijmegen, Nijmegen
⁹ Erasmus Medical Center, Rotterdam

Neuropsychology 2017; 31(4): 437-447
**ABSTRACT**

**Objective.** Despite early dietary treatment phenylketonuria patients have lower IQ and poorer executive functions compared to healthy controls. Cognitive problems in phenylketonuria have often been associated with phenylalanine levels. The present study examined the cognitive profile and mental health in adult phenylketonuria, in relation to phenylalanine levels and tetrahydrobiopterin treatment.

**Method.** Fifty-seven early treated adult patients with phenylketonuria and 57 healthy matched controls (18-40 years) performed IQ-subtests and executive function-tests from the Amsterdam Neuropsychological Tasks. They also completed the Adult Self-Report on mental health problems. Analyses of variance were performed to examine group differences.

**Results.** Patients with phenylketonuria had normal IQ’s although lower than controls. They performed poorer on working memory, inhibitory control, and sustained attention-tasks. Patients reported Depressive and Avoidant Personality problems more frequently. Specifically, patients with childhood and lifetime phenylalanine≥360 µmol/L had poorer cognitive and mental health outcomes than controls. In a subset of patients, comparisons between patients on and off tetrahydrobiopterin showed that non-tetrahydrobiopterin users (matched for childhood, pre-treatment phenylalanine) were slower (on number of tasks) and reported more mental health problems.

**Conclusions.** Adult patients had lower IQ and poorer executive functions than controls, resembling problems observed in younger patients with phenylketonuria, as well as more internalizing problems. Group differences and phenylalanine-outcome associations were smaller than those observed in younger populations. A subset of non-tetrahydrobiopterin users, matched for childhood phenylalanine level, had poorer outcome on some tests than tetrahydrobiopterin users, which might indicate an impact of tetrahydrobiopterin treatment beyond lowering phenylalanine. However, clinical relevance needs further investigation.
INTRODUCTION

Phenylketonuria (PKU; OMIM 261600) is an inborn error of phenylalanine (Phe) metabolism, characterized by an impaired conversion of Phe into tyrosine (Tyr). Left untreated, PKU leads to severe neurocognitive dysfunction, seizures, and behavioural problems. When diagnosed by neonatal screening and treated early and continuously, at least throughout childhood, patients develop and function (cognitively) within the lower normal range (Jahja, Huijbregts, de Sonneville, van der Meere, & van Spronsen, 2014; Smith & Knowles, 2000). On average, the intelligence quotient (IQ) of treated patients with PKU is approximately 8-10 points lower than that of their healthy counterparts (Waisbren et al., 2007). The cognitive profile of treated PKU is characterized by weaknesses in motor control and executive functioning, which are more pronounced in combination with sustained attention (Albrecht, Garbade, & Burgard, 2009; Christ, Huijbregts, de Sonneville, & White, 2010; Huijbregts et al., 2003; Leuzzi et al., 2004). Extensive neuropsychological testing has particularly taken place among children and young adolescents with PKU.

In addition to cognitive difficulties, internalizing behaviour problems have frequently been reported. Findings indicated that patients with PKU experience problems such as depression, mood swings, anxiety, phobic reactions, and poor self-image (Anjema et al., 2011; Arnold et al., 1998; Jahja et al., 2013; Smith & Knowles, 2000; Weglage et al., 2000). A recent study has also found poorer social (cognitive) skills and functioning in adult patients with PKU compared to controls. Poorer social functioning was related to higher lifetime Phe (Jahja et al., 2016).

Treatment of PKU is based on blood Phe reduction by a Phe-restricted diet, Large Neutral Amino Acid (LNAA) supplementation, and/or pharmacological treatment with tetrahydrobiopterin (BH$_4$) (Blau, van Spronsen, & Levy, 2010). BH$_4$ is a co-factor necessary for the phenylalanine hydroxylase (PAH) enzyme to convert Phe into Tyr. When there is an absence or loss of PAH activity, Phe concentrations increase to toxic levels for the brain. In approximately 20 to 50% of patients with PKU, BH$_4$ supplementation increases Phe tolerance, probably by enzyme stabilization. Generally those who are BH$_4$ responsive already have milder forms of PKU, caused by different types of gene mutations (Keil et al., 2013). By means of genotyping but also with a BH$_4$-loading test, it can be determined whether patients with PKU are BH$_4$ responsive (Anjema et al., 2013). The defined requirement for BH$_4$ responsiveness is a reduction of Phe of at least 30% during 48 hours of BH$_4$ use. Patients with a reduction in Phe below 20% are considered non-responders, while a reduction of more than 85% indicates BH$_4$ deficiency (Anjema et al., 2013; Blau et al., 2010). In BH$_4$ responsive patients, it has been shown that administration of BH$_4$ results in an increase of PAH activity and a subsequent decrease of Phe levels.
Further benefits include an increase in Phe tolerance, which has been associated with improved quality of life, as adherence to the very strict diet becomes less important (Keil et al., 2013). There are some indications that BH₄ is also beneficial for cognitive functioning (Burton et al., 2015; Christ, Moffitt, Peck, & White, 2013). With respect to the mechanisms that might underlie improvement of cognition, preliminary data showed that BH₄ decreased prolactin concentrations. As dopamine inhibits prolactin production, this result was hypothesized to indicate increased dopamine availability in the brain (van Vliet et al., 2015), which, in turn, might improve cognitive functioning. There is also evidence that BH₄ improves white matter integrity, but results regarding improvement of executive abilities are inconsistent. Whereas some researchers did not find such improvements (White et al., 2013), others did (Christ et al., 2013). Results from Christ et al. (2013) suggested that BH₄ treatment leads to better working memory and changes in brain activation. Based on the available evidence, however, it is too early to conclude that BH₄ improves cognition and mental health, and particularly, it is still unclear through which mechanisms it might do so.

It has convincingly been shown that the quality of cognitive function is negatively related to plasma Phe concentrations, again especially among children and young adolescents with PKU (Christ et al., 2010). There is also evidence indicating that Phe levels during childhood are more strongly related to cognitive functioning later in life than Phe levels at the day of neuropsychological testing, i.e. concurrent Phe levels (Huijbregts et al., 2002). One of the major, still outstanding questions in PKU concerns the upper target blood Phe concentrations to achieve the best outcome. The target Phe concentrations may be different for different ages, although it has also been claimed that target Phe concentrations should be the same for the whole lifespan (Camp et al., 2014). In most countries the recommended target range for plasma Phe levels, which are considered a surrogate marker of brain Phe levels, is 120 to 360 µmol/L during childhood (i.e. ≤12 years), with more lenient upper target levels thereafter, ranging from 600 to 1200 µmol/L (MRC, 1993).

This study aimed to investigate the cognitive profile and mental health of early diagnosed and treated adults with PKU compared to healthy controls. The focus of investigations into cognitive functioning was on cognitive domains shown to be deficient in early and continuously treated children and adolescents with PKU, i.e. IQ and executive functioning. Based on previous studies more internalizing mental health problems were expected. Cognitive and mental health outcomes were associated with Phe concentrations (concurrent, lifetime, and specifically in childhood). We also investigated whether, beyond lowering Phe levels, BH₄ influenced cognitive functioning and mental health.
METHODS

Participants
The PKU group consisted of 57 patients (24 males, 33 females, mean age 27.7 years ± 6.0, range 18.7-40.0 years), all were diagnosed by neonatal screening and treated (through dietary protein restriction and amino acid supplements) immediately upon diagnosis. Neonatal screening started in the Netherlands in 1974. All patients paid regular follow-up visits to a Dutch university medical centre and were treated according to National Institutes of Health guidelines (National Institutes of Health Consensus Development Panel, 2001). BH4 and non-BH4 users were both included in this study. Twelve BH4 responsive patients used BH4 doses up to 20 mg/kg with a max of 1400 mg/day. A healthy control group was recruited via patients or through the researchers and included 57 participants (19 males, 38 females, mean age 25.7 years ± 5.6, range 18.1-40.8 years). Four controls were family related, i.e. siblings or cousins, six controls were from the direct surrounding of the patient (friend, classmate, or partner). The PKU and control group did not differ significantly in age (t(112) = -1.9, p = .062), gender distribution (χ² = 0.9, p = .338), education (χ² = 4.3, p = .229), and yearly income (χ² = 9.8, p = .080).

On the day of testing, blood samples were taken from the patients preprandially in order to obtain concurrent Phe levels. Historical Phe levels were provided by the patients’ physicians. Lifetime Phe was computed as the mean of all half-year median Phe levels, from birth until the day of testing. In PKU literature, it is custom to use the mean of half-year median Phe levels to represent lifetime Phe. During certain periods of life (e.g. childhood) many more samples (i.e. blood Phe determinations) are taken than in other periods of life (e.g. adolescence), which could bias the mean. Therefore the mean of half-year medians is used. Also in the same manner, Phe levels between 0 and 12 years (i.e. childhood Phe) were calculated to improve understanding of the effect of Phe during this developmental stage.

The Medical Ethics Committee of the University Medical Centre Groningen provided national approval for the study, which is registered in the Dutch CCMO Register (NL38932.042.11). All participants and/or their parents signed an informed consent.

Instruments
This study is part of a larger Dutch multicentre study (PKU-COBESO: Jahja et al., 2013) focusing on cognitive, social, and behavioural sequelae of early and continuously treated patients with PKU in relation to (history of) metabolic control. A standardized testing protocol was used (Jahja et al., 2013), which took approximately 2.5 hours to complete.
The Block Design and Vocabulary subtests of the Wechsler Adult Intelligence Scale 3rd Edition (Wechsler, 1997) were used to estimate IQ. Executive functions (EF) were assessed using tasks from the Amsterdam Neuropsychological Tasks (ANT: De Sonneville, 2014). The ANT is a standardized test battery, of which the reliability and validity has been described extensively elsewhere (Huijbregts et al., 2002; Rowbotham, Pit-ten Cate, Sonuga-Barke, & Huijbregts, 2009). The selected tasks address two core executive functions, working memory (WM) and inhibitory control, with one selected task (mainly measuring inhibitory control) also measuring sustained attention. In one task the overall number of correct responses was used, whereas number of errors and reaction time (RT in ms) were measured in other tasks. All tasks started with practice trials in order to determine whether verbal task instructions were understood. Participants were instructed to respond as quickly and as accurately as possible.

Three tasks were included to measure working memory: the Visuo-Spatial Sequencing task (VSS), Memory Search 2 Dimensions task (MS2D), and Feature Integration task (FI). These working memory tasks have multiple levels of complexity in different task parts. In the VSS, a fixed constellation of nine open circles is presented on the screen. In each trial (n = 24) between three and seven circles are highlighted briefly. Participants must remember which circles were highlighted as well as the order in which they were highlighted, and they respond in each trial by clicking the remembered locations in the presented order. Quality of WM is represented by the number of correct responses, with a further contrast between number of correctly identified targets (i.e. visuospatial WM) and number correct targets in the correct order (i.e. visuospatial and visuotemporal WM), reflecting higher-order WM skills.

The MS2D task consists of two task parts. In the first part (48 trials, with 24 containing the target stimulus) participants have to identify one memorized target stimulus with two specific features (shape and colour) out of four displayed on the screen (e.g., a blue triangle), whereas in the second part (48 trials, with 24 containing a target stimulus), three different target stimuli had to be remembered (e.g., a blue square, a red circle, and a yellow triangle), of which one had to be present in the display for a yes-response (see also Rowbotham et al., 2009). A ‘yes’ response is given by pressing the mouse button under the index finger of the dominant hand, otherwise the ‘no’ button (under the index finger of the non-dominant hand) is pressed. The second task part thus has the higher WM demands. Number of errors and RT of correct responses were recorded.

In the FI task a fixed target pattern, a 3x3 matrix with three red-coloured and six white-coloured squares, has to be identified out of four such matrices/stimulus patterns displayed on the screen (80 trials, with 40 containing the target stimulus). The distracter stimuli (three when the target pattern was present, four when the target

62
pattern was absent) either strongly resemble the target stimulus (i.e. the “similar” condition) or were clearly different (the “dissimilar” condition) (Rowbotham et al., 2009). The similar condition requires top-down controlled search posing higher WM demands than the dissimilar condition. A ‘yes’ response, again consisting of a press of the mouse button under the index finger of the dominant hand, is required when the target pattern is present, a ‘no’ response when the target is absent. Again, number of errors and RT of correct responses were recorded and used in statistical analyses.

The Sustained Attention-Dots task (SAD) measures inhibitory control and sustained attention, and took about 10-12 minutes to complete (Huijbregts et al., 2002; Leuzzi et al., 2004). It includes 600 trials, where dots are presented in random configurations on the computer screen: 200 trials with three dots, 200 with four dots and 200 with five dots. When four dots appear on the screen, participants press the ‘yes’ button under the index finger of the dominant hand. When three or five dots are presented, the ‘no’ button under the index finger of the non-dominant hand must be pressed. When four dots appear, inhibitory control of the prepotent response is required, particularly when the task progresses, as the other response is required twice as often and progressively becomes the automatic or prepotent response. Auditory feedback (a beep) is given when an error is made. Inhibitory control was measured by the overall number of errors and Mean Series Time (MST, i.e. the mean RT of 50 series of 12 trials, each with three 3-, 4-, and 5-dot trials). A comparison was made regarding error rate and response speed (MST) during the first and last 120 trials in order to examine performance over the course of the task, hence measuring sustained attention.

All tasks are self-paced. The post-response interval, i.e. the period between the response and next stimulus onset, is 250 ms in the SAD task and 1200 ms in the FI and MS2D tasks. The valid response window, i.e. reaction times that are considered valid, was set at 200-6000 ms post-stimulus onset.

Mental health problems were measured by means of the Adult Self-Report (ASR) of the Achenbach System of Empirically Based Assessment (Achenbach & Rescorla, 2003). The ASR is a norm-referenced questionnaire consisting of 102 items (with 3-point rating scales) suitable for adults. The ASR provides six DSM-IV-oriented scales: Depressive problems, Anxiety problems, Somatic problems, Avoidant Personality problems, Attention Deficit/Hyperactivity problems, and Antisocial Personality problems. Higher scores represent more mental health problems.

Data analyses
IBM SPSS Statistics 22nd version was used to conduct analyses. For comparisons between patients with PKU and healthy controls analyses of variance were used. Univariate analyses were used to compare IQ estimates, overall response speed in the
VSS working memory and SAD inhibition tasks, and error rate in the SAD inhibition task. Multivariate analyses were used to compare levels of mental health problems. Repeated measures analyses of variance were used to compare performance between patients with PKU and healthy controls across conditions in the VSS, MS2D, and FI working memory tasks and the SAD sustained attention. In the VSS, the within-subjects factor was represented by “number correctly identified targets” versus “number correctly identified targets in the correct order”; in the MS2D this was represented by WM load, i.e. remembering one stimulus with two distinctive features in task part 1 versus remembering three stimuli with two distinctive features each in task part 2; in the FI the within-subjects factor was represented by WM demand, i.e. stimuli with dissimilar distracters versus stimuli with similar distracters; and in the SAD sustained attention the within-subjects factor was represented by inhibitory control at the beginning versus inhibitory control at the end of the task. Analyses were conducted with the whole PKU group and repeated without BH4 users.

Associations between Phe levels (concurrent, lifetime, and between 0 and 12 years of age) and the cognitive and mental health outcome variables were first examined using Pearson one-tailed correlations. Furthermore, groups were formed, among the patients with PKU, according to the most commonly recommended upper target limits of 360 µmol/L for lifetime, concurrent, and childhood Phe and were compared to each other and controls using analyses of variance. In order to provide more insight into the effects of BH4, BH4 users and non-users, matched on (generally low) Phe levels during childhood and on age, were compared on cognitive profile and mental health. These two groups were compared to each other using the same analyses of variance as described above.

RESULTS

Descriptive data

Concurrent, lifetime, and childhood Phe concentrations for the complete PKU group, for the BH4 and non-BH4 group are displayed in Table 1. Concurrent and lifetime Phe were significantly correlated ($r = 0.57$, $p < .001$).

Twelve out of 57 patients (21%) were BH4 responsive. The BH4 group (mean age 23.9 years ± 4.8) was significantly younger than the non-BH4 group (mean age 28.8 years ± 5.9): $t(55) = 2.6$, $p = .011$. Concurrent Phe ($t(31) = 2.6$, $p = .013$) and lifetime Phe ($t(34) = 4.2$, $p < .001$) were lower for the BH4 group (resp. 489 µmol/L ± 206 for concurrent Phe and 348 µmol/L ± 92 for lifetime Phe) than for the non-BH4 group (resp. 699 µmol/L ± 359 and 505 µmol/L ± 174). Phe concentrations in childhood
(0-12 years) did not significantly differ (BH\(_4\): 302 µmol/L ± 134; nonBH\(_4\): 372 µmol/L ± 151).

Table 2 shows mean scores and standard deviations (SD) on the tasks and questionnaire for the PKU and control group. An overview of all statistical analyses is given in Table 3.

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics for PKU and control group</th>
<th>PKU (n = 57)</th>
<th>PKU nonBH(_4) (n = 45)</th>
<th>PKU BH(_4) (n = 12)</th>
<th>Controls (n = 57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (male:female)</td>
<td>24:33</td>
<td>20:25</td>
<td>4:8</td>
<td>19:38</td>
</tr>
<tr>
<td>Mean age ± SD (range)</td>
<td>27.7 ± 6.0</td>
<td>28.8 ± 5.9</td>
<td>23.9 ± 4.8</td>
<td>25.7 ± 5.6</td>
</tr>
<tr>
<td></td>
<td>(18.7 - 40.0)</td>
<td>(19.7 - 40.0)</td>
<td>(18.7 - 32.7)</td>
<td>(18.1 - 40.8)</td>
</tr>
<tr>
<td>Mean Concurrent Phe ± SD (range)</td>
<td>655 ± 342</td>
<td>699 ± 359</td>
<td>489 ± 206</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(66 - 1550)</td>
<td>(66 - 1550)</td>
<td>(220 - 1030)</td>
<td></td>
</tr>
<tr>
<td>Mean Lifetime Phe ± SD (range)*</td>
<td>472 ± 172</td>
<td>505 ± 174</td>
<td>348 ± 92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(223 - 1001)</td>
<td>(253 - 1001)</td>
<td>(223 - 527)</td>
<td></td>
</tr>
<tr>
<td>Mean Childhood Phe 0-12y ± SD (range)*</td>
<td>357 ± 150</td>
<td>372 ± 151</td>
<td>302 ± 134</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(181 - 903)</td>
<td>(181 - 903)</td>
<td>(181 - 689)</td>
<td></td>
</tr>
</tbody>
</table>

* n = 56

**PKU versus healthy controls: IQ, EF, and mental health**

**IQ.** Although still within the normal range, patients with PKU had a significantly lower IQ (mean IQ: 100 ± 12) than the control group (mean IQ: 108 ± 11): \(t(107) = 3.3, p = .001\). The observed differences were unaffected by exclusion of BH\(_4\) users: \(t(96) = 3.0, p = .004\).

**Working Memory.** For all WM tasks, there was a strong effect of condition, i.e. participants performed worse in task parts with higher WM load. Both patients with PKU and healthy controls made more errors and were slower, with \(\eta^2\) ranging from 0.55 to 0.91 (see Table 3) and with strong observed power around 1.00 in the more demanding conditions of the tasks.

In the Visuo-Spatial Sequencing task (VSS), patients with PKU were less accurate and slower than controls (Table 3). A significant group by condition interaction indicated that the difference in accuracy between patients and controls was larger when both location and order of target stimuli were taken into account (see Figure 1).
Table 2. Descriptive statistics measurements for PKU and control group

<table>
<thead>
<tr>
<th></th>
<th>PKU (n = 57)</th>
<th>PKU nonBH4 (n = 45)</th>
<th>PKU BH4 (n = 12)</th>
<th>Controls (n = 57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ estimate*</td>
<td>100 ± 12</td>
<td>100 ± 12</td>
<td>99 ± 12</td>
<td>108 ± 11</td>
</tr>
<tr>
<td><strong>Working Memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean N correctly identified targets</td>
<td>103.8 ± 1.7</td>
<td>103.8 ± 1.6</td>
<td>103.9 ± 2.1</td>
<td>104.4 ± 0.7</td>
</tr>
<tr>
<td>Mean N correct in correct order</td>
<td>98.6 ± 5.6</td>
<td>98.2 ± 5.7</td>
<td>100.1 ± 4.8</td>
<td>100.7 ± 3.5</td>
</tr>
<tr>
<td>Mean RT in ms</td>
<td>1057 ± 207</td>
<td>1090 ± 213</td>
<td>934 ± 125</td>
<td>889 ± 141</td>
</tr>
<tr>
<td>MS2D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N errors Low memory load</td>
<td>1.7 ± 2.4</td>
<td>1.6 ± 2.1</td>
<td>2.3 ± 3.4</td>
<td>1.7 ± 2.2</td>
</tr>
<tr>
<td>N errors High memory load</td>
<td>10.3 ± 6.6</td>
<td>9.9 ± 6.5</td>
<td>11.9 ± 7.0</td>
<td>9.2 ± 8.0</td>
</tr>
<tr>
<td>Mean RT Low memory load in ms</td>
<td>605 ± 146</td>
<td>622 ± 144</td>
<td>542 ± 142</td>
<td>549 ± 80</td>
</tr>
<tr>
<td>Mean RT High memory load in ms</td>
<td>2144 ± 570</td>
<td>2260 ± 551</td>
<td>1708 ± 421</td>
<td>1875 ± 418</td>
</tr>
<tr>
<td>FI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N errors (dissimilar)</td>
<td>1.4 ± 2.2</td>
<td>1.5 ± 2.3</td>
<td>1.0 ± 1.7</td>
<td>1.3 ± 1.7</td>
</tr>
<tr>
<td>N errors (similar)</td>
<td>4.5 ± 2.6</td>
<td>4.5 ± 2.6</td>
<td>4.6 ± 2.9</td>
<td>3.9 ± 2.6</td>
</tr>
<tr>
<td>Mean RT dissimilar in ms</td>
<td>893 ± 200</td>
<td>919 ± 187</td>
<td>796 ± 225</td>
<td>834 ± 135</td>
</tr>
<tr>
<td>Mean RT similar in ms</td>
<td>1705 ± 457</td>
<td>1780 ± 455</td>
<td>1421 ± 355</td>
<td>1539 ± 318</td>
</tr>
<tr>
<td><strong>Inhibitory Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N errors</td>
<td>24.6 ± 14.7</td>
<td>24.3 ± 15.0</td>
<td>25.8 ± 13.8</td>
<td>17.3 ± 10.3</td>
</tr>
<tr>
<td>Mean Series Time in s</td>
<td>9.3 ± 1.8</td>
<td>9.6 ± 1.9</td>
<td>8.0 ± 1.1</td>
<td>8.7 ± 1.5</td>
</tr>
<tr>
<td><strong>Sustained Attention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAD - over the course of the task</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Errors first 120 trials</td>
<td>4.0 ± 3.0</td>
<td>4.0 ± 2.9</td>
<td>4.0 ± 3.3</td>
<td>2.4 ± 1.7</td>
</tr>
<tr>
<td>Errors last 120 trials</td>
<td>5.2 ± 3.1</td>
<td>5.2 ± 3.2</td>
<td>5.3 ± 2.7</td>
<td>4.1 ± 2.8</td>
</tr>
<tr>
<td>Mean Series Time first 120 trials in s</td>
<td>9.1 ± 1.5</td>
<td>9.4 ± 1.5</td>
<td>7.7 ± 1.0</td>
<td>8.7 ± 1.5</td>
</tr>
<tr>
<td>Mean Series Time last 120 trials in s</td>
<td>9.4 ± 2.2</td>
<td>9.7 ± 2.2</td>
<td>8.3 ± 1.4</td>
<td>8.6 ± 1.6</td>
</tr>
<tr>
<td><strong>Mental Health</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASR DSM-IV scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive problems</td>
<td>5.1 ± 4.4</td>
<td>5.5 ± 4.5</td>
<td>3.6 ± 3.5</td>
<td>3.4 ± 3.8</td>
</tr>
<tr>
<td>Anxiety problems</td>
<td>3.5 ± 2.7</td>
<td>3.6 ± 2.6</td>
<td>3.2 ± 3.1</td>
<td>3.2 ± 2.8</td>
</tr>
<tr>
<td>Somatic problems</td>
<td>1.9 ± 1.9</td>
<td>1.9 ± 1.8</td>
<td>1.8 ± 2.3</td>
<td>1.4 ± 2.0</td>
</tr>
<tr>
<td>Avoidant Personality problems</td>
<td>3.0 ± 2.7</td>
<td>3.1 ± 2.7</td>
<td>2.5 ± 2.5</td>
<td>1.8 ± 2.0</td>
</tr>
<tr>
<td>Attention Deficit/Hyperactivity problems</td>
<td>4.6 ± 3.5</td>
<td>4.6 ± 3.7</td>
<td>4.6 ± 2.6</td>
<td>4.6 ± 3.4</td>
</tr>
<tr>
<td>Antisocial Personality problems</td>
<td>2.4 ± 2.4</td>
<td>2.5 ± 2.5</td>
<td>1.8 ± 1.9</td>
<td>2.4 ± 2.5</td>
</tr>
</tbody>
</table>

* 2 IQ scores missing in nonBH4 group, 1 in BH4 group, 2 in control group
VSS: Visuo-Spatial Sequencing; MS2D: Memory Search 2 Dimensions; FI: Feature Integration; SAD: Sustained Attention-Dots; ASR: Adult Self-Report
Cognitive profile and mental health in adult PKU

In the Memory Search 2 Dimensions task (MS2D), no differences between patients with PKU and controls were observed regarding accuracy of task performance. Patients with PKU were, however, significantly slower than controls. Moreover, there was a group by condition interaction: results showed a greater decline in response speed for patients than for controls when WM load increased (see Figure 2).

For the Feature Integration task (FI), again with similar accuracy, patients with PKU performed slower than controls and the differences between the control and PKU group became larger with increasing difficulty in working memory condition, i.e. with similar distracter stimuli (see Figure 3).
Inhibitory Control. ANOVAs showed that across the entire Sustained Attention-Dots task (SAD), patients with PKU were less accurate and performed slower than controls (see Table 3).

Sustained Attention. Repeated measures analyses of variance across the SAD revealed a significant increase of errors made at the end of the task compared to the start for all participants (see Table 2 and Table 3). There was no significant group by time effect for accuracy, indicating that the difference between patients with PKU and controls in accuracy of task performance did not increase with time. For speed of task performance, however, such an effect was observed. This interaction was caused by slowing of task execution in the PKU group (from MST = 9.1 sec ± 1.5 to a MST of 9.4 sec ± 2.2), whereas response speed of controls remained stable towards the end of the task (from a Mean Series Time of 8.7 sec ± 1.5 to a MST of 8.6 ± 1.6) (see Figure 4, right panel).

Excluding the data from the BH₄ users did not change the abovementioned results with the ANT tasks.

Mental Health. The MANOVA showed that patients with PKU reported significantly more Depressive problems and Avoidant Personality problems than controls (see Table 3). When excluding the BH₄ group, results remained similar and effect sizes became somewhat larger (Depressive problems: $\eta_p^2 = .065$; Avoidant problems: $\eta_p^2 = .067$).
Table 3. Overview Repeated Measures/MANOVAs for PKU group (n=57) and controls (n=57)

<table>
<thead>
<tr>
<th></th>
<th>PKU versus controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F$</td>
</tr>
</tbody>
</table>

### Working Memory: VSS
- **N of correct responses**
  - Condition: 146.2, 1,112***, 0.566
  - Group: 7.2, 1,112**, 0.060
  - Condition*Group: 4.4, 1,112*, 0.038
- **Reaction time**
  - Group: 25.6, 1,112***, 0.186

### Working Memory: MS2D
- **N of errors**
  - Condition: 157.6, 1,112***, 0.585
  - Group: 0.5, 1,112, 0.005
  - Condition*Group: 0.8, 1,112, 0.007
- **Reaction time**
  - Condition: 1181.1, 1,112***, 0.913
  - Group: 9.2, 1,112**, 0.076
  - Condition*Group: 6.5, 1,112*, 0.055

### Working Memory: FI
- **N of errors**
  - Condition: 136.2, 1,112***, 0.549
  - Group: 1.1, 1,112, 0.009
  - Condition*Group: 1.3, 1,112, 0.012
- **Reaction time**
  - Condition: 930.1, 1,112***, 0.893
  - Group: 4.8, 1,112*, 0.041
  - Condition*Group: 4.6, 1,112*, 0.039

### Inhibitory Control: overall SAD (ANOVA)
- Total errors: 9.5, 1,112**, 0.078
- Mean Series Time (MST): 3.9, 1,112, 0.034

### Sustained Attention: over time SAD (Repeated Measures)
- **N of errors**
  - Time: 41.5, 1,112***, 0.270
  - Group: 9.5, 1,112**, 0.078
  - Time*Group: 1.0, 1,112, 0.009
- **Mean Series Time (MST)**
  - Time: 1.2, 1,112, 0.010
  - Group: 4.4, 1,112*, 0.038
  - Time*Group: 5.0, 1,112*, 0.043

### Mental Health: ASR (MANOVA)
- Depressive problems: 5.2, 1,110*, 0.045
- Anxiety problems: 0.4, 1,110, 0.004
- Somatic problems: 1.5, 1,110, 0.013
- Avoidant Personality problems: 6.4, 1,110*, 0.055
- Attention Deficit/Hyperactivity problems: 0.0, 1,110, 0.000
- Antisocial Personality problems: 0.0, 1,110, 0.000

*p<0.05  **p<0.01  ***p<0.001  +p<0.10
1 VSS Condition: N correctly identified targets versus N correctly identified targets in correct order
2 MS2D Condition: low (1 target) versus high (3 targets) working memory load
3 FI Condition: dissimilar versus similar target distracters
4 SAD Condition: first 120 trials versus last 120 trials
IQ as covariate. Including IQ in the analyses of variance affected one effect observed for accuracy to such an extent that it became non-significant (i.e. the FI group by condition effect) and also two effects observed for speed of responding (i.e. the MS2D group by condition effect, which became a trend \[ F(1,106) = 3.3, p = .071, \eta^2 = .030 \] and the SAD group by time effect, which also became a trend \[ F(1,106) = 3.0, p = .085, \eta^2 = .028 \]). After inclusion of IQ as a covariate, analyses of variance still showed significant group differences regarding speed of responding in the working memory tasks VSS \[ F(1,106) = 15.1, p < .001, \eta^2 = .124 \] and MS2D \[ F(1,106) = 4.9, p = .029, \eta^2 = .044 \], accuracy of inhibitory control (as measured by SAD: \[ F(1,106) = 5.9, p = .017, \eta^2 = .053 \]), and accuracy in sustained attention (also measured with SAD: \[ F(1,106) = 5.4, p = .021, \eta^2 = .049 \]). Group differences in Depressive \[ F(1,105) = 4.0, p = .048, \eta^2 = .037 \] and Avoidant Personality \[ F(1,105) = 4.9, p = .029, \eta^2 = .045 \] also remained significant after taking into account any possible IQ effects.

Correlations with concurrent Phe, lifetime Phe, and childhood Phe

IQ. For the whole PKU group, concurrent Phe was marginally (and negatively) related to the IQ estimate \( r = -0.19, p = .081 \). When examining the non-BH\(_4\) group only, lifetime Phe was marginally (and negatively) correlated with IQ \( r = -0.26, p = .051 \) and Phe between 0-12 years was significantly (and negatively) correlated with IQ \( r = -0.30, p = .029 \).

Working Memory. Concurrent Phe was positively related to RT of task MS2D high WM load condition \( r = 0.33, p = .006 \), and to RT of FI (low WM load condition: \( r = 0.26, p = .024 \); high WM load: \( r = 0.26, p = .026 \)), indicating slower performance with higher Phe levels. For task FI, the number of errors made under high WM load was also positively correlated with lifetime Phe \( r = 0.24, p = .039 \) and a trend was seen for childhood Phe \( r = 0.22, p = .051 \). When excluding the BH\(_4\) group, concurrent Phe was significantly related to RT of the MS2D high WM load condition \( r = 0.27, p \)}
Cognitive profile and mental health in adult PKU

Inhibitory Control and Sustained Attention. In the SAD task, concurrent Phe was significantly related to MST \( r = 0.36, p = .003 \) and to MST at the end of the task \( r = 0.41, p = .001 \): higher Phe levels were associated with slower task performance. These correlations remained significant when leaving out the BH\(_4\) group, although they became somewhat weaker.

Mental Health. No significant correlations were observed between Phe measures and mental health outcomes, regardless of the in- or exclusion of BH\(_4\) users.

Group comparisons based on concurrent Phe, lifetime Phe, and childhood Phe

We also compared groups on IQ, EF, and mental health, after patients with PKU were split into low and high Phe groups according to the upper target level most frequently used in the literature and in treatment guidelines, i.e. \( 360 \mu\text{mol/L} \).

IQ. With respect to IQ, patients with lifetime Phe \( \geq 360 \mu\text{mol/L} \) \( n = 34 \) differed from controls \( F(1,87) = 14.8, p < .001, \eta^2_p = .145 \) and from patients with Phe <\( 360 \mu\text{mol/L} \) \( n = 19 \) \( F(1,51) = 4.2, p = .046, \eta^2_p = .076 \). When concurrent Phe was used to create groups, again only patients with PKU with Phe \( \geq 360 \mu\text{mol/L} \) \( n = 42 \) differed from controls \( F(1,95) = 11.6, p = .001, \eta^2_p = .076 \). When childhood Phe was taken to create groups, controls differed from both PKU groups, and patients with Phe levels < and \( \geq 360 \mu\text{mol/L} \) did not differ from each other.

Working memory. In the VSS task, only patients with lifetime Phe \( \geq 360 \mu\text{mol/L} \) performed worse than controls (correctly identified targets: \( F(1,91) = 7.7, p = .007, \eta^2_p = .078 \); correctly identified in correct order: \( F(1,91) = 7.4, p = .008, \eta^2_p = .075 \)). Also, when childhood Phe was used to create groups, only patients with Phe \( \geq 360 \mu\text{mol/L} \) \( n = 21 \) were less accurate than controls when there was a high WM load \( F(1,76) = 5.4, p = .023, \eta^2_p = .066 \). In the MS2D task, high WM load condition, again only patients with Phe \( \geq 360 \mu\text{mol/L} \) differed from controls (lifetime: \( F(1,91) = 10.2, p = .002, \eta^2_p = .101 \); concurrent: \( F(1,100) = 8.9, p = .004, \eta^2_p = .082 \)), this time with respect to RT. In the FI task, when higher-order WM was required (i.e. the similar condition), patients with childhood Phe \( \geq 360 \mu\text{mol/L} \) made more errors than controls \( F(1,76) = 5.7, p = .020, \eta^2_p = .069 \) but also compared to those with Phe <\( 360 \mu\text{mol/L} \) \( F(1,54) = 4.1, p = .049, \eta^2_p = .070 \).

Inhibitory Control and Sustained Attention. In the SAD task, only patients with lifetime Phe \( \geq 360 \mu\text{mol/L} \) differed from controls in overall accuracy \( F(1,91) = 9.3, p = .003, \eta^2_p = .093 \), and only patients with concurrent Phe \( \geq 360 \mu\text{mol/L} \) were slower than controls at the end of the task \( F(1,100) = 7.1, p = .009, \eta^2_p = .067 \).
Mental Health. Only patients with lifetime Phe \( \geq 360 \ \mu mol/L \) had more Depressive \( F(1,90) = 8.3, p = .005, \eta_p^2 = .084 \) and Avoidant personality problems than controls \( F(1,90) = 6.2, p = .015, \eta_p^2 = .064 \). Also, only patients with PKU with childhood Phe \( \geq 360 \ \mu mol/L \) had more Depressive problems than controls \( F(1,75) = 8.4, p = .005, \eta_p^2 = .101 \). They also reported more Somatic problems than both controls \( F(1,90) = 6.2, p = .015, \eta_p^2 = .064 \) and patients with childhood Phe \( < 360 \ \mu mol/L \) \[ F(1,75) = 8.4, p = .005, \eta_p^2 = .101 \].

As these results may have been influenced by the inclusion of BH\(_4\) users, the analyses were repeated excluding the BH\(_4\) users, but this did not result in different outcomes, except for one result that gave a somewhat stronger indication that Phe levels below 360 \( \mu mol/L \) during the first 12 years of life may result in better adult outcomes than Phe levels above that level. Patients with Phe \( \geq 360 \ \mu mol/L \) (n = 19) differed from both controls (n = 57) \[ F(1,74) = 5.5, p = .021, \eta_p^2 = .070 \] and patients with Phe \( < 360 \ \mu mol/L \) (n = 25) \[ F(1,42) = 4.8, p = .035, \eta_p^2 = .102 \] regarding accuracy when a high WM load was in place, as measured by the FI task.

Comparison low Phe-nonBH\(_4\) versus low Phe-BH\(_4\)

Next comparisons were made between BH\(_4\) users (n=10) and non-users (n=25). Only patients with Phe concentrations in childhood \( < 360 \ \mu mol/L \) were included, as BH\(_4\) responsive patients generally already have lower Phe levels than non-responsive patients. Two BH\(_4\) responsive patients had Phe \( \geq 360 \ \mu mol/L \) and were therefore excluded from these analyses. There was no significant difference in gender \( (\chi^2 = 0.7, p = .392) \), childhood and concurrent Phe between the BH\(_4\) and non-BH\(_4\) group: respectively \( t(33) = 1.5, p = .14 \) and \( t(33) = 1.2, p = .24 \). Childhood and concurrent Phe was resp. 255 \( \mu mol/L \ \pm 43 \) and 497 \( \mu mol/L \ \pm 225 \) for the BH\(_4\) group, and for the non-BH\(_4\) group this was resp. 280 \( \mu mol/L \ \pm 44 \) and 643 \( \mu mol/L \ \pm 359 \). However, both groups differed in lifetime Phe \( (t(33) = 2.6, p = .013) \) and in age \( (t(33) = 2.1, p = .045) \). The BH\(_4\) group had a lifetime Phe of 316 \( \mu mol/L \ \pm 61 \), while the non-users had a lifetime Phe of 405 \( \mu mol/L \ \pm 99 \). And the BH\(_4\) group was younger than the non-BH\(_4\) group (mean age 24.0 years \( \pm 5.3 \) versus mean age 28.6 years \( \pm 6.0 \)).

The BH\(_4\) group had an IQ of 98 \( \pm 13 \), while this was 103 \( \pm 10 \) for the non-BH\(_4\) users. Results showed that IQ was not significantly different between both groups: \( t(32) = 1.2, p = .24 \). The group of non-users was slower on the WM tasks than the BH\(_4\) group: VSS \[ F(1,33) = 6.2, p = .018, \eta_p^2 = .158 \], MS2D \[ F(1,33) = 7.7, p = .009, \eta_p^2 = .188 \], with the non-BH\(_4\) group performing more slowly in the high WM condition \[ F(1,33) = 7.6, p = .009, \eta_p^2 = .187 \], and FI \[ F(1,33) = 3.3, p = .078, \eta_p^2 = .091 \], again with the non-users being slower in the high WM condition \[ F(1,33) = 4.6, p = .039, \eta_p^2 = .122 \]. A similar result was observed for response speed on the SAD inhibitory control task \[ MST overall: F(1,33) = 5.4, p = .027, \eta_p^2 = .140 \], and MST across the SAD task,
i.e. sustained attention \( [F(1,33) = 5.9, p = .021, \eta^2_p = .151] \). When only patients were compared whose lifetime Phe (as well as childhood) levels were <360 µmol/L, the group of non-users \((n = 10)\) reported more Depressive problems \( [F(1,17) = 4.4, p = .050, \eta^2_p = .207] \) and Anxiety problems \( [F(1,17) = 8.2, p = .011, \eta^2_p = .325] \) than the BH\(_4\) group \((n = 9)\).

Finally, because both groups differed in age, the BH\(_4\) group was matched to 10 non-BH\(_4\) users, after which both groups had similar age, concurrent, childhood, and lifetime Phe. There were trends for information processing speed with regards to the VSS \( [F(1,18) = 3.5, p = .076, \eta^2_p = .165] \), where non-BH\(_4\) users were slower, MS2D \( [F(1,18) = 3.9, p = .062, \eta^2_p = .180] \), where the non-BH\(_4\) group was again slower in the high WM condition \( [F(1,18) = 5.2, p = .034, \eta^2_p = .226] \), as they were in the high WM condition of the FI \( [F(1,18) = 3.4, p = .080, \eta^2_p = .160] \). Somatic problems were also reported more by the non-BH\(_4\) group \( [F(1,18) = 5.4, p = .031, \eta^2_p = .232] \).

**DISCUSSION**

This study showed that, compared to healthy controls, adult patients with PKU had a slightly lower IQ, albeit in the normal range, poorer working memory, and poorer inhibitory control, which became more evident when combined with sustained attention. Patients also had more internalizing mental health problems, specifically depressive and avoidant personality problems. The cognitive profile observed in adult patients with PKU thus resembled the cognitive profile observed in children and young adolescents with PKU (Albrecht et al., 2009; Christ et al., 2010; Huijbregts et al., 2003; Leuzzi et al., 2004; Waisbren et al., 2007). Patients were mainly slower in responding when working memory demands were high or when inhibitory control was required together with sustained attention, while also often making more errors. Several results, concerning both accuracy and speed of task performance, were irrespective of IQ. This may be considered as further support for specific impairments in (adult) PKU. Compared to studies with younger participants (e.g. Anderson et al., 2007; Huijbregts, de Sonneville, Licht, Sergeant, & van Spronsen, 2002; Huijbregts et al., 2002), effect sizes for group differences observed in the present study were smaller. This seems to contrast, to some extent, with the results from a meta-analyses by Moyle and colleagues, who reported Hedges’ g scores between 0.40-0.60 for group differences on several cognitive tests (including IQ and executive functioning) in groups of adolescents and adults (Moyle, Fox, Arthur, Bynevelt, & Burnett, 2007). However, it is not clear whether these strong effect sizes could be attributed to the adolescent group, as a distinction between adolescents and adults was not made.
In accordance with the meta-analysis by Albrecht et al. (2009), the strength of associations between Phe concentrations and cognitive outcomes seems to decrease with age. Albrecht et al. (2009) found strong correlations in studies with children and adolescents with PKU, but no such associations for adult patients, whereas we only found few significant correlations, regardless of whether concurrent, lifetime, or childhood Phe levels were studied. In our study, Phe concentrations in childhood were negatively related to IQ, whereas concurrent and lifetime Phe levels were negatively related to (several aspects of) working memory task performance, and concurrent Phe level was negatively associated with (several aspects of) inhibitory control and sustained attention.

Previous studies have demonstrated that patients with lifetime Phe below 360 µmol/L have better executive functioning (Huijbregts et al., 2002; Leuzzi et al., 2004). Although the results from correlational analyses were not very convincing, several indications were found, in accordance with previous studies (Huijbregts et al., 2002; Leuzzi et al., 2004), for the importance of keeping Phe levels below the upper target level of 360 µmol/L, which is recommended in most treatment guidelines. Overall group differences (on both cognitive and mental health outcomes) between patients with PKU and controls mainly resulted from differences between controls and patients with PKU who had Phe levels higher than 360 µmol/L. Patients with lifetime Phe levels ≥360 µmol/L most consistently differed from controls, but similar results were observed when groups were created based on childhood and concurrent Phe levels. It should be noted that patients with Phe <360 µmol/L generally did not differ from those with Phe levels ≥360 µmol/L. The exceptions were IQ, where patients with lifetime Phe ≥360 µmol/L had lower IQ scores than both controls and patients with Phe <360 µmol/L; one working memory task, where patients with childhood Phe ≥360 µmol/L had poorer scores than both controls and patients with childhood Phe <360 µmol/L; and internalizing mental health problems, where patients with childhood Phe ≥360 µmol/L reported more somatic problems than both controls and patients with childhood Phe <360 µmol/L.

Mental health issues, especially internalizing problems were again observed in our study, comparable to what has been reported in other studies (Anjema et al., 2011; Arnold et al., 1998; Jahja et al., 2013; Smith & Knowles, 2000; Weglage et al., 2000). Our patients showed more depressive and avoidant personality problems, which indicates avoidance of social interactions and feelings of inadequacy and inferiority. Mental health is strongly related to social cognition and socials skills, which have also been found to be suboptimal in adults with PKU (Jahja et al., 2016).

Finally, patients not using BH₄ who were matched to BH₄ users with respect to childhood (or pre-treatment) Phe concentrations, were slower than BH₄ users on working memory, inhibitory control, sustained attention, and reported more
internalizing problems. The fact that these differences were observed between patients matched on pre-treatment Phe concentrations, makes it tempting to believe that BH₄ treatment has a positive impact on cognitive and behavioural outcomes, through another mechanism than lowering Phe levels. The few studies focusing on BH₄ and cognitive outcome suggested that BH₄ improves cerebral dopamine concentrations (which are strongly associated with cognitive functions) and white matter integrity, and enhances brain activation (Christ et al., 2013; van Vliet et al., 2015; White et al., 2013). For example, Christ et al. (2013) reported associations between BH₄ treatment, neural activation patterns and working memory performance. However, these effects appeared to be related to a drop in Phe levels following BH₄ treatment (although no significant correlations between neural activation and Phe were found). There are also studies indicating that BH₄ increases the stability of Phe (Burton, Bausell, Katz, Laduca, & Sullivan, 2010), which in itself seems to be related to cognitive outcomes (Anastasoaie, Kurzius, Forbes, & Waisbren, 2008). To date, the mechanisms through which BH₄ treatment might improve cognition and mental health remain unclear, and mechanisms not directly associated with the pharmaceutical properties of BH₄ should also be considered, particularly because BH₄ responsiveness is already associated with milder disease. For example, psychological (or placebo) effects of BH₄ may also be present. For example, BH₄ users follow a more normal diet and/or feel better, have less internalizing thoughts and fewer social adjustment problems, than those patients on a strict diet. This might result in cross-over effects with respect to cognitive functioning.

Clinical relevance and Limitations
It should also be noted that the observed statistically significant differences between the BH₄ users and non-users regarding cognition and mental health may not be clinically relevant. When taking Cohen’s recommendations for interpreting effect sizes (with a partial η² of 0.01 deemed small, a partial η² of 0.06 deemed medium, and a partial η² of 0.14 deemed large) and power (which should be .80 or higher) (Cohen, 1988), results for only one working memory task (MS2D) could be considered clinically relevant, with partial η² of 0.19, and power close to .80 (.76). For the mental health differences between BH₄ users and non-users, only the result regarding anxiety could be considered clinically relevant (partial η² = .33, power = .78). Obviously, effect sizes and power are also largely determined by group sizes, and, whereas the PKU-COBESO study already has a relatively large number of participants, PKU remains a relatively rare disease which makes it difficult to achieve satisfactory power levels. Together with the results from other studies into BH₄ use and cognitive and mental health outcomes, our results warrant further research into both effects and potential working mechanisms of BH₄. Further to this theme, it seemed like
a number of our main effects indicate not only statistical significance but clinical relevance as well. Effect sizes were medium to large and power was satisfactory (around .80 or higher) for a number of overall group differences between controls and patients with PKU (accuracy and speed of working memory, as measured by different tasks, inhibitory control, and avoidant personality traits). However, clinical relevance became particularly evident when patients with PKU with Phe levels ≥360 µmol/L were compared to healthy controls, with medium to large effect sizes and power generally well above .80 (range .77 to .97) for IQ, two out of three working memory tasks (speed and accuracy), inhibitory control (speed and accuracy), and depressive behaviour. These clinically and statistically significant group differences were mainly observed when groups were created based on lifetime Phe levels, with some effects also present when childhood or concurrent Phe levels were used for group formation.

Following the above, it is clear that the first limitation of this study concerns the small sample sizes for subgroup comparisons. A second issue is the continued discussion of whether or not to include IQ as a covariate in analyses of variance comparing patients with PKU and controls on specific cognitive abilities. As IQ scores are made up of a number of different cognitive abilities, including the executive functions measured in the present study, including IQ as a covariate may obscure specific impairments. Although several group differences remained significant after inclusion of IQ as a covariate, it is possible that if we would have administered the full range of IQ tests, larger decreases of group differences would have been observed. The issue however remains as to whether this would be a good approach, as it may be better to identify and target specific weaknesses within a cognitive profile. A third issue or limitation of the present study is that we cannot rule out the possibility that particularly patients who had maintained treatment and regular monitoring participated in this study, which may have led to an underestimation of effects. Future studies should try to include the full spectrum of (adult) patients with PKU. Finally, and this in part repeats what has already been discussed, the small number of participants for specific subgroup analyses, in combination with a lack of normative data for the neuropsychological tasks makes it difficult to determine the clinical relevance of the observed effects. Also, there may be other factors involved in influencing cognitive and mental health outcome in adulthood such as experiencing social support and quality of life. However, in general, there are indications that the adult patients with PKU still function suboptimally compared to the healthy control group regarding executive functioning and mental health, which is important to consider in future research into optimal treatment and monitoring of PKU from birth onwards.
REFERENCES


data on blood prolactin concentrations suggest increased cerebral dopamine concentrations. *Molecular Genetics and Metabolism, 114*(1), 29-33.


