Molecular Genetics and Metabolism Reports 34 (2023) 100953

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism Reports

journal homepage: www.elsevier.com/locate/ymgmr

Neuropsychological endpoints for clinical trials in methylmalonic acidemia and propionic acidemia: A pilot study

Kimberly A. Chapman [a,*], Devon MacEachern [b], Gerald F. Cox [c,d], Mavis Waller [c], Jeanine Fogarty [c], Suzanne Granger [b], Miganush Stepanians [b], Susan Waisbren [d]

[a] Children's National Rare Disease Institute, 7125 13th Pl NW, Washington DC 20012, USA
[b] PROMETRIKA, LLC, 100 CambridgePark Drive, Cambridge, MA 02140, USA
[c] HemoShear Therapeutics Inc., 501 Locust Ave #301, Charlottesville, VA 22902, USA
[d] Boston Children's Hospital, Harvard Medical School, 300 Longwood Ave., Boston, MA 02115, USA

ARTICLE INFO

Keywords:
Methylmalonic acidemia
Propionic acidemia
Neuropsychological testing
Cookie theft picture task, Vineland adaptive behavior scale (VABS)

ABSTRACT

Introduction: This pilot study assessed instruments measuring relatively discrete neuropsychological domains to inform the selection of clinical outcome assessments that may be considered for interventional trials in methylmalonic acidemia (MMA) and propionic acidemia (PA).

Methods: Tests and questionnaires were selected for their possible relevance to MMA and PA and potential sensitivity to modest changes in functioning and behavior.

Results: Twenty-one patients (<18 years, $n = 10$;>18 years, $n = 11$) and/or their caregivers responded to video interviews and paper tests. Language deficits and significant motor deficits in some participants impacted scoring, especially in the verbal and processing speed sections of the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V) and the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV). However, all participants ≥12 years of age were able to complete the Cookie Theft Picture Task. Thus, verbal discourse remains a potentially useful endpoint for participants in this age group. The Vineland Adaptive Behavior Scales (VABS-3) Adaptive Behavior Composite and Communication Scores confirmed delayed or immature functioning in day-to-day activities in these participants. Significant motor deficits prevented completion of some tests. Computerized processing speed tasks, which require pressing a button or tapping a computer screen, may be easier than writing or checking off boxes on paper in this cohort. Sleep characteristics among MMA participants were within normative ranges of the Child and Adolescent Sleep Checklist (CASC), indicating that this measurement would not provide valuable data in a clinical trial. Despite their challenges, responses to the Metabolic Quality of Life Questionnaire indicated these patients and their caregivers perceive an overall high quality of life.

Conclusion: Overall, test and questionnaire results were notably different between participants with MMA and participants with PA. The study demonstrates that pilot studies can detect instruments that may not be appropriate for individuals with language or motor deficits and that may not provide a broad range of scores reflecting disease severity. It also provides a rationale for focusing on discrete neuropsychological domains since some aspects of functioning were less affected than others and some were more closely related to disease severity. When global measures are used, overall scores may mask specific deficits. A pilot study like this one cannot ensure that scores will change over time in response to a specific treatment in a clinical trial. However, it can avert the selection of instruments that do not show associations with severity or biomedical parameters likely to be the target of a clinical trial. A pilot study can also identify when differences in diagnoses and baseline functioning need to be addressed prior to developing the analytical plan for the trial.

Abbreviations: MMACHC, MMA with cobalamin C deficiency, CblC.

* Corresponding author at: 7125 13th Place NW, Washington DC 20012, USA.
E-mail address: KChapman@childrensnational.org (K.A. Chapman).

https://doi.org/10.1016/j.ymgmr.2022.100953
Received 2 October 2022; Received in revised form 27 December 2022; Accepted 28 December 2022

1. Introduction

The various enzymatic subtypes of methylmalonic acidemia (MMA, OMIM #251000, 251100, 251110) and propionic acidemia (PA, OMIM #606054) are intoxication-type metabolic disorders associated with developmental delay, intellectual disability, metabolic encephalopathy, and movement disorders as well as other significant medical complications [1–3]. Clinical trials for these disorders are investigating treatments that aim to improve disease-related morbidities, including neuropsychologic; consequently, measures of neuropsychologic ability need to be evaluated for their feasibility and utility as clinical outcome assessments in interventional studies.

Several studies have demonstrated risk for, or presence of, development delay in MMA and PA [2,4,5]. Complicating the picture is that 75% of PA infants and 56% of MMA infants have a severe neonatal onset, which can cause irreversible brain damage leading to static abilities [4]. In this pilot study, we aimed to inform the selection of clinical outcome assessments that may be considered for interventional clinical trials in MMA and PA. Although often discussed together, studies of neuropsychological abilities are frequently published separately for MMA and PA. Thus, we will present our findings for each disorder separately.

Individuals with MMA are at risk for developmental delay and cognitive difficulties with diminished functioning [6]. One study followed 20 patients with MMA, whose mean IQ was 85 (the normative mean is 100). Scores on the processing speed subtest were lower than scores on subtests measuring verbal comprehension, perceptual reasoning, and working memory [7]. Another study noted fine motor weaknesses [8]. Investigators did not generally allude to behavioral, emotional or psychiatric symptoms associated with MMA (except for cobalamin C deficiency). However, one early study noted instances of aggression, attention deficit disorder, and behaviors on the autistic spectrum [9]. A few reports document single cases of late-onset MMA identified because of altered behavior or psychosis [10].

Studies of children and adults with PA document a broad range of neuropsychological outcomes. More than half to three-quarters of cases performed in the range of intellectual disabilities in the largest follow-up studies, as measured by the Snijders-Oomen Test (SON) and the Culture-Free-Test (CFT 20-R) [5,11]. Motor delays and/or deficits occurred in >50% of the 55 children in one series, although most affected individuals were ambulatory [5]. In the same study, 55% had speech delay. Among 12 consecutively diagnosed patients in another study, 6 demonstrated speech delay and 4, who also exhibited behaviors on the autistic spectrum, remained nonverbal at age 4 years and older. Four other children in this study exhibited behaviors commonly seen in autism [12]. De la Batie et al. [13,14] also described autism in patients with PA. Mentioned in some studies were attention deficits, sleep disorders, depression, anxiety, visual hallucinations, and psychosis [5,11,13,15].

Quality of life as reported on generic health questionnaires (KidKindl questionnaire; Strengths and Difficulties questionnaire; Familien-Belastungs-Fragebogen [Families' Burden Scale]) revealed that the majority of affected children and adolescents considered themselves healthy and feeling similar to peers [5]. Quality of life for families became the focus of studies primarily after the introduction of liver transplantation for MMA. As predicted by Pascoal [16], generic quality of life measures (most frequently, the 36-Item Short Form Health Survey [SF-36], the EuroQoL- Five Dimension [EQ-5D] questionnaire, the Pediatric Quality of Life Inventory [PedsQL], and the Child Heath Questionnaire CHQ]) resulted in scores not dissimilar to scores from families of healthy children, while qualitative interviews revealed significant reductions in anxiety and impressions of the burden of illness following transplantation [17].

Studies identifying specific neuropsychological deficits apart from overall intelligence in MMA and PA remain sparse. Multiple studies in MMA and PA demonstrate that there are neuropsychological differences, but given the impact of motor and speech delays, it is unclear whether these would be amenable to short term assessment of stabilization or change relative to baseline or to a control in a clinical trial.

The goal of this study was to explore the use of several neuropsychological tests and questionnaires as clinical outcome measures tobe considered for interventional clinical trials in MMA and PA.

The specific aims were to:

- Determine the applicability of instruments for use in a clinical trial of MMA and PA in terms of ease of administration, participant burden, ability of participants to complete the tests, and range of scores obtained
- Measure associations between scores on the various tests and with perceived severity of disease
- Identify potential differences in scores for participants with MMA compared to PA

2. Materials and methods

Procedures: All research was performed in accordance with the Declaration of Helsinki and was approved by WCG IRB (IRB00000533; Study Tracking #20203003; Study number 1293015) and the Children's National Institutional Review Board (IRB00009016; Pro00015072). Individuals were identified through out-reach via patient support organizations including the Organic Acidemia Association, the Propionic Acidemia Association, and Metabolic Support UK, and directed to a website maintained by HemoShear Therapeutics, Inc. to learn more about the study and initiate enrollment, if desired. Potential participants were contacted by the study coordinator, who conducted the informed consent process. Data collection began after participants returned the signed consent form. All interviews and neuropsychological testing were conducted via video conferencing and conformed to remote administration methods as recommended by the test publishers, where this information was available. Testing materials for timed paper and pencil tasks were mailed ahead of time and completed during the testing session. All questionnaires were administer by one of the co-authors of this study (JF) who also served as study coordinator. All tests were administered by another co-author (SW), a psychologist. Scoring was automated or performed manually by SW.

Instruments: Tests and questionnaires were selected for their possible relevance to MMA and PA. Overall intellectual functioning as expressed by Full Scale IQ tends not to change over the short term. Therefore, we chose to examine several domains and subdomains of the selected instruments, with the intent of identifying endpoints associated with severity of disease and hence more likely to show change in the time frame of a typical clinical trial.

A primary aim of this study was to assess potential instruments. To do so required a heterogeneous sample in order to determine if scores differentiated varying degrees of disabilities. Only tests yielding a broad range of scores would be appropriate for a clinical trial since an endpoint would need to measure degree of change. A test would not be useful if all participants, no matter their underlying cognitive abilities, performed at the same level. Therefore, we included participants with MMA at risk for more severe impacts (such as MMUT) and those facing less risks MUT^{-} (Waisbren, 2021). For our analyses, however, we excluded the 3 cases with CblC because of vision and psychiatric issues.

The instruments were administered in English. Table 1 lists the instruments and their basic characteristics.

Subtests from the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V; [18]) and the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV; [19]) provide well-validated measures of verbal comprehension and processing speed.

The Vineland Adaptive Behavior Scales-Third Edition (VABS-3, [20]) is a 189-item parent questionnaire assessing the typical skills of individuals who are ages 3 years and older. It has been used extensively in the assessment of individuals with developmental delay, intellectual disabilities, and other conditions affecting performance in day-to-day

Table 1
List of piloted tests and questionnaires.

Instrument	Subtests	Domain assessed	Age range, years	Time to complete, minutes
Tests Administered				
WISC-V WAIS-IV	Vocabulary & Similarities	Verbal Comprehension	6–17 18+	10
WISC-V WAIS-IV	Coding & Symbol Search	Processing Speed	6–17 18+	10
Boston Diagnostic Aphasia Examination	Cookie Theft Picture Task	Verbal Discourse	12+	5
Questionnaires administered				
Vineland Adaptive Behavior Scales-Third Edition	Structured Interview for parent/caregiver	Communication, Daily Living, Socialization, Adaptive Behavior	3+	25
Metabolic Health Quality of Life Questionnaire	Verbally administered to children	Child perception of Quality of Life	8–18	10
Child and Adolescent Sleep Checklist	Verbally administered to parent/caregiver	Bedtime, sleep, movement, daytime sleep problems	3–18	10

WISC-V=Wechsler Intelligence Scale for Children, Fifth Edition
WAIS-IV=Wechsler Adult Intelligence Scale, Fourth Edition

activities. Remote administration has been evaluated and concluded to be valid [21]. In this study, the VABS Communication subscale standard score represented the primary score of interest from the VABS, although other scores for subscales measuring aspects of daily living and socialization are also presented here. The Adaptive Behavior Composite Score incorporates these subscales and provides a measure of overall functioning.

This exploratory pilot study included other instruments that could be considered for analyses in a clinical trial. Of particular interest was the Cookie Theft Picture Task from the Revised Boston Diagnostic Aphasia Examination [22]. This test provides a standardized stimulus for eliciting spontaneous speech, also called verbal discourse. The test was designed for individuals ages 12 years and older. The test simply involves showing a picture of a mother washing dishes and two children, one of whom is reaching for a cookie while standing on a stool that is about to tip over. The examinee is instructed to, "Tell me everything you see going on in this picture." Participant responses are recorded, transcribed and then rated according to various measures of language complexity. Most adults complete this test in 1–2 min.

The stimulus test picture has been criticized for its use in assessment of stroke victims and for not reflecting a diverse, linguistically rich, and multicultural population [23]. Despite this flaw, the test has shown potential to be applied in other circumstances. A ten-year longitudinal study designed to identify early markers of Alzheimer's Disease reported that change in scores on the Cookie Theft Picture Test was a better predictor of later Alzheimer's Disease than results from traditional neuropsychological tests [24]. In participants from the longitudinal Framingham Heart Study, significant predictive power for identifying later onset of Alzheimer's disease was obtained, with accuracy of 0.70, using only linguistic variables derived from the Cookie Theft Picture Task. Results from this test proved to have stronger predictive power than any of the other neuropsychological tests [25]. Researchers employed machine learning in conjunction with MRI and CT scans to determine the utility of the Cookie Theft Picture Task in differentiating individuals with mild aphasia compared to a healthy control group, with a predictive value of 91.67% [26]. They found that participants with any degree of aphasia produced fewer words, fewer different words, fewer open-class words, and fewer nouns than controls. In addition, they produced fewer sentences overall and per unit of time. The scores on the Cookie Theft Picture Task correlated significantly with differences in brain correlates identified through MRI and CT scans. Moreover, scores on the Cookie Theft Picture Task did not correlate with simple language tasks, such as confrontation naming or verbal fluency, which were relatively conserved in mild aphasia. We reasoned that if the Cookie Theft Picture Test was sensitive to subtle declines that it might also detect subtle improvements in verbal discourse. This would make it useful in clinical trials.

The Metabolic Health Quality of Life Questionnaire (MetabQoL v.1; [27]) is a 28-item self-report check-list for children with organic acidemias and urea cycle disorders. The questions focus on intoxication-type disorders placing individuals at risk for metabolic decompensations and requiring dietary restrictions with medical foods and special formulas. Responses are rated on a 5-point Likert Scale. Parents/caregivers completed this questionnaire for children who were not able to do so independently.

The Child and Adolescent Sleep Checklist (CASC; [28]) is a 24-item parent/caregiver-completed questionnaire to identify sleep problems in children, ages 3 to 18 years. Responses to the questions about sleep habits are scored from 0 (never or don't know) to 3 (always), yielding an overall score of 0 to 72. Scores >18 indicate sleep problems.

We did not have complete data regarding enzymatic subtype for the isolated MMA participant since we relied entirely on caregiver or participant report. Therefore, perceived severity of the disorder was obtained during the background interview. The parent/caregiver or adult participant was asked to rate the child's or their own overall metabolic condition as 1) asymptomatic/mild, 2) moderate or 3) severe. Respondents based their answers on a subjective impression, taking into account various factors such as hospitalizations, cognitive abilities, dietary restrictions, and impact on day-to-day life. They were not asked to explain their ratings, so information about whether their response indicated current or general severity of the condition was not available.

Data analyses: Descriptive analyses focused on the time required for participants to complete the test(s), the number of participants able to complete each of the tests, and the ranges of scores. Pearson correlations or non-parametric correlations were performed to determine associations between each of the tests, and analysis of variance was used to describe associations with perceived disease severity as reported by participants or their parents/caregivers. After confirming homogeneity of variance and normality, *t*-tests were used to determine the differences between the MMA and PA subgroups.

Three individuals with MMACHC (cobalamin C deficiency, MMA-CblC) participated in the interviews. Because they do not have isolated MMA the analyses presented here exclude those with MMA-CblC. Their data are included in the supplemental material for reference.

3. Results

3.1. Sample characteristics

The study included 24 participants with a self-reported diagnosis of

MMA or PA. The characteristics of the three participants with MMA-CblC are provided in the supplemental material. The remaining 21 participants are discussed here and the population characteristics are summarized in Table 2.

Twenty-one participants with isolated MMA ($n = 9$) or PA ($n = 12$) had a median age of 18.25 years (range 3.25 to 56.17). Twelve (57.14%) participants were male. Three (14.29%) individuals with MMA enzymatic subtypes had received liver transplants; one of these also had received a kidney transplant. Sixteen (76.2%) participants self-identified as White, 1 as Black/African American, 1 as Hispanic, 2 as Asian, and 1 was unknown. They represented 11 US states, 3 European countries (Scotland, Belgium and Spain), and Australia. All were fluent in English. Seventeen (80.95%) participants were right-handed.

MMA enzymatic subtypes were as follows: 7 (MMUT); 1 cobalamin A (CblA); and 1 cobalamin B (CblB). Self- or caregiver-reported severity of disease was mild in 8 (40%) participants, moderate in 4 (20%) participants, and severe in 8 (40%) participants; severity was not reported in 1 participant. In this study, we do not have reliable information regarding the enzymatic MMA or PA subtypes.

Overall, 13/21 (61.90%) had been diagnosed as having intellectual disabilities. Behaviors on the autistic spectrum were reported in 4 (33.33%) participants with PA. In terms of psychiatric symptoms, 2 (28.57%) participants with MMA and 4 (33.33%) participants with PA had anxiety. Attention deficit hyperactivity disorders (ADHD) were not mentioned as problems by parents/caretakers or participants; no formal assessment of ADHD was performed.

3.2. Aim 1: Applicability of instruments

In general, participants and parents/caregivers cooperated enthusiastically with the neuropsychological tests and questionnaires. None complained about the length of the interview and testing session. Total time for participants was <30 min and for parents/caregivers less than one hour. All tests and questionnaires were easily administered via Zoom audio-visual sessions and all participants returned the paper protocols needed to score the processing speed paper and pencil tests. No sessions needed to be re-scheduled due to internet or other technical issues. Table 3 presents the number and percentage of age-eligible participants who were able to complete each study task. Individuals who were unable to complete the Wechsler tasks received the basal score for each subtest. Three children (ages 3–5 years) were too young to complete the neurocognitive tests: 2 with MMA (5 and 3 years old) and 1 with PA (5 years). Two children (ages 10 and 11 years old), but none of the adults, were unable to complete the tests. Both had PA and exhibited autistic behaviors. All participants ages 12 years and older had enough language to complete the Cookie Theft Picture Task.

Table 2
Demographic and Disease Characteristics for MMA and PA Participants.

	PA $N = 12$	MMA-MMUT $N = 7$	MMA-CblA N = 1	MMA-CblB N = 1	Total $N = 21$
Age (years)					
Mean (SD)	23.71 (15.59)	12.43 (7.36)	6.75(−)	38.67 (−)	19.85 (14.30)
Median	20.54	14.58	6.75	38.67	18.25
Min, Max	5.08, 56.17	3.25, 20.00	6.75, 6.75	38.67, 38.67	3.25, 56.17
Gender, n (%)					
Male	5 (41.7)	5 (71.4)	1 (100)	1(100)	12 (57.1)
Female	7 (58.3)	2 (28.6)	0	0	9 (42.9)
Transplanted, n (%)					
Yes	0	3 (42.9)	0	0	3 (14.3)
No	12 (100)	4 (57.1)	1(100)	1 (100)	18 (85.7)
Perceived Disease Severity					
Asymptomatic/ mild	3 (25.00)	4 (57.14)	1 (100)	–	8 (40.0)
Moderate	3 (25.00)	1(14.29)	0	–	4 (20.0)
Severe	6 (50.00)	2 (28.57)	0	–	8 (40.0)

All instruments yielded a wide range of scores in the combined MMA and PA cohort (Table 3). The Wechsler Verbal Composite and Processing Speed Scores confirmed that cognitive difficulties were prevalent in this population. The VABS Adaptive Behavior Composite Score indicated delayed or immature functioning in day-to-day activities.

The MetabQoL 1.0 total score indicates a perception of a high quality of life. In our sample, 3 patients responded to the questionnaire and the parent/caregiver completed the questionnaire for the other 7 patients. Only 2 of the 28 items were endorsed as more than sometimes "bothersome" by >2 respondents: "Having regular check-ups" and "Having trouble keeping up in school or work."

Only 4 participants were rated as having sleep problems on the CASC. The mean and median scores across all participants were well below the normative cut-off score of 18. In PA participants only, the mean score was 18.83 and the median was 20.

The Cookie Theft Picture Task was completed by 12 participants over 12 years of age. The mean (SD) and median numbers of words used by participants were 70 ± 50 and 59, respectively. The range was broad; 32 to 212 words. In a study of subjects with mild to moderate Alzheimer's disease, the healthy control group had a mean number of words of 99.0 ± 50.6 [29].

3.3. Aim 2: Associations between scores on the various tests and with perceived severity of disease

Associations between the neuropsychological tests and questionnaires (Table 4) and between the tests and perceived disease severity (Table 5) were evaluated. As noted in Table 4, The Wechsler Verbal Composite Score, Wechsler Processing Speed Composite Score, and VABS Adaptive Behavior Composite Score were highly intercorrelated, reinforcing the validity of each test as a measure of functioning. The Cookie Theft Picture Task was associated with the Wechsler Verbal Comprehension Composite Score (the only other direct measure of verbal expression) but not with other measures. This suggests that it may tap into a more discrete functional ability.

Scores from the MetabQoL 1.0 Questionnaire were also associated with other test scores except for the Cookie Theft Picture Task. Scores from the CASC questionnaire were not associated with any other measures, indicating that sleep disorders may represent a discrete symptom.

As noted in Table 5, scores on the Wechsler and Vineland scales differed by perceived severity rating. However, perceived severity was not associated with the number of words used to describe the Cookie Theft Picture Task or with quality of life or sleep disturbance scores.

3.4. Aim 3: Differences in scores for participants with PA compared to MMA

Participants with PA performed below participants with MMA on the Wechsler subtests (Verbal Comprehension $p = 0.0495$; Processing Speed $p = 0.0099$), and the VABS (Fig. 1 and Fig. 2).

On the Cookie Theft Picture Task, participants with PA responded with a mean of 50 ± 20 (median = 43) words compared to 98 ± 68 (median = 81) words used by the MMA group ($p = 0.1952$, $t = 1.52$). The total number of words used by the MMA group did not differ from what was found in the study that included a healthy control group [29]. The mean total score on the MetabQoL was 72 ± 13 (median = 67) for the PA group and 83 ± 11 (median = 81) for the MMA group ($p = 0.1808$, $t = 1.47$), where low scores indicate a more "bothersome" impact of the metabolic disorder and scores <50 suggest a severe impact on quality of life. No participant received a total score < 55, but mean scores on items related to having blood draws, repeated check-ups, and difficulties in school were < 50. On the CASC, only 1 child was described as "Not getting enough sleep" and no domain had an average rating > 2 (indicating an occasional problem). However, 4 children, all with PA, received total scores >18, indicating sleep problems.

Table 3
Age-eligible participants able to complete test/questionnaire and scores.

Instruments & Subtests	Number of Age-eligible Participants	Percent Completed by Participant	Mean ± SD	Range	Median	Normative Mean ± SD/ Cut-offs
Tests						
WISC-V/WAIS-IV Verbal Composite (Vocabulary & Similarities)	18	17 (99.44%)	78± 26	45–133	72	100 ± 15
WISC-V/WAIS-IV Processing Speed Composite (Coding & Symbol Search)	18	17 (99.44%)	71 ± 21	45–124	68	100 ± 15
Cookie Theft Picture Task Total Number of Words	12[a]	12 (100%)	70 ± 50	32–212	59	99 ± 50.6[b]
Questionnaires						
VABS Adaptive Behavior Composite Score	21	21 (100%)	76 ± 23	39–114	73	100 ± 15
VABS Communication Score	21	21 (100%)	74 ± 25	30–109	73	100 ± 15
MetabQoL 1.0 Questionnaire Total Score (High scores = good QoL)	10	10 (100%)	77 ± 13	58–94	80	84 ± 13 (Self) 71 ± 19 (Proxy)
CASC Total Score (High scores = sleep problems)	12[c]	12 (100%)	15 ± 6	3–23	16	Abnormal sleep is score > 18

CASC=Child and Adolescent Sleep Checklist; MetabQoL = Metabolic Quality of Life; VABS=Vineland Adaptive Behavior Scale; WISC-V=Wechsler Intelligence Scale for Children, Fifth Edition; WAIS-IV=Wechsler Adult Intelligence Scale, Fourth Edition.
Note: One participant was unable to complete the WISC-V test and was given a basal score (lowest score) of 45 for the Verbal Composite Score and Processing Speed Composite Score.
[a] Cookie Theft Picture Task is administered to participants over 12 years old.
[b] The normative sample refers to results from [29].
[c] In this test, age 18 years is considered adolescent.

Table 4
Pearson's Correlation for Neurocognitive Tests and Questionnaires.

r *p*-value n	WISC-V/ WAIS-IV Verbal Comprehension Composite Score	WISC-V/ WAIS-IV Processing Speed Composite Score	VABS: Communication Score	VABS: Adaptive Behavior Composite Score	Cookie Theft Picture Task: Total Number of Words	MetabQoL Questionnaire: Total Score	CASC: Total Score
WISC-V/ WAIS-IV Processing Speed Composite Score	0.70870 0.0014 17						
VABS: Communication Score	0.83191 <0.0001 17	0.86929 <0.0001 17					
VABS: Adaptive Behavior Composite Score	0.82534 < 0.0001 17	0.88322 <0.0001 17	0.96808 <0.0001 21				
Cookie Theft Picture Task[a]: Total Number of Words	0.79929 0.0032 11	0.53776 0.0880 11	0.52593 0.0790 12	0.47727 0.1166 12			
MetabQoL Questionnaire: Total Score	0.74477 0.0548 7	0.89960 0.0058 7	0.77893 0.0079 10	0.83862 0.0024 10	−1.00000 2		
CASC: Total Score	−0.43818 0.2381 9	−0.42482 0.2544 9	−0.35425 0.2586 12	−0.27229 0.3919 12	−0.99120 0.0845 3	−0.25701 0.4735 10	

CASC=Child and Adolescent Sleep Checklist; MetabQoL = Metabolic Quality of Life; VABS=Vineland Adaptive Behavior Scale; WISC-V=Wechsler Intelligence Scale for Children, Fifth Edition; WAIS-IV=Wechsler Adult Intelligence Scale, Fourth Edition.
Note: Pearson correlation was used to describe the relationship between each neurocognitive test. The correlation coefficient (r), *p*-value, and number of participants are reported for each comparison.
[a] The Cookie Theft Picture Task was administered to participants over 12 years of age.

4. Discussion

This pilot study assessed instruments measuring a variety of functional domains as a prelude to designing a clinical trial for the treatment of MMA and PA. The results demonstrate what can and what might not be learned through this approach in selecting tests and questionnaires that have the potential to serve as neuropsychological endpoints in a clinical trial.

Here we hoped to 1) identify instruments that could yield a broad range of scores reflecting the abilities of a heterogenous sample of individuals with MMA and PA and 2) assess domains that could be reliably measured, even in individuals with cognitive, language or motor challengesIf during the pilot stage, test scores are similar for all participants, who probably differ in terms of treatment history, severity of the condition, or other parameters, it is unlikely that the test will be sensitive enough to detect differences among groups in the clinical trial.

Table 5
Comparison of neurocognitive tests and questionnaires with perceived severity.

	WISC-V/ WAIS-IV Verbal Comprehension Composite Score	WISC-V/ WAIS-IV Processing Speed Composite Score	VABS: Communication Score	VABS: Adaptive Behavior Composite Score	Cookie Theft Picture Task: Total Number of Words	MetabQoL Questionnaire: Total Score	CASC: Total Score
Severity[a]	9.59	7.18	5.78	7.55	1.53	0.61	0.56
f-value	0.0028	0.0079	0.0122	0.0045	0.2736	0.5676	0.5922
p-value	16	16	20	20	11	10	12
n							

CASC=Child and Adolescent Sleep Checklist; MetabQoL = Metabolic Quality of Life; VABS=Vineland Adaptive Behavior Scale; WISC-V=Wechsler Intelligence Scale for Children, Fifth Edition; WAIS-IV=Wechsler Adult Intelligence Scale, Fourth Edition.

[a] Analysis of Variance (ANOVA) was used to assess the relationship between severity (Mild, Moderate, or Severe) and each neurocognitive test and questionnaire. The f-value, p-value, and number of participants in each comparison are reported.

In this cohort, the MetabQoL and the CASC did not meet these criteria, since scores for only a few participants differed from normative samples.

We also found that on most measures, scores from MMA andPA cohorts differed. The median scores on the WISC-V/WAIS-IV for participants with PA were lower than for participants with MMA. The same was true for VABS scores. This result could have occurred due to sample bias. Nonetheless, it will be important to determine whether these subgroups differ at baseline prior to analyzing results from a clinical trial.

The finding that the perceived severity scores from caregivers and participants correlated with Wechsler verbal and processing speed indices and with the VABS Communication and Composite scores, suggests that these tests hold promise for assessing individuals with MMA and PA in a clinical trial.

Language deficits and significant motor deficits impacted performance. In terms of applicability, this study revealed that a primary verbal outcome measure would be appropriate for a subsample of individuals with basic language skills. Due to significant language deficits, 2/18 (11.1%) participants failed to complete any item on the Wechsler Vocabulary and Similarities subtests comprising the Verbal Comprehension Score. However, all participants ages 12 years and older were able to complete the Cookie Theft Picture Task. Thus, verbal discourse remains a potential endpoint for participants in this age group, although additional studies are needed to validate its use in this population.

Due to significant motor deficits, 2/18 (11.1%) participants were unable to complete any item on the Wechsler Coding and Symbol Search subtests, comprising the Processing Speed Score. These pencil and paper tasks, although focused on measuring processing speed, require significant fine motor abilities. Computerized processing speed tasks, such as those from the Cambridge Neuropsychological Test Automated Battery (CANTAB) [30] or the National Institute of Health (NIH) Cognition Toolbox [31] would be more appropriate for this population. These tasks generally require participants to simply press a button or tap a computer screen, which was well within the functional capacities of participants in our study.

Data collection from the questionnaires was 100% successful, although samples were small for the child-focused sleep and quality of life questionnaires. The VABS-3, administered via the standardized interview format, was well accepted and provided a rich source of information.

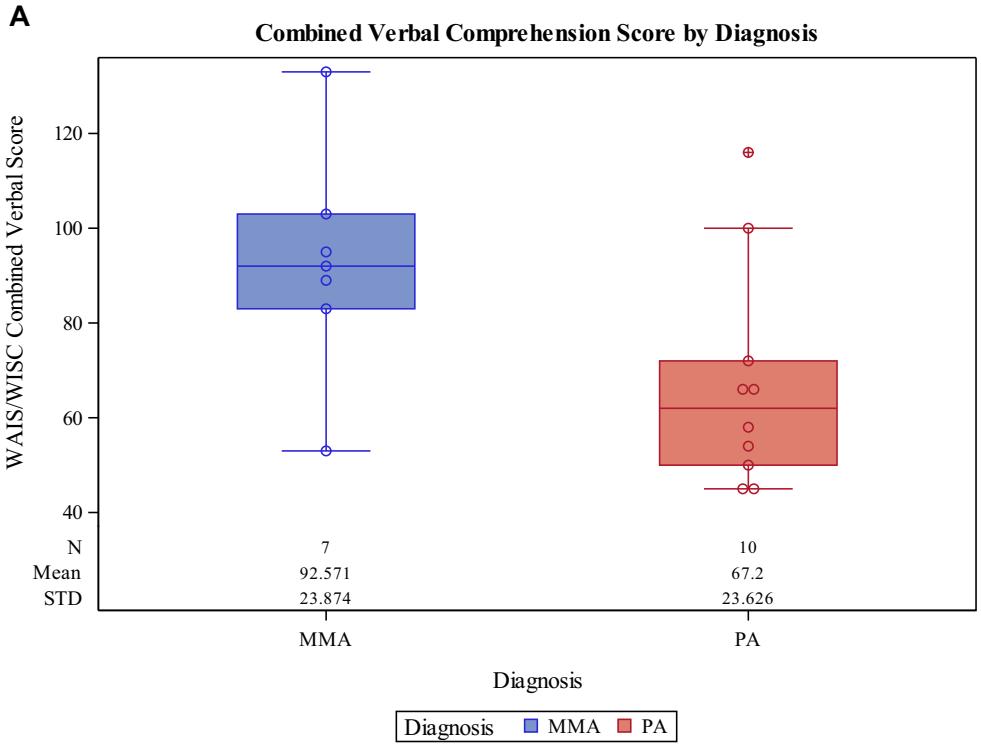
The range of scores for all the tests directly measuring cognitive abilities appeared adequate to allow for detection of higher and lower functioning. The MetabQoL, completed by parents/caregivers for all but 3 of the older children, produced scores that were generally above the population mean (indicating a good quality of life) and only 2 items were endorsed as more than sometimes bothersome. On the CASC, the mean and median scores were well below the cut-off indicating sleep problems. Therefore, these questionnaires are unlikely to be useful in a clinical trial since it is difficult to show improvement when baseline scores are within the normative range.

Scores from the various tests and questionnaires showed strong intercorrelations. The high correlations with the VABS Adaptive Behavior Composite Score reflect the global nature of this score, comprised of scales for communication, daily living, and socialization skills. Despite the limited variability of scores (as reflected by the small standard deviation), the MetabQoL also correlated with scores from the other tests. This may again reflect the broad nature of items included in this questionnaire. Scores from the VABS and Wechsler scales were associated with perceived severity of condition.

As expected, the score from the Cookie Theft Picture Task tended to be correlated with the other verbal tasks. However, it was not correlated with severity, which might reflect the more subtle nature of verbal discourse. Individuals participating in the longitudinal studies of Alzheimer's disease were found to have underlying deficits in discourse long before any signs of dementia were noted [24].

Until recently, little attention has been given to linguistic features characterizing spontaneous speech in individuals with metabolic disorders. However, this may be a domain that is sensitive to modest changes in toxic metabolite levels on a day-to-day basis and may have an important influence on day-to-day functioning. Specifically, linguistic features of verbal discourse have the potential to serve as markers that reveal subtle symptoms in MMA and PA that may not be readily apparent. Thus, verbal discourse may provide a unique measure of functional change in clinical trials. Studies in linguistics reveal that such speech involves motor abilities as well as coordination and speed of cognitive processes and rapid use of working memory for syntax production [24,32]. These include domains noted as particularly affected in MMA [7] and PA [5]. For this pilot study, responses to the Cookie Theft Picture Task were scored manually. Recently, machine learning has been used in identifying linguistic parameters that predicted later onset of Alzheimer's disease in cognitively healthy adults [25]. Another study using machine learning identified individuals with mild aphasia compared to a healthy control group, with a predictive value of 91.67% [26]. Moreover, scores on the Cookie Theft Picture Task did not correlate with simple language tasks, such as confrontation naming or verbal fluency. Linguistic markers derived from machine learning also differentiated patients with Parkinson's disease from controls [32] and detected ON/OFF medication states in people with Parkinson's disease [33]. This suggests that subtle differences in verbal discourse related to brain chemistry can be identified within the reasonable time frame of a clinical study. If true in MMA and PA, verbal discourse measures have the potential to be powerful tools in clinical trials. For this study, we simply used “number of words” to characterize the response to the picture since this variable was associated with overall discourse scores in one study [26].

Comparisons between the MMA and PA cohorts revealed striking differences in scores on the Wechsler Verbal and Processing Speed Tests as well as the Cookie Theft Picture Task. The VABS Communication and Adaptive Behavior Composite Scores were also significantly different, indicating greater challenges in every domain for individuals with PA. Moreover, only children with PA received scores indicating sleep disturbances of below normative levels and lower quality of life. As noted in the recent revisions of the guidelines for the diagnosis and management of MMA and PA, treatments leading to reductions of metabolic



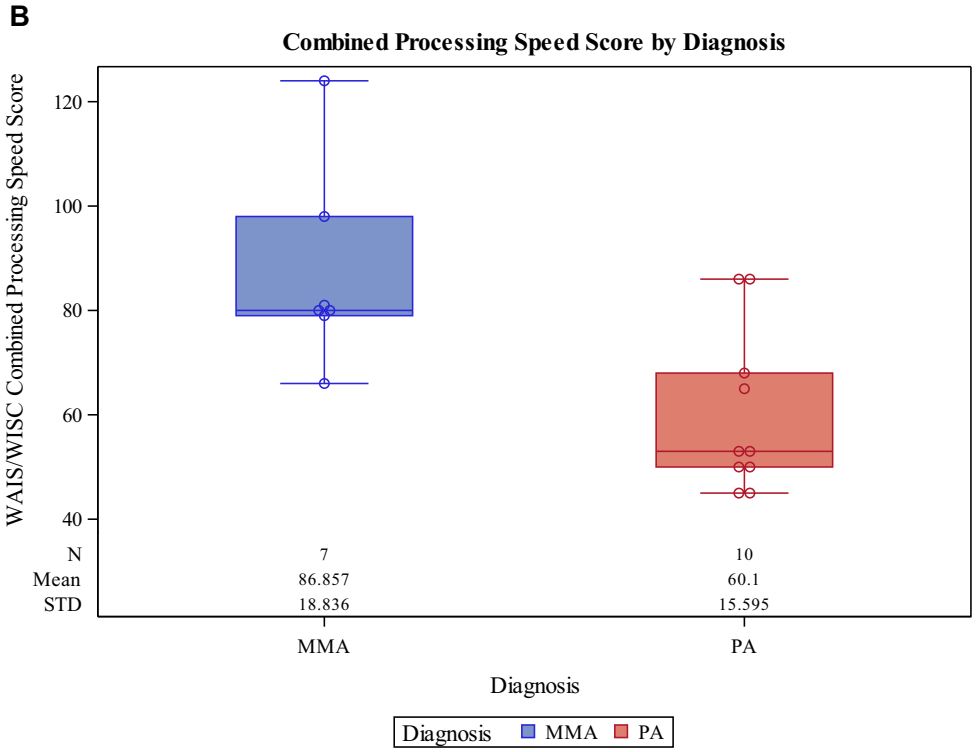


Fig. 1. WAIS-IV/WISC-V Combined Verbal Comprehension and Processing Speed Scores by Diagnosis (MMA or PA).
(A) Combined scores for WAIS-IV verbal comprehension composite with WISC-V verbal comprehension composite in MMA (excluding MMA-CblC) and PA. A *t*-test was used for comparison of scores by diagnosis ($p = 0.0495$). (B) WAIS-IV and WISC-V processing speed scores for MMA (excluding MMA-CblC) and PA. A t-test was used for comparison of scores by diagnosis ($p = 0.0099$). Data for MMA-CblC participants are included in the supplemental material.

episodes may mitigate the neurocognitive impact of MMA, while early intervention and treatment have heretofore had little impact on cognitive outcomes in PA [34]. The panel of experts publishing these guidelines offered no explanation for this discrepancy in outcomes. Nonetheless, these findings should be considered when selecting endpoints for clinical trials.

5. Conclusions

This exploratory pilot study focused only on a few neuropsychological tests. In addition, the small cohort of participants in the MMA and PA cohorts and their disparate ages limit the generalizability of the results. Remote administration and self-reported perceptions of severity further limit the strength of the results. Despite these limitations, the study demonstrates that pilot studies can detect instruments that may

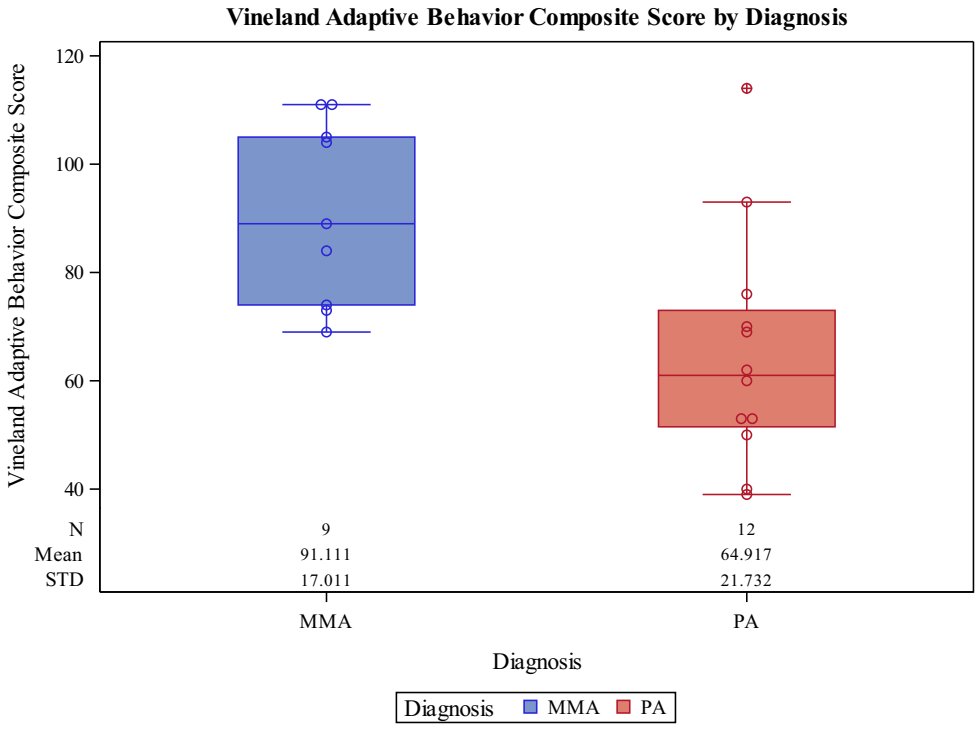


Fig. 2. VABS-3 Adaptive Behavior Composite Score by Diagnosis (MMA or PA).
Vineland Adaptive Behavior – 3rd Ed. Adaptive Behavior Composite Scores in MMA (excluding MMA-CblC) and PA. A t-test was used for comparison of scores by diagnosis ($p = 0.0059$). Data for MMA-CblC participants is included in the supplemental material.

not be appropriate for individuals with language or motor deficits and that may not provide a broad range of scores reflecting disease severity. It also provides a rationale for focusing on discrete neuropsychological domains since some aspects of functioning were less affected than others and some were more closely related to disease severity. When global measures are used overall scores may mask specific deficits. A pilot study like this one cannot ensure that scores will change over time in response to a specific treatment in a clinical trial. However, it can avert the selection of instruments that do not show associations with severity or biomedical parameters likely to be the target of a clinical trial. A pilot study can also identify when differences in diagnoses and baseline functioning need to be addressed prior to developing the analytical plan for the trial.

Author statement

KAC played a role in conceptualization, methodology, investigation, resources, Writing (original draft, reviewing and edition), visualization, and supervision.

DM played a role in conceptualization, data curation, formal analysis, investigation, writing (review and editing), and visualization.

GFC played a role in conceptualization, resources, review and editing, visualization, supervision, funding, and project administration.

MW played a role in conceptualization, resources, review and editing, visualization, supervision, funding, and project administration.

JF played a role in investigation, resources, data curation, review and editing.

SG played a role in methodology, conceptualization, data curation, formal analysis, and investigation.

MS played a role in methodology, conceptualization, data curation, formal analysis, and investigation.

SW played a role in conceptualization, methodology, investigation, resources, Writing (original draft, reviewing and edition), visualization, and supervision.

Declarations

GFC, MW, and JF all employees of HemoShear Therapeutics who also paid for the testing material and hired ProMetrika to do statistical analysis which was provided by DM, SG, and MS. SW is reimbursed as a consultant to HemoShear Therapeutics for her time. KAC is a collaborator and received no reimbursement for time for this study.

Data availability

Data will be made available on request.

References

[1] F. Hörster, A.T. Tuncel, F. Gleich, T. Plessl, S.D. Froese, S.F. Garbade, et al., Additional Contributors from *E*-IMD Delineating the clinical spectrum of isolated methylmalonic acidurias: cblA and mut, J. Inherit. Metab. Dis. 44 (1) (2021 Jan) 193–214, https://doi.org/10.1002/jimd.12297. Epub 2020 Sep 15. PMID: 32754920.

[2] L. Pena, B.K. Burton, Survey of health status and complications among propionic acidemia patients, Am. J. Med. Genet. A 158A (7) (2012 Jul) 1641–1646, https://doi.org/10.1002/ajmg.a.35387. Epub 2012(Jun 7). PMID: 22678880.

[3] J. Schreiber, K.A. Chapman, M.L. Summar, N. Ah Mew, V.R. Sutton, E. MacLeod, et al., Neurologic considerations in propionic acidemia, Mol. Genet. Metab. 105 (1) (2012 Jan) 10–15, https://doi.org/10.1016/j.ymgme.2011.10.003. Epub 2011 Oct 19. PMID: 22078457.

[4] C. Dionisi-Vici, F. Deodato, W. Röschinger, W. Rhead, B. Wilcken, Classical' organic acidurias, propionic aciduria, methylmalonic aciduria and isovaleric aciduria: long-term outcome and effects of expanded newborn screening using tandem mass spectrometry, J. Inherit. Metab. Dis. 2–3 (2006) 383–389, https://doi.org/10.1007/s10545-006-0278-z (Apr-Jun;29). PMID: 16763906.

[5] S.C. Grünert, S. Müllerleile, L. De Silva, M. Barth, M. Walter, K. Walter, et al., Propionic acidemia: clinical course and outcome in 55 pediatric and adolescent patients, Orphanet J Rare Dis. 8 (2013 Jan 10) 6, https://doi.org/10.1186/1750-1172-8-6. PubMed PMID: 23305374; PubMed Central PMCID: PMC3568723.

[6] S.E. Waisbren, Review of neuropsychological outcomes in isolated methylmalonic acidemia: recommendations for assessing impact of treatments, Metab. Brain Dis. 37 (5) (2022 Jun) 1317–1335, https://doi.org/10.1007/s11011-022-00954-1. PMID: 35348993.

[7] C.J. O'Shea, J.L. Sloan, E.A. Wiggs, M. Pao, A. Gropman, E.H. Baker, et al., Neurocognitive phenotype of isolated methylmalonic acidemia, Pediatrics. 129 (6) (2012) e1541–e1551.
[8] B. Han, W. Nei, M. Sun, Y. Liu, Z. Cao, Clinical presentation, molecular analysis and follow-up of patient with mut methylmalonic acidemia in Shandong province, China, Pediatr Neonatol. 61 (2) (2020 Apr) 148–154.
[9] S.B. van der Meer, F. Poggi, M. Spada, J.P. Bonnefont, H. Ogier, P. Hubert, et al., Clinical outcome of long-term management of patients with vitamin B12-unresponsive methylmalonic acidemia, J. Pediatr. 125 (6:1) (1994) 903–908.
[10] F. Sheikhmoonesi, A. Shafaat, S. Moarefian, T. Zaman, Affective Disorder as the First Manifestation of Methymalonic Acidemia: A Case Report, 2013.
[11] M. Nizon, C. Ottolenghi, V. Valayannopoulos, J.-P. Arnoux, V. Barbier, F. Habarou, et al., Longterm neurological outcome of a cohort of 80 patients with classical organic acidurias, Orphanet J Rare Dis. 8 (2013) 148–159.
[12] P. Witters, E. Debbold, K. Crivelly, K. Vande Kerckhove, K. Corthouts, B. Debbold, et al., Autism in patients with propionic acidemia, Mol. Genet. Metab. 119 (4) (2016 Dec) 317–321, https://doi.org/10.1016/j.ymgme.2016.10.009. Epub 2016 Oct 31. PMID: 27825584.
[13] C.D. de la Bâtie, V. Barbier, V. Valayannopoulos, G. Touati, A. Maltret, A. Brassier, et al., Acute psychosis in propionic acidemia: 2 case reports, J. Child Neurol. 29 (2014) 274–279.
[14] C.D. de la Bâtie, V. Barbier, C. Roda, A. Brassier, J.B. Arnoux, V. Valayannopoulos, et al., Autism spectrum disorders in propionic acidemia patients, J. Inherit. Metab. Dis. 41 (4) (2018 Jul) 623–629.
[15] T. Shuaib, N. Al-Hashmi, M. Ghaziuddin, E. Megdad, D. Abebe, A. Al-Saif, as., Propionic acidemia associated with visual hallucinations, J. Child Neurol. 27 (2012) 799–803.
[16] Pascoal C, Brasil S, Francisco R, Marques-da-Silva D, Rafalko A, Jaeken J, et al. Patient and observer reported outcome measures to evaluate health-related quality of life in inherited metabolic diseases: a scoping review. Orphanet J Rare Dis. 2018 Nov 2813(1):215. doi: https://doi.org/10.1186/s13023-018-0953-9. PMID: 30486833; PMCID: PMC6263554.
[17] K. Splinter, A.K. Niemi, R. Cox, J. Platt, M. Shah, G.M. Enns, et al., Impaired health-related quality of life in children and families affected by methylmalonic acidemia, J. Genet. Couns. 25 (5) (2016 Oct) 936–944, https://doi.org/10.1007/s10897-015-9921-x. Epub 2015(Dec 14). PMID: 26667650.
[18] D. Wechsler, Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V), NCS Pearson, Bloomington, MN, 2014.
[19] D. Wechsler, Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV), NCS Pearson, Bloomington, MN, 2008.
[20] S.S. Sparrow, D.V. Cicchetti, C.A. Saulnier, Vineland Adaptive Behavior Scales, Third Edition (Vineland-3), NCS Pearson, Bloomington, MN, 2016.
[21] C.M. Doyen, M.J. Oreve, E. Desailly, V. Goupil, K. Zarca, Y. L'Hermitte, et al., Telepsychiatry for children and adolescents: A review of the PROMETTED project, Telemed. J. E Health 24 (1) (2018 Jan) 3–10.
[22] H. Goodglass, The Assessment of Aphasia and Related Disorders, Third edition, Pro-ed, Austin, TX, 2001.
[23] A. Steinberg, P.D. Lyden, A.P. Davis, Bias in stroke evaluation: rethinking the cookie theft picture, Stroke. 53 (6) (2022 Jun) 2123–2125, https://doi.org/10.1161/STROKEAHA.121.038515. Epub 2022 May 6. PMID: 35514285.
[24] K.D. Mueller, B. Hermann, J. Mecollari, L.S. Turkstra, Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks, J. Clin. Exp. Neuropsychol. 40 (9) (2018 Nov) 917–939, https://doi.org/10.1080/13803395.2018.1446513. Epub 2018(Apr 19). PMID: 29669461; PMCID: PMC6198327.
[25] E. Eyigoz, S. Mathur, M. Santamaria, G. Cecchi, M. Naylor, Linguistic markers predict onset of Alzheimer's disease, EClinicalMedicine 28 (2020 Oct 22) 100583, https://doi.org/10.1016/j.eclinm.2020.100583.
[26] J.A. Matias-Guiu, P. Suárez-Coalla, M. Yus, V. Pytel, L. Hernandez-Lorenzo, C. Delgado-Alonso, et al., Identification of the main components of spontaneous speech in primary progressive aphasia and their neural underpinnings using multimodal MRI and FDG-PET imaging, Cortex. 146 (2022 Jan) 141–160, https://doi.org/10.1016/j.cortex.2021.10.010. Epub 2021 Nov 13. PMID: 34864342.
[27] N.A. Zeltner, M.R. Baumgartner, A. Bondarenko, R. Ensenauer, D. Karall, S. Kolker, et al., Development and psychometric evaluation of the MetabQoL 1.0: A quality of life questionnaire for pediatric patients with intoxication-type inborn errors of metabolism, JIMD Rep. 37 (2017) 27–35.
[28] Y. Oka, F. Horiuchi, T. Tanigawa, S. Suzuki, F. Kondo, S. Sakurai, et al., Development of a new sleep screening questionnaire: Child and Adolescent Sleep Checklist (CASC), Japanese J Sleep med. 3 (2009) 404–408.
[29] J.O. de Lira, T.S.C. Minett, P.H.F. Bertolucci, K.Z. Ortiz, Analysis of word number and content in discourse of patients with mild to moderate Alzheimer's disease, Dement Neuropsychol. 8 (3) (2014 Jul-Sep) 260–265, https://doi.org/10.1590/S1980-57642014DN83000010. PMID: 29213912; PMCID: PMC5619403.
[30] T.W. Robbins, M. James, A.M. Owen, B.J. Sahakian, L. McInnes, P. Rabbitt, Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers, Dementia (Basel, Switzerland). 5 (5) (1994) 266–281.
[31] R.C. Gershon, M.V. Wagster, H.C. Hendrie, N.A. Fox, K.F. Cook, C.J. Nowinski, NIH toolbox for assessment of neurological and behavioral function, Neurology. 80 (11 Suppl 3) (2013 Mar 12) S2–S6, https://doi.org/10.1212/WNL.0b013e3182872e5f. PMID: 23479538; PMCID: PMC3662335.
[32] A.M. García, F. Carrillo, J.R. Orozco-Arroyave, N. Trujillo, J.F. Vargas Bonilla, F. Fittipaldi, et al., How language flows when movements don't: an automated analysis of spontaneous discourse in Parkinson's disease, Brain Lang. (2016) 19–28, https://doi.org/10.1016/j.bandl.2016.07.008. Nov;162. Epub 2016 Aug 5. PMID: 27501386.
[33] R. Norel, C. Agurto, S. Heisig, J.J. Rice, H. Zhang, R. Ostrand, et al., Speech-based characterization of dopamine replacement therapy in people with Parkinson's disease, NPJ Parkinsons Dis 6 (2020 Jun 12) 12, https://doi.org/10.1038/s41531-020-0113-5. PMID: 32566741; PMCID: PMC7293295.
[34] P. Forny, F. Hörster, D. Ballhausen, A. Chakrapani, K.A. Chapman, C. Dionisi-Vici, M. Dixon, S.C. Grünert, S. Grunewald, G. Haliloglu, M. Hochuli, T. Honzik, D. Karall, D. Martinelli, F. Molema, J.O. Sass, S. Scholl-Bürgi, G. Tal, M. Williams, M. Huemer, M.R. Baumgartner, Guidelines for the diagnosis and management of methylmalonic acidaemia and propionic acidaemia: First revision, J. Inherit. Metab. Dis. 44 (3) (2021 May) 566–592, https://doi.org/10.1002/jimd.12370. Epub 2021 Mar 9. Erratum in: J Inherit Metab Dis. 2022 Jul;45(4):862. PMID: 33595124; PMCID: PMC8252715.