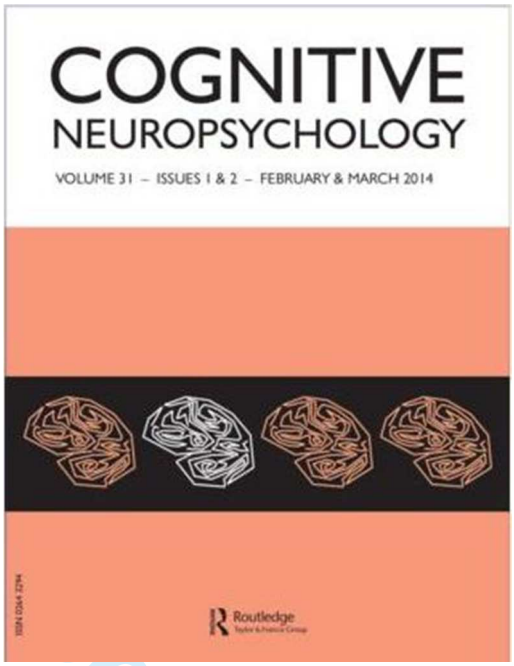


Markers of cognitive function in patients with metabolic disease: Morquio Syndrome and Tyrosinemia Type III

Journal:	*Cognitive Neuropsychology*
Manuscript ID	CN-FA 38.17.R1
Manuscript Type:	Full Article
Keywords:	Tyrosinemia, inherited metabolic disease, language, attention, developmental disorder, MPS-IVa

Markers of cognitive function in patients with metabolic disease: Morquio Syndrome and Tyrosinemia Type III

James Blundell[1], Dr Steven Frisson[1], Dr Anupam Chakrapani[2], Shauna Kearney[2], Dr Suresh Vijay[2], Professor Anita MacDonald[2], Professor Paul Gissen[3], Dr Chris Hendriksz[4], Dr Andrew Olson[1]

[1]University of Birmingham, Birmingham, UK

[2]Birmingham Children's Hospital, Birmingham, UK

[3]Great Ormond Street Hospital, London, UK

[4]University of Pretoria, Steve Biko Academic Unit, Pretoria, South Africa

Acknowledgements. The authors would like to thank all of the patients and their families who supported this research, the many school children who participated in the control sample and the teachers and school administrators who made this possible. We gratefully acknowledge the support of an ESRC CASE studentship award to James Blundell in collaboration with the MPS Society and by the Birmingham Children's Hospital Charities award BCHRF 230.

Correspondence concerning this article should be addressed to Dr James Blundell, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

Abstract

We characterise cognitive function in two neurodegenerative metabolic diseases where cognitive effects have been considered mild or non-existent. We asked whether cognitive effects were present and whether the cognitive effects of neurodegeneration were hetero- or homogeneous across diseases. Thirteen Morquio syndrome patients, 11 Tyrosinemia Type III patients and 104 controls were assessed using tasks for attention (simple RT, feature and conjunction visual search), language (BPVS, BNT) and oculomotor function (fixation, pro-saccade, anti-saccade, smooth pursuit). There were different patterns of deficits. In Morquio syndrome, visual search was slow, differences did not increase in proportion to display sizes and were not driven by the longest reaction times in the distirbution. In addition, maintaining gaze on a target in an oculomotor task was difficult. These patterns point to problems with control processes for attention. Language was relatively spared. In Tyrosinemia Type III, in contrast, there were modest effects in attention, but language was clearly affected, and effects increased with age. Both diseases presented clear evidence of cognitive effects, but with different functional profiles. The different markers (attentional control in Morquio and language in Tyrosinemia) are good candidates for disease tracking and for establishing functional/biological links and they suggest that there may be stronger impacts in different regions of the brain (frontal vs temporal areas).

Keywords: Morquio, Tyrosinemia, inherited metabolic disease, language, attention, developmental disorder

Markers of cognitive function in patients with mild metabolic disease: Morquio Syndrome and Tyrosinemia Type III

Inherited metabolic diseases (IMDs) are large and heterogeneous class of genetic disorders that are caused by dysfunction within a single pathway of intermediary metabolism. In the majority of these diseases the dysfunction of metabolic enzymes leads to the accumulation of toxic metabolites, disrupting the normal development of multiple systems. The severity of symptoms associated with IMDs can vary widely. Mild symptoms can include physiological abnormalities such as skeletal dysplasia and impaired endurance (Davison, Kearney, & Horton, 2013; Wraith, 2006). Severe consequences include mental retardation, central nervous system (CNS) complications, and reduced life expectancy (Bendadi et al., 2014; De Laet et al., 2011; Masurel-Paulet et al., 2008; Thimm et al., 2011, 2012). Research into the cognitive impact of IMDs has largely been limited to standardised intelligence tests, achievement tests and adaptive behaviour scales (Bax & Colville, 1995; Biernacka, Jakubowska-Winecka, & Tylki-Szymanska, 2010; Davison et al., 2013; Shapiro et al., 2009), that are not suitable for characterising impairments in specific cognitive domains or accurately tracking impact over time (Martin et al., 2008). In the current study we present results from two IMD groups, Morquio syndrome and Tyrosinemia type III, where cognitive impairments have been considered mild (using standardised tests). We evaluate Morquio and Tyrosinemia type III patients against the developmental trajectories defined by a sample of typically developing controls (Thomas, Annaz, & Ansari, 2009) to characterise specific patterns of affected and preserved cognitive abilities. We concentrate on the domains of language, attention and oculomotor control.

Morquio Syndrome (MPS-IVa , OMIM 253000)

Morquio syndrome (MPS-IVa) is lysosomal storage disorder that is caused by the deficiency of the lysosomal enzyme N-acetylgalactosamine-6-sulfase (GALNS, EC 3.1.6.4; encoded by GALNS gene at 16q24.3) which has a role in the sequential degradation of the glycosaminoglycans (GAG) keratan sulphate and chondroitin-6-sulfate (Neufeld & Muenzer, 2001; Wraith, 2006). Both keratan sulphate and chondroitin-6-sulfate are essential constituents of connective tissue, including cartilage and vessel walls. The resultant accumulation of these two GAGs leads to a classic phenotype

defined by severe skeletal dysplasia, hip dysplasia, marked short stature, genu valgum, and cornea clouding (Hendriksz et al., 2013; Wraith, 2006). Treatment to reduce substrate burden in MPS-IVa include enzyme replacement therapy (ERT) and haematopoietic stem cell transplantation (HSCT) as a means to alleviate the majority of the skeletal and coronary complications.

In contrast to other lysosomal storage disorders (e.g. Hurler-Scheie syndrome, Hunter's syndrome, Niemann-Pick Type C), MPS-IVa patients have been reported to have no neurological or neurocognitive impairments (Dvorak-Ewell et al., 2010; Wraith, 2006) and neuroimaging results typically find no neuroanatomical abnormalities (Koto, Horwitz, Suzuki, Tiffany, & Suzuki, 1978). However, recent neurocognitive findings (Davison et al., 2013) from 8 MPS-IVa patients (aged 5 -17 years) suggested that mild / borderline cognitive impairments do exist. Age appropriate standardised intellectual tests (e.g. WASI, WISC) revealed full scale IQ scores either in the lower average range (80 – 90), borderline range (70 – 80), or extreme low range (<70) in 4 patients. The remaining 4 patients obtained normal Full Scale IQ scores (85 – 115). Attention problems in the majority of children were reported by parents using the Child Behavioural Checklist.

Mild cognitive impairments in MPS-IVa were further supported by MRS findings in the same study. In 3 patients with cognitive impairments there was a correlation between white matter metabolite concentrations (N-acetylaspartate) and cognitive indices. In addition, MRI findings revealed neuroanatomical abnormalities which were apparent in more than half the patients. These included mild asymmetry of the lateral ventricles, prominent perivascular spaces and high signal white matter areas of the right frontal lobe. Unlike the MRS findings, there was no correlation with cognitive indices, but this could be because the behavioural measures lacked sensitivity. A formal assessment of attention is needed, especially in the context of the attentional difficulties reported by parents.

Tyrosinemia Type III (T3, OMIM 276710)

Tyrosine is an amino acid that is catabolised through the tyrosine metabolic pathway into fumarate and acetoacetate, both of which are important for gluconeogenesis and ketogenesis. Dysfunction at one of the sequential enzymatic reactions of this pathway will lead to one of the 3 identified hypertyrosinemia disorders: tyrosinemia types I to III, which result in the accumulation of plasma tyrosine levels and increased urine excretion of tyrosine (Chakrapani, Gissen, & Mckiernan, 2012). For

example, tyrosinemia type I is caused by the deficiency the fumarylacetoacetate hydrolase, the final enzyme in the tyrosine catabolic pathway, which results in an accumulation of the highly toxic metabolites fumaryl- and maleylacetoacetate in the liver. As a result, patients experience organ dysfunction and carcinogenesis, with hepatocellular carcinoma a frequent cause of death in childhood. Tyrosinemia type II, results from a defect in tyrosine transaminase, the first enzyme in the tyrosine pathway and effects, in particular, the eyes and skin, but also mental development.

Tyrosinemia type III (T3) is caused by the deficiency of 4-hydroxyphenylpyruvate dioxygenase, the second enzyme within the tyrosine metabolic pathway. There is an extreme accumulation and increased excretion of tyrosine, but not an increase in the same toxic metabolites (fumarylacetoacetate) that occur from enzyme defects later in the pathway (Chakrapani et al., 2012; Scott, 2006). Treatment of T3 patients consists of a low-protein diet and administration of ascorbic acid to control tyrosine levels (Scott, 2006). In contrast to tyrosinemia type I, T3 is a more benign form, not associated with the liver or renal complications which lead to reduced life expectancy.

T3 is the rarest of the three tyrosinemia disorders and the effects of elevated concentrations of tyrosine in the central nervous system are not well established. Only 15 cases have been reported in the literature to date (Ellaway et al., 2001; Heylen et al., 2012; Szymanska et al., 2015). However, neurological and intellectual difficulties are commonly reported characteristics. A review of T3 patients by Ellaway et al. (2001) provides the best description of cognitive functioning for a collection of T3 patients (13 patients). Eight out of 13 patients demonstrated neurological symptoms such as developmental delay or mental retardation, attention deficit and behavioural disturbance, acute ataxia, tremor, hypotonia and absent deep tendon reflexes. The most common long-term complication was intellectual impairment, found in 75% of reported cases (Ellaway et al., 2001). It was speculated that impairment results from neurotoxic effects linked to elevated tyrosine levels. The results, however, are based on aggregated scores from tests that are not designed to distinguish specific types of cognitive impairment (e.g. e.g. Stanford-Binet, WISC, WASI), so further research is required to determine how specific cognitive capacities are affected.

Further evidence of cognitive impairment in T3 comes indirectly from research with T1 patients who have been treated with 2-nitro-4-trifluromethylbenzoyl (NTBC). NTBC gives T1 patients the

biochemical profile of T3 patients by inhibiting the enzyme 4-hydroxyphenylpyruvate dioxygenase. This greatly increases T1 patient survival by preventing the subsequent accumulation of toxic metabolites that affect the liver and renal systems, but elevates tyrosine concentrations as a side-effect. T1 is a more common form of tyrosinemia and, in contrast to T3, more detailed reports of neurocognitive function in this group exist (Bendadi et al., 2014; Thimm et al., 2011; Van Ginkel et al., 2016) Using age-appropriate test batteries, different studies have documented lower IQ scores and specific language production and comprehension impairments in NTBC-treated T1 patients (Bendadi et al., 2014; Thimm et al., 2012). The most detailed description of cognitive function in T1 was recently provided by (Van Ginkel et al., 2016), measuring nineteen NTBC-treated patients with both age-appropriate psychometric tests (e.g. WISC and WASI) and additional measures of executive functioning. Lower average IQ performance (median 85: range 55 – 111) were reported. Executive function tests revealed specific working memory deficits (in reaction time and error rates) in T1 patients when compared to age-matched controls. Inhibition was also assessed but no differences were found between patients and controls. The cognitive deficits from the T1/NTBC studies could be due to the elevated neurotoxic levels of tyrosine that are a side effect of NTBC therapy, but they could also be a direct effect of NTBC (Van Ginkel et al., 2016). Results from T3 patients will help to clarify this issue.

Based on these previous findings (Bendadi et al., 2014; Ellaway et al., 2001; Thimm et al., 2012; Van Ginkel et al., 2016), we expect to see some cognitive impact in tyrosinemia type III resulting from high levels of tyrosine. As with Morquio Syndrome, however, we will characterise, in more detail, capacities which are affected and preserved.

Methods

Participants

Morquio Syndrome (MPS IVa) and Tyrosinemia Type III (T3)

Thirteen patients diagnosed with Morquio syndrome (8 male; mean age: 9.59 years, range: 5.27 – 14.39 years) and eleven patients diagnosed with tyrosinemia type III (3 male; mean age: 12.53 years, range: 4.40 – 19.58 years) were recruited at Birmingham Children's Hospital, UK (demographics in Appendix 1). Diagnosis in Morquio patients was confirmed via genetic testing, all were considered to have a severe phenotype of the disease, but no patient presented with corneal clouding and none was

ventilated. Diagnosis of Tyrosinemia Type III was confirmed via genetic testing after elevated tyrosine levels were detected during newborn screening. All patients were native English speakers. A number were bilingual (see Appendix 1). Both patient groups were tested in 2012-13 at the Wellcome Trust Clinical Research Facility at the Birmingham Children's Hospital. Consent was obtained from parents/guardians prior to testing.

Typically Developing Controls (TD)

To compare patients with typically developing controls (TD), we used a developmental trajectory approach. We tested a large sample of controls across a range of ages between 6 and 20 to establish the trajectory of typical development in each cognitive domain (Attention tasks N = 104, Language tasks N = 104, Oculomotor tasks N = 265). Prior to testing, informed consent was taken from undergraduate student participants or from parents of child participants.

Apparatus & Procedure

We used a battery of neuropsychological tests in three cognitive domains: language, attention and oculomotor control. Patients took took approximately 1 – 1.5 hours to complete the battery, depending on ability and need for breaks.

Attention and oculomotor function were assessed with computerised tasks created using Experiment Builder (SR Research Ltd., Mississauga, Ontario, Canada). Participants viewed stimuli from 60 cm on a 36 x 27 cm CRT-monitor with a resolution of 1024 by 768 pixels and a refresh rate of 60Hz, producing a viewing area of 33.5° x 25.5° (width x height) of visual angle. Reaction time (RT, measured in milliseconds, msec) for manual responses were recorded using a Cedrus button box (http://cedrus.com/). Language tasks were table-top tests.

Eye movements were recorded via an EyeLink® 1000 Tower Mount (SR Research Ltd., Ontario, Canada). The location and duration of fixations are measured using corneal reflection via an infrared camera. Head movement was minimized using a forehead and chin rest. Eye movements were calibrated to an accuracy of at least 1^{o} using a nine-point calibration array. Drift correction was employed before each trial. Participants were recalibrated if measured fixation of a central point was inaccurate. Saccade detection was based on default Eyelink 1000 settings (19-sample window): samples were classified as part of a saccade if eye velocity exceeded 22^{o}/second and eye position changed more

than 0.3 °, otherwise samples were classified as fixations. Prior to analysis trials from all tasks were visually inspected to ensure that participants were engaged in the task and artefacts had been removed.

Attention tasks

Attention was measured with a simple reaction time and visual search task. We familiarised participants with the visual search targets (red ladybirds) by introducing them in the simple reaction time task, which was always presented first.

For simple reaction time we measured speed to respond to appearance of a visual target. Participants completed 20 trials where a red ladybird (5.5° x 7.5°) was presented in a box in one of the four quadrants around the centre (upper left, upper right, lower left, lower right; box size = 10° x 9°; 12° diagonal distance from screen centre). The target appeared in each quadrant 5 times, with the order of the target locations randomised between participants. Each trial began with a centrally presented fixation cross (1° x 1°), after a variable delay of 1500 – 4000ms the red ladybird target stimulus was presented in one of the four quadrants. Participants were instructed to respond with their preferred hand, using a single button press, as quickly as possible after the target appeared. The stimulus remained on the screen until the participant made a response or for 3000ms. There was an inter-trial interval of 1000ms prior to the presentation of the fixation cross for the following trial.

The visual search task consisted of 3 feature search blocks and 3 conjunction search blocks. In all blocks participants indicated whether the red ladybird target character was present with a "Yes" button-press, or absent using a "No" button-press. Feature search asked the participants to search for the red ladybird among a set of green ladybird distracters. The target was defined by colour only. Conjunction search asked participants to search for a red ladybird among green ladybirds and red beetle distractors. The target is defined by a *conjunction* of colour and shape, since neither identifies the target in isolation (Eckstein, 2011; Treisman & Gelade, 1980). Each block contained 12 trials with 3 display sizes (4, 8, and 12 items). The target was present in half the trials in each block, and absent in the other half. Block order was randomised between participants. Search displays were created by dividing the screen into a 4 x 4 grid (each grid location = 8.5° x 6.5°). Search items (4° x 5° of visual angle) were randomly assigned to the 16 grid locations. In order to make displays less regular, each item's position

was jittered by a random amount (x maximum ±2° and y maximum ±1°). For target-present conjunction trials one red beetle distractor was replaced with the target to ensure the number of red and green elements were equal.

Each trial began with a centrally presented fixation cross for 1000ms, followed by the stimulus. Participants used a Cedrus button box to make a response (left button = target absent; right button = target present). Stimuli disappeared after a response or if no response was registered within 10 seconds. A blank screen was displayed for 1000ms prior to the following trial. At the beginning of each block participants were informed of the block type (feature or conjunction). Targets and distractors were shown on screen and verbal instructions given by the experimenter. There were four practice trials with feedback, at the beginning of the experiment, to familiarise participants with the task.

Response times for correct responses were used to calculate an overall visual search response time and a measure of visual search efficiency. Search efficiency was defined as slope of the change in RT with increasing display size, i.e. the change in response speed as a function of display size. When search efficiency is zero, RT does not increase with display size. Values above zero indicate that additional items in larger displays take more time. Feature search normally has a search efficiency near zero because targets "pop out" regardless of the number of distractors.

Language tasks

Measures of verbal production (Boston Naming Task or BNT; Kaplan, 2000) and comprehension (British Picture Vocabulary Scale or BPVS; Dunn et al., 1997) were used to assess children's language abilities.

The BNT is a 60-item picture naming test with items that increase in difficulty. If a participant was unable to name an item, a semantic cue was offered. A phonetic cue was provided if the semantic cue failed to produce a correct response. An item was counted correct if the participant was able to produce the name with or without a semantic cue, and incorrect if the participant failed to produce the name or produced it after a phonetic cue. The test finishes if participants make more than 5 consecutive errors.

The BPVS is a receptive vocabulary test. Participants listen to a word and select a matching picture from four line drawings. There is no spoken response. There were 14 sets of 12 items which increase in difficulty. The test is terminated if there are eight or more errors on a single set.

Oculomotor function tasks

Oculomotor function was tested using two paradigms: a fixation and pro-saccade task (Fischer & Weber, 2010; Klein, 2001; Munoz & Everling, 2004; Salman et al., 2006). These tasks assess distinct components of the saccadic eye movement system and frontal executive functions that contribute to oculomotor control (e.g. inhibition and sustained attention during fixation).

Fixation Task

In the fixation task, participants were asked to fixate a central target, move their eyes to follow the target when it moved to one of four possible positions and then maintain their gaze until the target disappeared. There were 20 trials. An elephant face target (1.5 ^{o} in size) appeared in the centre of the screen and then moved to a location 10^{o} to the left, right, above or below the central fixation point. Possible target locations were marked with small circles (0.5 ^{o} in size) to indicate where the target might appear. Trials began with the presentation of the target centrally for 1000 milliseconds (msec). The target disappeared and immediately reappeared (no gap or overlap) randomly at one of the four surrounding locations. The target remained at this location for 5000ms and participants were asked to maintain fixation on the target until it disappeared. A 3 x 3° box surrounding each target was defined as a region of interest (ROI) and fixations within the ROI were counted as fixations on the target. Dwell time was defined as the length of time participants maintained eye position within the target's ROI (*FixDwell*). The frequency of saccades away from the target (greater than 2°) that moved the participant's gaze outside the target ROI were also measured (*FixSacc*).

Pro-saccade Task

Participants viewed 48 trials where a target elephant face (1.5^{o} in size) randomly appeared at one of eight positions around a central starting position. The four locations from the fixation task were used (see above), along with four additional locations, closer to the centre (eccentricity = 5^{o}). Each trial began with the stimulus appearing at the centre of the screen for a random amount of time between 1000 and 2000ms. The stimulus then disappeared and reappeared at one of the possible target locations

for 1000ms without any gap or overlap. Participants were asked to look as quickly and accurately as possible directly at the target.

To be included in the analysis, saccades had to start at the centre of the screen and the eye movement had to be toward the peripheral target. We measured the onset of the first movement toward the target and its peak velocity. Only data from targets at 10^{o} are presented here.

Data Analysis

Measures that were greater than 3 SDs away from individual participant means were defined as outliers and removed. Patients and typically developing (TD) control children were compared using a developmental trajectory approach (Thomas et al., 2009). Developmental change was modelled as a linear or quadratic function of age. We were interested in differences in the *rate* of developmental change (differences in the slope of trajectories) and differences in absolute levels of performance (offsets between patient and control trajectories). We fit linear mixed effects models (LME) with fixed factors for *Group*, *Age* and *Condition* (e.g. target location). Main effects of *Group* or *Condition* indicated higher or lower performance across all ages. Interactions with *Age* indicated that the rate of developmental change differed either between patients and controls (e.g. *Group* X *Age*), or between conditions (e.g. *Condition* X *Age*), or that there were condition-specific trajectory differences between patients and controls (e.g. *Group* X Age^2 X *Condition*). Akaike's Information Criterion (AIC) was used to compare models (Burnham & Anderson, 2002). AIC is preferred for model comparison because, unlike likelihood ratio p-values, AIC balances fit and the number of model parameters when choosing models.

Model selection using AIC is different from p-values, but not difficult to understand. In brief, better models produce smaller AIC values, but the absolute AIC values are not interpretable. Instead, the change in AIC (ΔAIC) between models is meaningful and captures the weight of evidence for each model (rather than being subject to a cut-off, like p-values). Evidence for a model starts to be clear if the ΔAIC exceeds 2. If ΔAIC between the "best" model and alternative models is less than 2 then the two models are substantially equivalent. When ΔAIC is between 2 and 10 there is decreasing support for an alternative model. A model with a ΔAIC > 10 has essentially no support. For models where the ΔAIC is less than 2, it is reasonable to favour the least complex model (i.e. model with fewest

parameters/variables). Comparisons can be assisted by calculating Akaike weights (AICw; Burnham & Anderson, 2002). AIC*w* expresses the relative probability that a model is the best *in a particular set*, considering only the models from that set. It measures the weight of evidence for the models being compared. When values are relatively equal across two or more models, they are all relatively good models of the data. If one model has a high value and the others are low, there is a model that is clearly better. Favoured models contain the terms that would be significant in a traditional analysis. For example, if a highly-rated model has a term for *Group* but no interaction, this corresponds to a traditional analysis where the main effect of *Group* is significant and there is no interaction.

We report results from a large sample of controls and a much smaller sample of patients, as in inevitable in the study of rare diseases. This means that models that do not involve the *Group* term are largely determined by the data from normally developing control participants. We model control data without patients initially to characterise normal development. When we report results from patients and controls we will concentrate on evaluating the influence of the term for *Group*.

Because of possible heterogeneity among patients within groups we also compared individual patients to typically developing controls (TD). The question was whether individual patients were inside or outside the range of typically-developing values. To define the range of typically developing values, ninety-five percent prediction intervals were defined at each age, setting a point above and below the TD means. Ninety-five percent of individual control values should fall between these points. Cutoffs for each age were smoothed to a boundary that applied across ages by fitting a curve to the set of upper or lower boundary points. This method minimises the impact of idiosyncratic estimates, smoothing the boundary position across ages by borrowing information from adjacent age groups. After defining limits, individual patients were compared to controls using the equivalent of z-scores, but using the upper or lower endpoints of the smoothed prediction intervals, rather than the noisier age-specific prediction interval points. We label these z_{pi}.[1]

Results

[1] Specifically, for values above the mean, $SD_{pi} = (PI_{upper} - mean)/1.96$ and $z_{pi} = Y_{patient}/SD_{pi}$, where PI_{upper}= upper boundary of the smoothed prediction interval; *mean* = the mean predicted by the fit to control values; and $Y_{patient}$ = patient's observed data value. Values below the mean were calculated in the same way except the 95% prediction interval boundary (PI_{lower}) used the lower boundary since boundaries were not necessarily symmetric.

Table 1 summarises domains where Morquio (MPS-IVa) and Tyrosinemia (T3) patients had difficulties (shaded in grey). The number of patients who performed worse than the control mean, worse that 1 SD from the control mean and worse than 2 SD from the control mean are listed. MPS-IVa patients displayed clear deficits on tasks with elements of attentional control: visual search and holding fixation. T3 patients exhibited clear deficits on language tasks (*BNT* and *BPVS*). Mild attention deficits were also observed (*RTmean, VSmean*, *FixDwell)* for T3 patients.

Table 1 about here

Attention tasks - Simple Reaction Time Task

Typically developing Controls (TD)

Reaction times were log transformed because, as is often the case, the raw RT distribution was right-skewed and a log transform produces a more balanced distribution. The best model describing the development of simple reaction time included an effect of age, distinct trajectories for targets presented to the left and right and an interaction. The majority of developmental change occurred during the first years of development, with little change after age 12 (~300 msec over the range from 6 to 12 years). Reaction times for younger children were 51 msec slower when targets were on the right of the screen (Figure 1) and then the difference decreases with age.

Figure 1 about here

Morquio patients (MPS-IVa)

RT means for individual MPS-IVa patients are displayed in Figure 1. They were within the confidence limits of healthy development for individual patients (0/12 with $z_{pi} > 2$; black points inside dashed black line, Figure 1). The group mean was shifted towards slower reaction times, with eleven of 12 patients slower than the control mean, but not far enough to be outside the control range.

In the analysis of group differences a model with a quadratic term for age was used because the developmental trajectory is curved, not linear. The model with an Age^2 *X Condition* interaction was

essentially equivalent to the best model (ΔAIC = 0.23; best model: *Condition X Age*^{2} + *Group X Age*^{2}). Evidence of differences between patients and controls was weak (AIC results for the best models of MPS-IVa patients and controls are summarised in Appendix 2).

Tyrosinemia III

Individual T3 patient RT means are displayed in Figure 2. They were evenly distributed across the control range (solid grey lines) for all target locations.

Figure 2 about here

Only three patients had response times that were reliably slower than controls ($z_{pi} > 2$) when performance was averaged across target locations. These were the two youngest patients (patient 1 - 4.40 years, patient 2 - 5.16 years) and an older patient (patient 9 - 17.64 years). One patient (Patient 11; 19.10 years) was on the border ($z_{pi} = -1.98$).

In group comparisons, the best model included a *Condition* X *Age* X *Group* interaction (AIC results for the best models for Tyrosinemia III patients and controls are summarised in Appendix 3). The effect of condition was not particularly strong. A model with just main effects of *Age* and *Group* was nearly as good as the best model (ΔAIC = 1.97). There were, however, strong effects of *Age* and *Group*. The first models without these terms had essentially no support (ΔAIC=11.12 and ΔAIC=17.19). The 3-way interaction occurs because T3 patients are slower than the control mean at the earliest ages, closer to the mean in the middle of the age range and slow again when they are older, but by much less, and then the differences are somewhat different in the different quadrants. These interaction effects are, however, subtle and the small T3 sample makes them uncertain. The effect of group is stronger and occurs because 7/11 patients are slower than the control mean.

Visual Search

Typically developing Controls (TD)

Mean reaction time is plotted in the top panels and search efficiency in the bottom panels of Figure 3. TD controls increased their overall speed with age (top panels) in both conjunction and feature search. The biggest change occurred during the early years, with changes slowing after about

10 years old. Feature search times decreased with age, but were not longer for larger displays (i.e. efficiency did not change). Search efficiencies were close to zero at all ages (although there was some decline in variability). The best model included an effect of age but no effect of display size. The best model of conjunction search included an interaction between display size and age. The time to search each item declined steadily from around 70 to around 32 milliseconds (bottom right panel) over the range between 6 and 19 years. There was also a large drop in variability.

Accuracy supported the same pattern. Accuracy in feature search increased across ages (from 94% at age 6 to 98% at age 19). There was a difference in accuracy on present vs absent trials (90% vs 98% at age 6) that narrowed with age (99% vs 98% at age 19) and models supported an interaction between present/absent and age. There was no effect of display size. The best model of conjunction search accuracy included an interaction between display size and target present/absent and an interaction between age and present/absent. The interaction with age was very weak because a model with just a main effect of age was nearly equivalent (ΔAIC = 1.23). The interaction with display sizes occurs because more targets are missed in larger displays. Accuracy increased with age (85% at age 6 vs 97% at age 19). These patterns do not modify the interpretation of the reaction time results (where larger displays take longer).

Figure 3 about here

Morquio Syndrome (MPS-IVa)

The biggest difference between patients and controls occurred in the simpler task of feature search, although there was also some effect on conjunction search (Figure 3). In feature search, z_{pi} was > 2 in 6 patients and between 1 and 2 in one other patient. Eleven of 12 patients were slower than the control mean. Six would be expected by chance. In conjunction search, z_{pi} was > 2 in 3 patients and between 1 and 2 in 4 others. Ten of 12 patients were slower than the control mean.

General slowing was not usually associated with larger effects of display size. In feature search, searching additional items took longer in three individuals (Patient 1, z_{pi} = 2.46; Patient 8, z_{pi} = 5.99; patient 11, z_{pi} = 11.25), and, in conjunction search, in two (Patient 8, z_{pi} = 2.57; Patient 11, z_{pi} = 3.55).

In sum, a substantial subset of the patient group were generally slower than controls, but only a much smaller subset produced disproportionate increases in search time as display size increased (2-3/12). For most patients the increase with display size was normal even if overall search times were shifted upward by a constant amount.

Group differences between MPS-IVa patients and controls were analysed separately for feature and conjunction search. The best model of feature search did not include an effect of *Display Size* in either patients or controls and there were no two or three-way interactions. The lack of a *Display Size* effect means that usually the search for a feature difference happens in parallel across all locations in the display no matter how many items there are to search, which is the normal pattern in feature search (Treisman & Gelade, 1980). There were main effects for *Age* and *Group.* Younger participants were slower. The main result, for our purposes, was that patients were slower than controls (by 445ms). Terms for *Group* and *Age* were clearly necessary because models without these terms were poor.

There was no interaction between *Group* and *Age*, which means that, despite general slowing, patients make progress at a normal rate, neither catching up with controls or falling behind. We need to be somewhat cautious about this result, however, because of the small number of patients at each age.

In the analysis of accuracy, *Group* was added to the *Age* by *Present/absent* interaction from the control model. Instead of the increasing accuracy for both present and absent trials seen in controls, MPS-IVa patients initially show a bias to "present" and later a bias to "absent" with little consistent change to percent correct. This is not a theoretically important pattern in the present context. Averaging over present/absent conditions, the best model included an interaction between *Age* and *Group* that resulted from increasing overall accuracy in controls, but flat or slightly decreasing accuracy in patients. Patients and controls had equivalent accuracy when averaging across ages (both 94%). The accuracy results do not change the interpretation of reaction times. Accuracy is high in both patients and controls and there was no indication that longer RTs were associated with higher accuracy in the patient group ($R = -0.46$, $p=.13$; a negative R value results when higher accuracy is associated with *lower* RT).

The top four models of conjunction search were all very similar in their ability to account for reaction times (Display size X Age + Group X Age; Display size + Group X Age; Display size X Age + Group; Display Size + Age + Group; max change in AIC = 0.8). Good models all included terms for

Display Size, *Age* and *Group*. Where there was an interaction between *Group* and *Age,* this resulted from MPS-IVa patients who were slower than controls being concentrated at older ages, but the effect was based on small numbers. Some good models also included an interaction between *Display Size* and *Age* based on more efficient search at older ages. All these interactions, however, were small effects because the simpler model with only main effects was a close equivalent (ΔAIC=0.8). It is important to note that there was no interaction between *Display Size* and *Group*. Patients were as quick as controls to process each item, even when efficiencies were above zero, which shows that each item *did* require additional processing time. The display-size effect for both patients and controls was ~42 ms per item. The *Group* effect occurs because patients are ~378 ms slower than controls (for both large and small displays). Models without main effects of *Display Size*, *Age* or *Group* were poor.

This overall pattern shows that patients were generally slower than controls, but differences were not concentrated on larger displays, which means patients did not have slower item-by-item processing.

Once again, analysis of accuracy did not change this pattern. The best model of accuracy was not different from the control model and did not include a term for *Group*.

In sum, MPS-IVa patients were clearly slower in visual search. This difference, however, was constant across display sizes. Patients did not process each item in the search displays more slowly. The upward shift in reaction times without display size effects could have at least two origins. One possibility is that patients occasionally lose concentration or are distracted. If so, reaction times at the fast end of the range should be similar to controls, but there should be more difference in slower reaction times. Lapses could be distributed across trials with small and large displays. A second possibility is that a stage that is always present, but not sensitive to the attentional demands of the display, is affected (e.g. a decision stage is one possibility). This would result in higher reaction times across the full range of fast and slow times, not just a difference at the slow end of the scale.

To explore these alternatives, we binned reaction times into quintiles for each participant and compared quintiles for patients and controls using the vincentizing method (Ratcliff, 1979; see Romani, MacDonald, De Felice, & Palermo, 2017 for this kind of analysis in a population of PKU patients). Figure 4 shows reaction times and standardised differences between means for patients and controls in

each quintile. To standardise, we divided the difference between patient and control means by the standard error of the control mean. Because reaction time distributions have an extended tail associated with the longest reaction times, the variance in the final bin, especially, is higher than in the other bins, and analysis of quintile differences needs to take this increased variance into account. We will concentrate two aspects of the difference curves. We ask if they are sloped upwards, particularly in quintiles 2, 3 and 4, showing increasing differences with longer reaction times. We will also look at where patients means are slower than controls.

Figure 4 about here

The two possible sources of longer reaction times are well illustrated by comparing MPS-IVa and Tyrosinemia patients, as we will see below. With two exceptions, the pattern in MPS-IVa patients is for relatively stable differences across the full range of quintiles, not increasing differences. This is the case for feature search at ages 5-8, conjunction search at ages 8-10.5 and both feature and conjunction search at ages 10.5-14.5. Conjunction search at ages 5-8 were not reliably different from controls in any bins (all $t < 1.8$, all $p > 0.08$). Feature search for ages 8-10.5 is the only condition where there were steadily increasing differences across the range of RTs. Apart from the two exceptions noted, patient reaction times were always slower, even in the fastest bins ($t > 2.3$; $p < .03$) and not consistently increasing. The overall pattern does not support distraction/failures of attentional control or the accumulating effects of general slowing as the primary source of longer reaction times. Instead, there appears to be a problem with processing at a stage that is not sensitive to the attentional demands of the task. A decision stage is one possibility.

Tyrosinemia III (T3)

In feature search, four of eleven T3 patients had $z_{pi} > 2$. An additional two patients had z_{pi} between 1 and 2 (Figure 5). In conjunction search, five of eleven patients had $z_{pi} > 2$ and two patients had z_{pi} between 1 and 2. Search efficiency slopes had $z_{pi} > 2$ in only one patient and between 1 and 2 in three others.

Figure 5 about here

Group differences were analysed separately for feature and conjunction search. The best model for feature search included an *Age* by *Group* interaction. There were more patients who were further from controls at older ages. There was a ~76 ms difference between groups when participants were under 12 years old, with as many patients below the control mean as above it. Patients older than 15 years were, on average, ~187 ms slower than controls, and all four patients were above the mean. There was no effect of *Display Size*, indicating that feature search was efficient for both groups.

The best model for accuracy in feature search included an interaction between *Present/absent, Age* and *Group,* adding *Group* to the interaction from the control model. As with MPS-IVa patients, Tyrosinemia III patients show a different pattern of *Present/absent* effects compared to controls. Accuracy is relatively constant, but *Present* trials show some decrease, mainly due to results from one patient. What is more relevant, for our purposes, is that, overall, patients were less accurate than controls (89% vs 93%). There was no evidence of a speed/accuracy trade-off that would modify interpretation of the reaction time results.

In conjunction search, the best model of reaction times included main effects of *Display Size*, *Group* and *Age*, but no interactions. The group difference was an average of 71 msec when participants were below 12 years old and 133 ms when participants were above 15 years old. The lack of interactions with group shows that effects of *Display Size* and *Age* were not reliably different in patients compared to controls.

Models of accuracy in conjunction search added an interaction with *Group* to the *Display Size* by *Age* by *Present/absent* interaction from the control model. Present/absent differences were unstable in patients because of the small number of individuals and they are not theoretically important. We accumulated over present/absent conditions and put patients into only two age groups (less than or greater than 12 years old) to increase N. Models still required a *Group* by *Display size* by *Age* interaction. Tyrosinemia patients were generally less accurate than controls (80% vs 90%) and the display size effect was larger in younger patients only (accuracy dropped by 12% when young patients viewed larger displays, but only by 1-3% in older patients, young controls and old controls).

Results from analysis of reaction time quintiles showed that differences between patients and controls concentrated on longer reaction times (Figure 6). At ages 4.5-6, differences in both search tasks increased in bins 1-3 and then levelled out. At ages 6-8.5 patients were not slower than controls. At ages 15-19, differences increased across all bins in both search tasks. In the youngest and oldest patients, reaction times were different even in the first bin (all $t > 2.3$; $p < .03$), showing that patients in these groups were always slower, but the differences increased for longer reaction times. Different patterns at different ages should be interpreted with caution because the number of patients at each age is small and individual vs. age-related differences cannot be clearly separated.

Figure 6 about here

In sum, considering both accuracy and reaction time together, effects on accuracy were more pronounced for younger patients, but reactions times slowed more in older patients. Differences in visual search, in other words, were sometimes observed more in accuracy and sometimes more in reaction times, but there was a relatively consistent difference.. Turning to the level of individual patients, there was considerable variability. Some patients showed clear differences, and others did not. Seven of 11 individuals recorded reaction times that were slower than controls in feature search and seven of 11 in conjunction search (not always the same individuals). This is not a marked difference from controls because the value expected by chance is 5.5. By way of contrast, in MPS-IVa feature search the whole group was shifted up in relationship to control values, with nearly all MPS-IVa patients above the control mean. There was some indication that Tyrosinemia III patients get worse with age, but this was a preliminary result that should be treated with caution. It was based on small numbers, so that patient-specific effects will be difficult to distinguish from a general effect of age.

Language

Typically developing Controls (TD)

Control participants showed clear effects of age in both the BPVS ($\Delta AIC = 163$) and the BNT ($\Delta AIC = 58$). There was improvement across the range of ages tested (Figure 7).

Figure 7 about here

Morquio Syndrome (MPS-IVa)

All MPS-IVa patients completed the British Picture Vocabulary Scale (*BPVS*). Patient 12 did not complete the Boston Naming Test (*BNT*) because of time constraints. Individual *BPVS* and *BNT* scores are shown in Figure 7. Two individual patients had z_{pi} that was reliably different from controls on the Boston Naming Test; one below the control mean (patient 6; $z_{pi} = -3.17$) and one above (patient 7; $z_{pi} = 2.74$). Three other patients had z_{pi} between -1 and -2. The best model for the Boston Naming Test included only a term for changes with age. There was no evidence that MPS-IVa patients produced systematically lower BNT scores than controls.

No *BPVS* scores had $z_{pi} < -2$, but 7/12 patients were between -1 and -2 (Figure 7). The best model for *BPVS* scores included an interaction between *Group* and *Age*, but evidence for the interaction was weak because the model without the interaction was nearly equivalent ($\Delta AIC = 0.7$). The estimated score difference between Morquio patients and controls was -9.4 points (based on the model without an interaction).

In sum, MPS-IVa patients showed no difference from controls in productive vocabulary and only minor differences in receptive vocabulary with very few or no patients outside the control range on either task.

Figure 8 about here

Tyrosinemia III (T3)

Just over half of the T3 patients were more than 2 z_{pi} below controls on the *BNT* (6/10 patients). The best model for *BNT* results included an interaction between *Age* and *Group*. The difference between patients and TD controls widened with age (Figure 8). The difference at the earliest ages was 18, but widened to 56 in the oldest patients.

The majority of patients were more than 2 z_{pi} below controls on the *BPVS* ($z_{pi} < -2$; 8/11 patients). Similar to *BNT* results, the best model included an interaction between *Age* and *Group*. Once

again, the difference between T3 patients and TD controls got larger with age (Figure 8). At the earliest ages (6 years) the estimated difference was 32 points. By 19 the estimated difference had grown to 63 points. Models without terms for either *Age* or *Group* were poor.

Unlike MPS-IVa, where the differences between controls and patients were small, T3 patients show clear problems with language tasks that become more marked with age.

Oculomotor Tasks

Oculomotor tests included a fixation task and a saccade task. Morquio patients 5 and 7 did not want to attempt the saccade task and patients 2, 3, 4 and 11 did not complete all the saccade trials due to fatigue. T3 patient 2 was unable to complete the oculomotor tasks due to postural difficulties which prevented eye tracking.

Fixation Task – Dwell Time

Typically Developing Controls

Typical development of fixation dwell time was described best by a quadratic model that included an interaction between *Condition* and *Age*. Systematic checking of location revealed that the model was improved by having a distinct trajectory for the top target condition, with a shared trajectory for left, right and bottom targets (Separate Condition Model, ΔAIC = 0.87; AIC*w* = 0.39 ; Top Condition Model, ΔAIC = 0.00; AIC*w* = 0.61). This means that typically development occurs at a different rate for top targets in comparison to targets presented in the other three locations. Figure 9 shows that is largely due to differences between target conditions at the youngest ages (5 – 6 years) where participants are able to hold fixation on top targets longer than left, right and bottom targets (max 4345 ms top vs min 3903 bottom).

Figure 9 about here

Morquio Patients (MPS-IVa)

The majority of MPS-IVs patients exhibited fixation dwell time deficits ($z_{pi} < -2$) for the left (7/11 patients), right (8/11), and bottom target locations (9/11; Figure 9). Both younger and older patients had difficulties maintaining fixation. Fewer patients produced clear deficits for the top target location (4/11). However, even in this condition all patients were below the TD mean (solid line, Figure 9). When fixation time was averaged across the 4 target locations, 8/11 patients exhibited clear fixation duration deficits ($z_{pi} > 2$).

The best model describing group differences between MPS-IVa and TD groups was a model that included all 2-way interactions: *Group* X *Age, Group* X *Condition,* and *Condition* X *Age*. The *Condition* X *Age* interaction was already apparent in the TD data. The *Group X Condition* interaction reflects differences between quadrants that were larger for MPS-IVa patients than controls. Control differences were relatively modest (maximum difference, top vs bottom = 160 msec), but patient differences were more pronounced (maximum difference, also top vs bottom, 906 msec). The *Group* X *Age* interaction occurred because there were larger differences between groups at younger ages, but the model without this interaction was nearly equivalent, so its influence is weak (ΔAIC=2.1). The strongest difference, which is also reflected in model values, however, is that patients held fixation on the target for considerably less (1082 msec) than would be predicted for controls, and a model without the *Group* term was poor (*Condition X Age* model, ΔAIC = 55).

Figure 10 about here

Tyrosinemia Type III (T3)

T3 patient fixation dwell times are presented in Figure 10. As a group, the patients were shifted below the TD mean (Average $z_{pi} < 0$ in 9/10 patients), but fewer patients were clearly outside the control range (Average $z_{pi} < -2$ in 2/10 patients). Shorter fixations were more common to top and bottom targets (4 / 10 patients $z_{pi} < -2$), but the short fixation times were not always in the same individuals.

The best model for data from patients and TD controls included 2 two-way interactions: *Condition* X *Age* and *Condition* X *Group*. The *Condition* X *Group* interaction reflects differences between quadrants that were larger in patients (maximum difference of 158 msec in controls compared to 528 msec in patients). In general, the differences between T3 patients and controls were smaller and less prevalent compared to MPS-IV. In fact, if the three T3 patients with the shortest fixation times are removed, the T3 patient group is not different from controls (the best model includes a *Condition* X *Age* interaction, but no term for *Group*). When these patients are included, however, there is a group effect (model without Group, ΔAIC = 17). Together, both the individual and group results show that most T3 patients were not different from controls, but there were some exceptions.

Fixation Task – Intrusive Saccades

Typically Developing Controls

The frequency of intrusive saccades during the period in which participants were supposed to hold fixation was defined by a quadratic function. Figure 11 shows that intrusive saccades decrease with age, with the majority of age-related change occurring before 10 years. A model which included a *Condition* X *Age* interaction (ΔAIC / AIC*w* = 0 / .99) was clearly better than a model that did not (ΔAIC / AIC*w* = 8 / .01). Examination of systematically grouped trajectories revealed that, similar to the related measure of fixation time, there was a distinct developmental trajectory for the top target location (Top vs others model, ΔAIC / AIC*w* = 0.00 / 0.54; separate quadrant model, ΔAIC / AIC*w* = 0.38 / .44). The average number of intrusive saccades at 6 years was fewest for the top target location (*top* position, 1.72 fixations, vs maximum of 2.03 fixations for the *right* position). The difference between locations was smaller by 15 years (*right* position, 0.92, vs maximum of 1.07 fixations for the *left* position).

Figure 11 about here

Morquio syndrome (MPS-IVa)

MPS-IVa patients often exhibited more intrusive saccades than TD controls (Figure 11). Five of the 11 patients were more than 2 z_{pi} from controls. These 5 patients, not surprisingly, also had shorter fixation durations. Fixation durations would be expected to be shorter when there are intrusive saccades. Three patients, however, did not have deficits in both measures: patient 2 (6.44 years), 7 (9.84 years) and 8 (11.71 years). They were within the control range for intrusive saccades but not fixation time. This means the first fixation on the target was shorter than in controls, but there were not more saccades in general, and patients did not go back and forth between the target and other locations. They held fixation for a shorter time on the target and then moved away. Sample plots from horizontal and vertical trials in two patients and age-matched controls are shown in Figure 12. These illustrate the saccade and fixation deficits reported in these patients.

Figure 12 about here

The best model included an interaction between *Age* and *Condition,* and a main effect of *Group*. Intrusive saccade suppression was poorer in MPS-IVa patients compared to TD controls, regardless of the target location. On average, MPS-IVa patients made 0.67 more intrusive saccades than would be predicted for TD controls.

Figure 13 about here

Tyrosinemia III (T3)

Individual T3 intrusive saccade data is presented in Figure 13. Unlike MPS-IVa, intrusive saccade frequencies were normal in the majority of T3 patients (9/10 patients z_{pi} range -2 to 2). One older patient (patient 9) had more substantial problems across target locations and, therefore, for performance collapsed across locations (z_{pi} = -5.64). Importantly, patient 9 also exhibited fixation time deficits, and it is clear from the eye movement traces that fixation on the target was disrupted by intrusive saccades made towards the screen centre (Figure 14).

Figure 14 about here

T3 and control fixation count data was best explained a model with *Group* X *Age* and *Condition* X *Age* interactions, but the weakness of the group effect is evident, because a model with only a *Condition* X *Age* interaction is essentially equivalent (ΔAIC = .05). The *Group* X *Age* interaction occurs because the young T3 patients make somewhat fewer saccades than controls, while the older T3 patients make somewhat more. The *Condition* X *Age* interaction was evident in the control data.

Pro-saccade Task – Saccade Onset time

Typically Developing Controls

Figure 15 plots saccade onset times for TD controls (solid grey line). Similar to simple reaction times, saccade onsets decreased with age and the majority of change occurred between 5 and 12 years. Saccade onset times were log transformed because residuals for raw values increased at higher values. There was a clear preference for a model that included an *Age* X *Condition* interaction. Inspection of systematic differences between target locations revealed that development was best described by a model with independent trajectories for top and bottom target locations, and a shared trajectory for horizontal target locations. Bottom target onsets were longest at 6 years (max bottom, 307 ms vs min right, 247 ms). At older ages the differences narrow and the order changes (max top, 186 ms vs min right, 161 ms at 19 years).

Figure 15 about here

Morquio patients (MPS-IVa)

Individual MPS-IVa saccade onset values are shown in Figure 15. Averaged across target locations, all patients had z_{pi} less than 2. Borderline deficits were found for 3 patients on left targets (patient 2, 10 and 11) and 2 patients for top targets (patients 8 and 11). The best model for saccade onset included an interaction between *Group* and *Condition*, in addition to the interaction between *Age* and *Condition* that was noted in controls. The interaction results from control/patient differences that differ

by quadrant, although the means across quadrants are very similar (patients 19 msec faster than controls in the bottom quadrant and 34 msec slower in the left quadrant; overall means differ by less than 1 msec). There was no evidence for a main effect of group independent of the interaction with condition. The model with just an *Age* X *Condition* interaction was better than the model that included an *Age* X *Condition* interaction and a main effect of *Group* (ΔAIC = 1.39).

Figure 16 about here

Tyrosinemia type III

Individual T3 saccade onset values are shown in Figure 16. Averaged across target locations, only one patient (patient 9, 17.98 years) had a $z_{pi} > 2$. Patient 9 also had fixation time and intrusive saccade deficits. Models of saccade onset did not support differences between patients and controls. The best model did not include a *Group* term.

In sum, neither patient group was systematically slower to initiate saccades.

Pro-saccade Task – Saccade Peak Velocity

There was no evidence of differences in saccade velocity when either patient group was compared to controls. The preferred model for MPS-IVa patients and controls included a *Condition* X *Age* interaction, but no effect of group. For T3 patients, the model with only a *Condition* X *Age* interaction was essentially equivalent to the best model, which was more complex and, as a result, less preferred (ΔAIC=0.6; best model = *Condition* X *Age* + *Condition* X *Group*).

Discussion

We examined cognitive functioning in two rare inherited metabolic diseases – Morquio syndrome (MPS-IVa) and tyrosinemia type III. We compared patients to controls using formal tools for model selection and a developmental trajectory approach. This provides a richer way of evaluating developmental differences than traditional age-matched comparisons. We were able to separate different types of developmental effects, including changes to the *rate* of development and constant

decrements in performance in the presence of normal rates of developmental change. We included both analysis of group performance and individual comparisons (whether individuals are inside or outside the control range defined by model prediction intervals), which contribute different, but complimentary information (see discussion below).

In both diseases we found cognitive effects, but with different functional profiles. Attentional control was consistently affected in Morquio patients and language in Tyrosinemia. These are good candidates for sensitive markers of disease progression and they document selective functional impacts that can be integrated with descriptions of biological change. A summary of effects across cognitive domains is shown in Table 1.

MPS-IVa (Morquio Syndrome)

Few studies have investigated cognitive performance in Morquio syndrome (MPS-IVa), primarily because clinical observations report normal intellectual function (Dvorak-Ewell et al., 2010; Wraith, 2006). A recent study, however, reported mild cognitive changes (Davison et al., 2013) and highlighted difficulties with attention. We did more detailed cognitive testing and found attention deficits in several tasks. Fixation was disrupted by intrusive saccades and visual search times were slower, but without a decrease in search efficiency, which is consistent with a problem at a control stage rather than a problem in allocating attention during the search process. An analysis of reaction time distributions showed that the problem was not occasional lapses of attention, but slowing that affects both fast and slow reaction times. This implicates a stage that is always present, but is not sensitive to the attentional demands of the display (e.g. a decision stage, rather than a search stage). Two patients had elevated search slopes but most did not. These findings support and expand on the Davison et al. (2013) study, where parents of MPS-IVa patients reported difficulties with concentration and attention.

The fixation task, in particular, may index the same phenomena reported by parents, since concentration problems could be the everyday consequence of the same issue that is evident in maintaining attention on a target during formal testing. Intrusive saccades and shorter fixation times may not simply reflect a problem that is isolated to eye-movement control. The cortical areas that are responsible for maintaining fixation (Krauzlis, Goffart, & Hafed, 2017), including areas in dorso-lateral prefrontal cortex, the frontal eye fields and the superior colliculus, are areas that have connections to

other modalities: auditory or proprioceptive, in the case of the frontal eye fields (Medendorp, Buchholz, Van Der Werf, & Leoné, 2011); movement control, in the case of the superior colliculus (Krauzlis et al., 2017). The frontal eye fields, for example, are thought to "encode 'supramodal' representations to guide attention and behaviour" (Medendorp et al., 2011). Problems in these areas would plausibly lead to both difficulties with eye-movement control and to more general attentional issues.

Several areas implicated in fixation are also involved in visual search, where MPS-IVa patients had slowed responses in the presence of mostly normal search efficiencies. The frontal eye fields are associated with the production of exploratory eye movements during search (Booth et al., 2003), and dorsolateral and dorsomedial prefrontal cortex are involved in response decisions (Venkatraman, Rosati, Taren, & Huettel, 2009). The network of common areas would be a candidate system to examine for changes to brain function in MPS-IVa.

In sum, our results highlight difficulties with control aspects of attentional tasks in MPS-IVa. They extend the preliminary results reported by Davison et al. (2012) and they show that the differences in attentional tasks occur without commensurate changes in other cognitive tasks (e.g. language) or general slowing (as reflected in normal or near normal simple reaction time, saccade onset and saccade velocity).

Tyrosinemia III (T3)

In Tyrosinemia III, there were clear deficits in that both language production (*BNT*) and comprehension (*BPVS*), where 6 / 10 (55%) and 8 / 11 (73%) patients were clearly outside the control range and nearly all patients produced low scores compared to controls. These differences appeared to widen with age because the largest differences were found in older patients. There were also some differences in visual search, where overall visual search time was slower in 5 / 11 (45%) patients. The additional time needed for complex displays was not different from controls (like MPS-IVa), but the differences increased for slower reaction times, suggesting occasional attentional failures (unlike MPS-IVa). Differences in the attention tasks were present in the youngest and eldest patients.

To our knowledge, this study represents the most detailed cognitive assessment of T3 patients to date and highlights, especially, language impairments and some attentional effects along

with possible age-related decline. Our results provide formal tests to confirm and extend the case-based findings of Ellaway et al. (2001), who described mild to moderate intellectual impairments in a sample of 12 patients, but without specific information about cognitive testing.

The prevalence language deficits in T3 is interesting since this is a frequently reported cognitive feature of NTBC treated T1 patients (Bendadi et al., 2014; Thimm et al., 2011). NTBC inhibits 4-hydroxyphenylpyruvate dioxygenase, blocking the biochemical pathway in T1 at the same place where it is blocked in T3 and leading to greatly increased blood tyrosine levels. In essence, T1 patients become T3 patients with this treatment, except that T1 patients are also subjected to other potential collateral effects of NTBC.

The biochemical mechanism that explains the cognitive deficits of T3 patients is currently unknown, but in T1 patients, impairments are only present after NTBC treatment, which raises tyrosine levels. Learning impairments are also present in the mouse model of T1 when treated with NTBC, but not in wild type mice treated with NTBC, suggesting that the metabolic result of treating T1 with NTBC causes the impairment, rather than treatment with NTBC directly (Hillgartner et al., 2016). Since T3 results in the same metabolic outcome as T1 treated with NTBC, our results support the same conclusion. In addition, Ellaway et al. (2001) reported that T3 patients who were diagnosed and put on a low tyrosine diet when they were older presented with more severe symptoms than patients who began a low tyrosine diet earlier and they voiced the suspicion that elevated tyrosine levels in cerebrospinal fluid (CSF) are particularly damaging during infancy. Our study did not analyze biochemical data, but some of the variance in the cognitive performance in patients in our sample could result from individual differences in tyrosine levels that, in turn, are based on variations in early exposure or dietary compliance. Relating tyrosine concentrations at different ages to cognitive performance in future work will be an important step to understanding how and when tyrosine affects the brain and when dietary control is critical for effective treatment.

The behavioural consequences for T1 mice treated with NTBC are worth noting in relationship to the language deficits we found. Hillgartner et al (2016) found that NTBC treated T1 mice were slower to learn in a maze task. Both the BNT and the BPVS are tests of vocabulary. That is, they test the facility with which words are stored for the long term in the mental dictionary. Both tests

would be sensitive to problems with word learning. This means it may be that word *learning*, rather than language, per se, is what is impaired in our T3 patients.

The cognitive neuropsychology of metabolic disease

Our results show that neurodegeneration from metabolic disease is not a uniform process. We did not find a homogeneous set of general effects, nor an additive set of effects that increased with severity. Instead, different diseases had specific selective impacts on language or attention. Since the different metabolic diseases affect different chemical systems in the brain, this shows that neuro-cognitive systems are not just differentiated by location. Populations that are critical for different cognitive functions are differentially sensitive to cellular support systems. Cognitive neuropsychological assessment of metabolic disease (or other diseases that affect the cellular environment) can begin to ask "why?" Are there systems, for example, that are particularly sensitive to neural transmission speed, network synchronization, particular neurotransmitters, growth or membrane changes—all properties that would be directly influenced by the biochemical environment? What is the *functional* aspect of the system that makes these characteristics critical? This is a relatively unexplored dimension, within neuropsychology, that is orthogonal to the understanding of anatomical networks. These questions will be critical, however, to understanding the brain as an integrated biochemical and anatomical system and they will be critical to our understanding of disease.

For MPS-IVa, what we can say about the relationship between metabolic processes and cognition is limited because it has typically been described as a condition that doesn't have cognitive impacts. As a result, there has been little effort to explain how a deficiency in *N*-acetylgalactosamine-6-sulfatase (GALNS) might affect the brain. Davison et al. (2013) were probably the first to offer some possibilities. GALNS breaks down karatan sulfate and chondroitin-6-sufate, and these help coordinate neuroaxonal connection formation (Miller, Sheppard, & Pearlman, 1997). In addition, Davison speculate that MPS IVa may affect calcium signalling, and connections between calcium signalling, mitochondria and neurodegeneration have been made for Huntington's disease, amyotrophic lateral sclerosis, Alzheimer's disease and Parkinson's disease (Marambaud, Dreses-Werringloer, & Vingtdeux, 2009). Calcium signalling plays an important role in long term potentiation and excitotoxicity, two different routes to cognitive effects (Marambaud et al., 2009).

In tyrosinemia type III, there is a more interesting story and a more developed context. T3 involves a defect in the enzyme, 4-hydroxyphenylpyruvate dioxygenase, which is needed to break down tyrosine, the precursor of dopamine and norepinephrine. There are several possible ways this might lead to cognitive effects, as outlined by Hillgartner et al. (2016), who set out to explain cognitive effects in NTBC treated patients with tyrosinemia type I (T1; recall that NTBC blocks tyrosine metabolism at the same point affected in T3). With a defective enzyme of tyrosine catabolism, tyrosine accumulates. This could produce several primary changes in the brain: 1) increased tyrosine; 2) increased dopamine (as a result of conversion from tyrosine); 3) decreased large amino acids (like tryptophan, the precursor of serotonin) that are outcompeted by tyrosine; 4) decreased serotonin (due to lack of precursor). Each of these could create its own effects. Hillgartner et al., however, propose that the real source of cognitive problems in treated T1 is not a consequence of any of these primary effects, but via two by-products of defective tyrosine break-down: one that is neurotoxic (succinylacetone) and the other that is both neurotoxic and causes demyelination (δ-ALA).

This is a perfectly reasonable hypothesis for cognitive effects in NTBC treated T1. Hillgartner et al. propose that NTBC is poorly transported into the brain, so that tyrosine catabolism proceeds past the point where NTBC stops it in the body and neurotoxic by-products will result. The problem is that this hypothesis cannot explain cognitive effects in T3. The enzyme that is defective in T3 acts at the same point that NTBC acts, but it will be defective in all cells, including those in the brain. This means the toxic by-products that Hillgartner et al., identify will not be produced in T3, but cognitive decline still occurs. There must be a separate mechanism for cognitive decline to explain T3 (perhaps related to the primary effects), and if this yet-to-be-identified mechanism explains cognitive decline in T3 it could also explain what appear to be similar cognitive problems in T1. The mechanism, however, can not be based on catabolism that proceeds past the point where the defect occurs in T3.

The tyrosinemia study illustrates how specific information about cognitive impairments in metabolic disease (the similarity of impairments in T3 and NTBC-treated T1), together with information about the metabolic pathways affected, might influence our understanding of the link between cognitive and biochemical mechanisms, even if there are many outstanding questions (e.g. Are T1 and T3 really showing the same pattern of cognitive impairments when both are tested in sufficient

detail? What is the link between primary biochemical effects and the specific cognitive problems, concentrated on learning or language, that we observed?)

Clinically, our results, and, more generally, the results from cognitive testing reported in this special issue, are important because they can be used to anticipate the kinds of problems patients and families might expect and can inform choices about how to support patients. Specific cognitive capacities that are impacted by disease are also important to identify when choosing measures to use for monitoring patients or for assessing new treatments. Since oculomotor or attentional tasks show the biggest impacts in MPS-IVa, for example, and language tasks show the largest changes in T3, these would be the best initial candidates for monitoring.

Cognitive neuropsychology of development

Even though detailed cognitive assessment of the populations we have presented is at an early stage, it is clear that the cognitive neuropsychology of development presents both challenges and opportunities (see also, for example, Pitchford & Funnell, 1999). Development involves both changes over time and potential dependencies among changing capacities. Theories of development must specify the time course of intrinsic or environmentally conditioned change, the environmental inputs that are required for development, and potential dependencies among capacities. The pattern of dependencies that support successful development is not guaranteed to persist at the same level, or persist at all, when the system is mature, which means that development cannot be transparently inferred from either the mature system or insults to that system.

The complexity of the relationship between adult and developmental models can be illustrated, in terms that will be familiar to readers of *Cognitive Neuropsychology,* in the domain of reading. It is relatively clear, from cognitive neuropsychological results, that sub-lexical and lexical mechanisms for reading are largely independent components in the mature system and all major reading models include them (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 1999; Zorzi, 2010). The relative independence of these systems in adulthood, however, does not mean that they are independent during development. This is apparent from developmental dyslexia (e.g. Hulme & Snowling, 2016) or more directly, in the context of the current discussion, from acquired dyslexia after a stroke in childhood (Pitchford & Funnell, 1999). As they have in other areas of neuropsychology, the

consequences of brain damage can reveal properties of components and dependencies in a *developmental* (rather than a static) model that are not available in other ways. Although there are many studies of developmental cognitive difficulties, the approach that involves characterising components and dependencies in a detailed *developmental* cognitive model, of the sort that readers of *Cognitive Neuropsychology* would find familiar, is relatively underdeveloped. Our results are too preliminary to show this promise except in outline, but they do show that developmental effects from metabolic disease can be selective enough to allow tests of functional models and they illustrate both some methodological issues and some approaches that offer solutions.

One of the basic problems is that development is a moving target. The statistical models of developmental trajectories that we have introduced here provide some solutions. They allowed us to compare patients and controls across a range of ages when performance is changing and to separate offsets, which create differences across all ages, from changes to developmental rates. Offsets characterised MPS-IVa attentional changes (in both visual search and fixation tasks) while a change to the rate of development characterised T3 language results, where there was age-related decline.

Capturing developmental trajectories becomes an even more powerful tool when longitudinal data are available. Individual patient trajectories can be compared to control trajectories and variance in rates of development in the patient group will show whether there is, for example, consistent decline, or heterogeneity in the effects of disease over time. If patients need to be tracked clinically, individual longitudinal trajectories could conceivably show that patients are not following the normal course of development far before their performance is different from controls at any single point in time, providing a more sensitive test of disease progression.

The issue of sensitivity and the contrast between what is apparent in individual data and group data is already clear with our samples. Effects in some individuals that do not occur in all individuals in the group are common. Data from a group of patients, however, can also show that there are deficits even when the individual data do not. Conjunction search results for the MPS-IVa patients were a case in point. Only 3 of the patients were outside the control range as individuals, but the population, as a whole, was clearly shifted toward slower RTs, with only two patients faster than the control mean. There is a very low probability that so many scores would be slower the control mean in an unimpaired

sample. This is clinically important because it highlights the limits of individual samples for some questions (i.e. is there impairment from disease in this group or not) and also reinforces the point that techniques like longitudinal sampling may be needed to detect decline at the earliest possible moment at the individual level.

The statistical approach we have adopted is a powerful method to deal with developmental data and with relatively small samples. There are, however, limits that small samples, inevitable in rare diseases, impose. We cannot, for example, say very much about whether a group trajectory derived from cross-sectional patient data is a good representation of how diseases progress in *individual* patients, and we have avoided fitting a trajectory to the patient group. In a small sample, with relatively high variability, the precise shape, inclination or location of the trajectory can be highly unstable and shows, again, the need for longitudinal data. When the changes over time are clearer, as in the T3 language data, the statistical models allow clear conclusions based on the interaction between group and age. Even when this is the case, however, the precise location or shape of the trajectory may be relatively uncertain.

Conclusion

Our study introduces new data and methods to illustrate the promise of characterising neuropsychological effects in inherited metabolic diseases. Two patient cohorts showed that the cognitive effects of disease were not homogeneous. Specific measures that were identified as sensitive for each disease could be developed further into new tools that help track disease progress or quantify treatment benefits. In addition, our data are a promising basis for understanding the biochemical and individual factors that influence the severity and time course of developmental effects. Finally, they make several extensions possible. Cognitive performance can be related to changes in structure, activity or biochemistry as measured by MRI, fMRI, MRS or EEG to triangulate the biological basis of cognitive changes. A fuller understanding of population and individual profiles, extending the methods described here to longitudinal data, will be a necessary next step for understanding the cognitive impacts of disease, the best methods for tracking disease progression and the most sensitive indicators of treatment success.

References

Bax, M. C., & Colville, G. a. (1995). Behaviour in mucopolysaccharide disorders. *Archives of Disease in Childhood*, *73*(1), 77–81. https://doi.org/10.1136/adc.73.1.77

Bendadi, F., De Koning, T. J., Visser, G., Prinsen, H. C. M. T., De Sain, M. G. M., Verhoeven-Duif, N., … Van Hasselt, P. M. (2014). Impaired cognitive functioning in patients with tyrosinemia type i receiving nitisinone. *Journal of Pediatrics*, *164*, 398–401. https://doi.org/10.1016/j.jpeds.2013.10.001

Biernacka, M., Jakubowska-Winecka, A., & Tylki-Szymanska, A. (2010). The development of cognitive functions in children with Hurler phenotype mucopolysaccharidosis type I on enzyme replacement therapy with laronidase. *Pediatric Endocrinology, Diabetes and Metabolism*, (1), 249–254.

Booth, J. R., Burman, D. D., Meyer, J. R., Lei, Z., Trommer, B. L., Davenport, N. D., … Mesulam, M. M. (2003). Neural development of selective attention and response inhibition. *NeuroImage*, *20*(2), 737–51. https://doi.org/10.1016/S1053-8119(03)00404-X

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (Vol. 2). New York: Springer.

Chakrapani, A., Gissen, P., & Mckiernan, P. (2012). Disorders of Tyrosinemia Metabolism. In J. . Saudubray, van den Berghe, & J. H. Walter (Eds.), *Inborn Metabolic Diseases: Diagnosis and Treatment* (Fifth Edit, pp. 233–243). Springer. https://doi.org/10.1007/978-3-642-15720-2

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037//0033-295X.108.1.204

Davison, J., Kearney, S., & Horton, J. (2013). Intellectual and neurological functioning in Morquio syndrome (MPS IVa). *Journal of Inherited Metabolic Disease*, *32*(2), 323–328.

De Laet, C., Munoz, V. T., Jaeken, J., François, B., Carton, D., Sokal, E. M., … Goyens, P. J. (2011). Neuropsychological outcome of NTBC-treated patients with tyrosinaemia type 1. *Developmental Medicine and Child Neurology*, *53*(10), 962–4. https://doi.org/10.1111/j.1469-8749.2011.04048.x

Dvorak-Ewell, M., Wendt, D., Hague, C., Christianson, T., Koppaka, V., Crippen, D., … Vellard, M. (2010). Enzyme replacement in a human model of mucopolysaccharidosis IVA in vitro and its biodistribution in the cartilage of wild type mice. *PloS One*, *5*(8), e12194. https://doi.org/10.1371/journal.pone.0012194

Eckstein, M. P. (2011). Visual search : A retrospective. *Journal of Vision*, *11*(5), 1–36. https://doi.org/10.1167/11.5.14.

Ellaway, C. J., Holme, E., Standing, S., Preece, M. a, Green, a, Ploechl, E., … Leonard, J. V. (2001). Outcome of tyrosinaemia type III. *Journal of Inherited Metabolic Disease*, *24*(8), 824–32.

Fischer, B., & Weber, H. (2010). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*(03), 553. https://doi.org/10.1017/S0140525X00031575

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491–528. https://doi.org/10.1037//0033-295X.106.3.491

Hendriksz, C. J., Harmatz, P., Beck, M., Jones, S., Wood, T., Lachman, R., … Tomatsu, S. (2013). Review of clinical presentation and diagnosis of mucopolysaccharidosis IVA. *Molecular Genetics and Metabolism*. https://doi.org/10.1016/j.ymgme.2013.04.002

Heylen, E., Scherer, G., Vincent, M.-F., Marie, S., Fischer, J., & Nassogne, M.-C. (2012). Tyrosinemia Type III detected via neonatal screening: Management and outcome. *Molecular Genetics and Metabolism*, *107*(3), 605–607. https://doi.org/10.1016/j.ymgme.2012.09.002

Hillgartner, M. A., Coker, S. B., Koenig, A. E., Moore, M. E., Barnby, E., & MacGregor, G. G. (2016). Tyrosinemia type I and not treatment with NTBC causes slower learning and altered behavior in mice. *Journal of Inherited Metabolic Disease*, *39*(5), 673–682. https://doi.org/10.1007/s10545-016-9949-6

Hulme, C., & Snowling, M. J. (2016). Reading disorders and dyslexia: *Current Opinion in Pediatrics*, *28*(6), 731–735. https://doi.org/10.1097/MOP.0000000000000411

Klein, C. (2001). Developmental functions for saccadic eye movement parameters derived from pro-and antisaccade tasks. *Experimental Brain Research*, 1–17. https://doi.org/10.1007/s002210100711

Koto, A., Horwitz, A. L., Suzuki, K., Tiffany, C. W., & Suzuki, K. (1978). Morquio syndrome: Neuropathology and biochemistry. *Annals of Neurology*, *4*(1), 26–36. https://doi.org/10.1002/ana.410040106

Krauzlis, R. J., Goffart, L., & Hafed, Z. M. (2017). Neuronal control of fixation and fixational eye movements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160205. https://doi.org/10.1098/rstb.2016.0205

Marambaud, P., Dreses-Werringloer, U., & Vingtdeux, V. (2009). Calcium signaling in neurodegeneration. *Molecular Neurodegeneration*, *4*(1), 20. https://doi.org/10.1186/1750-1326-4-20

Martin, H. R., Poe, M. D., Reinhartsen, D., Pretzel, R. E., Roush, J., Rosenberg, A., … Escolar, M. L. (2008). Methods for assessing neurodevelopment in lysosomal storage diseases and related disorders: a multidisciplinary perspective. *Acta Paediatrica. Supplementum*, *97*(457), 69–75. https://doi.org/10.1111/j.1651-2227.2008.00651.x

Masurel-Paulet, a, Poggi-Bach, J., Rolland, M.-O., Bernard, O., Guffon, N., Dobbelaere, D., … Touati, G. (2008). NTBC treatment in tyrosinaemia type I: long-term outcome in French patients. *Journal of Inherited Metabolic Disease*, *31*(1), 81–7. https://doi.org/10.1007/s10545-008-0793-1

Medendorp, W. P., Buchholz, V. N., Van Der Werf, J., & Leoné, F. T. M. (2011). Parietofrontal circuits in goal-oriented behaviour: Sensorimotor integration in parietofrontal cortex. *European Journal of Neuroscience*, *33*(11), 2017–2027. https://doi.org/10.1111/j.1460-9568.2011.07701.x

Miller, B., Sheppard, A. M., & Pearlman, A. L. (1997). Developmental expression of keratan sulfate-like immunoreactivity distinguishes thalamic nuclei and cortical domains. *The Journal of Comparative Neurology*, *380*(4), 533–552. https://doi.org/10.1002/(SICI)1096-9861(19970421)380:4<533::AID-CNE9>3.0.CO;2-2

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews. Neuroscience*, *5*(3), 218–28. https://doi.org/10.1038/nrn1345

Neufeld, E. F., & Muenzer, J. (2001). The Mucopolysaccharidoses. In C. Scriver, A. Beaudel, W. Sly, & D. Valle (Eds.), *The Metabolic and Molecular Basis of Inherited Disease* (8th Edn, pp. 3421–3452). New York: McGraw-Hill.

Pitchford, N. J., & Funnell, E. (1999). An acquired form of developmental phonological dyslexia. *Cognitive Neuropsychology*, *16*(6), 573–587. https://doi.org/10.1080/026432999380735

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*(3), 446–461. https://doi.org/10.1037/0033-2909.86.3.446

Romani, C., MacDonald, A., De Felice, S., & Palermo, L. (2017). Speed of processing and executive functions in adults with phenylketonuria: Quick in finding the word, but not the ladybird. *Cognitive Neuropsychology*, 1–28. https://doi.org/10.1080/02643294.2017.1320278

Salman, M. S., Sharpe, J. a, Eizenman, M., Lillakas, L., Westall, C., To, T., … Steinbach, M. J. (2006). Saccades in children. *Vision Research*, *46*(8–9), 1432–9. https://doi.org/10.1016/j.visres.2005.06.011

Scott, C. R. (2006). The Genetic Tyrosinemias. *American Journal of Medical Genetics. Part C*, *142C*, 121–126. https://doi.org/10.1002/ajmg.c

Shapiro, E., Bjoraker, K., Guler, E., Delaney, K., Charnas, L., Tolar, J., … Thomas, K. (2009). 130. Neuropsychological outcomes in Hurler syndrome: 7–20 years post-hematopoietic stem cell transplantion. *Molecular Genetics and Metabolism*, *96*(2), S40–S41. https://doi.org/10.1016/j.ymgme.2008.11.131

Szymanska, E., Sredzinska, M., Ciara, E., Piekutowska-Abramczuk, D., Ploski, R., Rokicki, D., & Tylki-Szymanska, A. (2015). Tyrosinemia type III in an asymptomatic girl. *Molecular Genetics and Metabolism Reports*, *5*, 48–50. https://doi.org/10.1016/j.ymgmr.2015.10.004

Thimm, E., Herebian, D., Assmann, B., Klee, D., Mayatepek, E., & Spiekerkoetter, U. (2011). Increase of CSF tyrosine and impaired serotonin turnover in tyrosinemia type I. *Molecular Genetics and Metabolism*, *102*, 122–125. https://doi.org/10.1016/j.ymgme.2010.11.003

Thimm, E., Richter-Werkle, R., Kamp, G., Molke, B., Herebian, D., Klee, D., … Spiekerkoetter, U. (2012). Neurocognitive outcome in patients with hypertyrosinemia type I after long-term treatment with NTBC. *Journal of Inherited Metabolic Disease*, *35*(2), 263–8. https://doi.org/10.1007/s10545-011-9394-5

Thomas, M., Annaz, D., & Ansari, D. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language and Hearing Research*, *52*(2), 336–358.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Van Ginkel, W. G., Jahja, R., Huijbregts, S. C. J., Daly, A., Macdonald, A., De Laet, C., … Van Spronsen, F. J. (2016). Neurocognitive outcome in tyrosinemia type 1 patients compared to healthy controls. *Orphanet Journal of Rare Diseases*, *11*, 1–9. https://doi.org/10.1186/s13023-016-0472-5

Venkatraman, V., Rosati, A. G., Taren, A. A., & Huettel, S. A. (2009). Resolving response, decision, and strategic control: Evidence for a functional topography in dorsomedial prefrontal cortex. *Journal of Neuroscience*, *29*(42), 13158–13164. https://doi.org/10.1523/JNEUROSCI.2708-09.2009

Wraith, J. E. (2006). Mucopolysaccharidoses and Oligosaccharidoses. In *Inborn Errors of Metabolism: Diagnosis and Treatment* (4th Edition, pp. 495–507). Springer.

Zorzi, M. (2010). The connectionist dual process (CDP) approach to modelling reading aloud. *European Journal of Cognitive Psychology*, *22*(5), 836–860. https://doi.org/10.1080/09541440903435621

Table 1: Summary of cognitive deficits across cognitive domains

Task	MPS-IVa				Tyrosenimia III			
	Number of patients with scores worse than control mean	Number of patients with $z_{pi} > 1$	Number of patients with $z_{pi} > 2$	Evidence of group difference?	Number of patients with scores worse than control mean	Number of patients with $z_{pi} > 1$	Number of patients with $z_{pi} > 2$	Evidence of group difference?
Simple reaction time	11/12	3/12	0/12	weak	7/11	4/11	3/11	**yes**
Visual search								
Feature search								
Overall search time	11/12	7/12	6/12	**yes**	7/11	6/11	4/11	**older patients**
Effect of display size	7/12	5/12	3/12	no	1/11	1/11	0/11	no
Conjunction search								
Overall search time	10/12	7/12	3/12	**yes**	7/11	7/11	5/11	**yes**
Effect of display size	6/12	2/12	2/12	no	4/11	3/11	1/11	no
Language								
Boston Naming Test	7/11	4/11	1/11	no	9/10	8/10	6/10	**yes**
BPVS	9/11	7/11	0/11	weak	11/11	10/11	7/11	**yes**
Oculomotor tasks								
Fixation task								
Dwell time	11/11	10/11	8/11	**yes**	9/10	6/10	2/10	weak
Intrusive saccades	9/11	7/11	5/11	**yes**	3/11	2/11	1/11	no
Prosaccade task								
Onset time	5/9	1/9	0/9	no	5/9	2/9	1/9	no
Saccade velocity	6/9	1/9	0/9	no	1/9	0/9	0/9	no

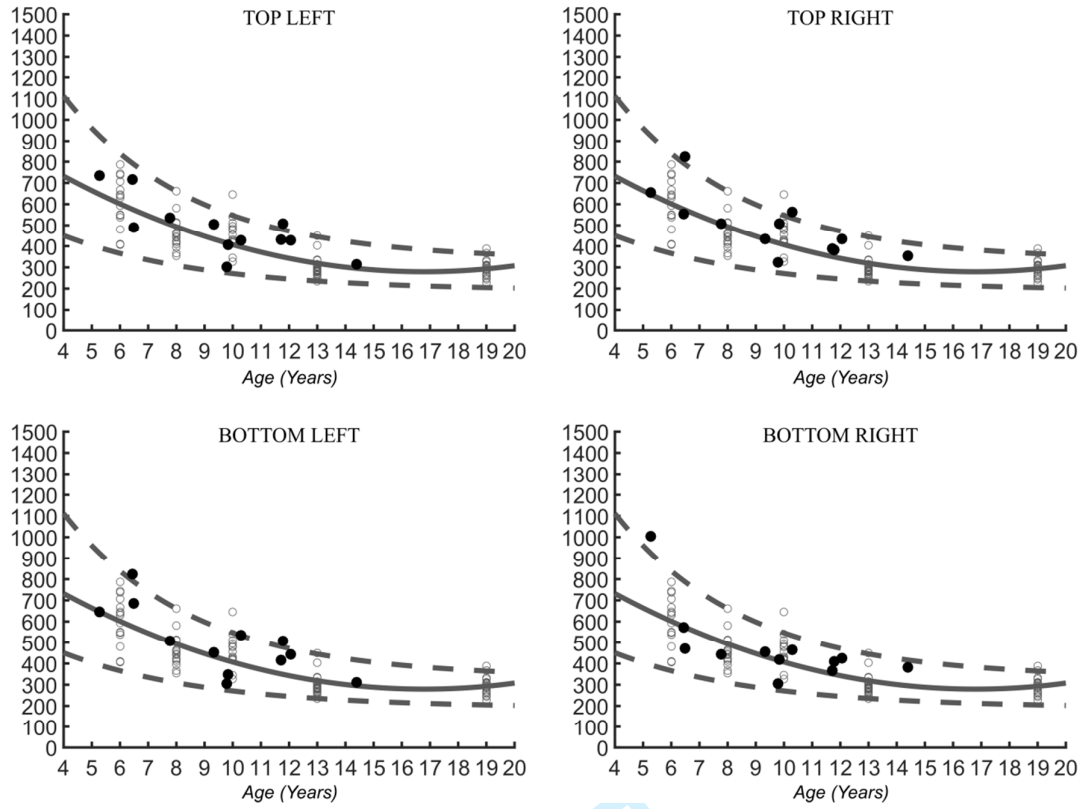


Figure 1. Mean simple reaction time for MPS-IVa patients (filled circles) and typically developing controls (open circles). The developmental trajectory for controls is plotted as a solid line with 95% prediction intervals (dashed lines).

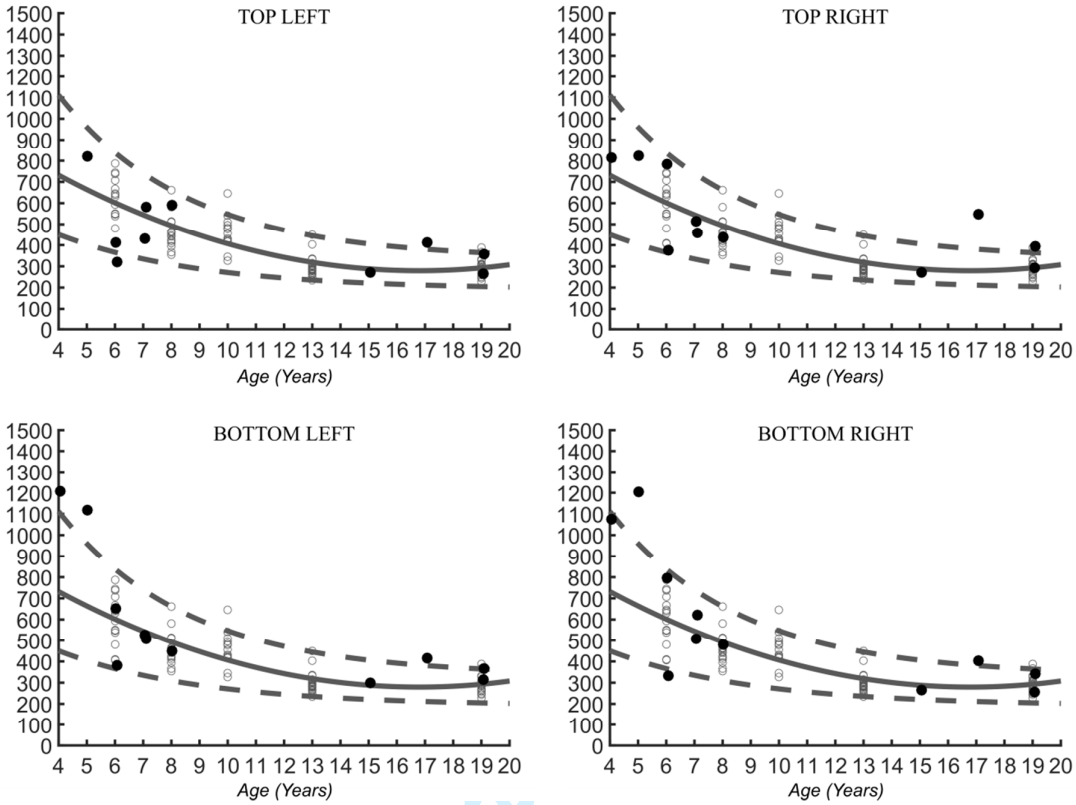


Figure 2. Mean simple reaction time for T3 patients (filled circles) and TD controls (open circles). The developmental trajectory for controls is plotted as a solid line with a 95% prediction intervals (dashed lines).

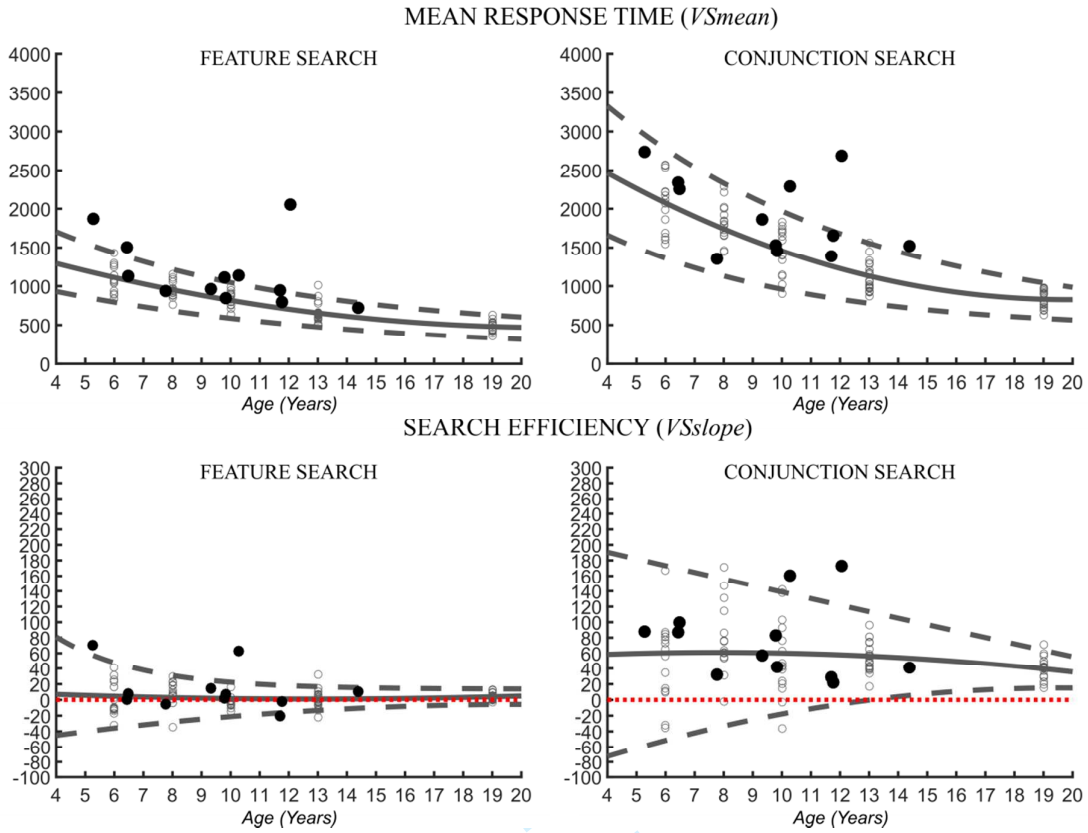
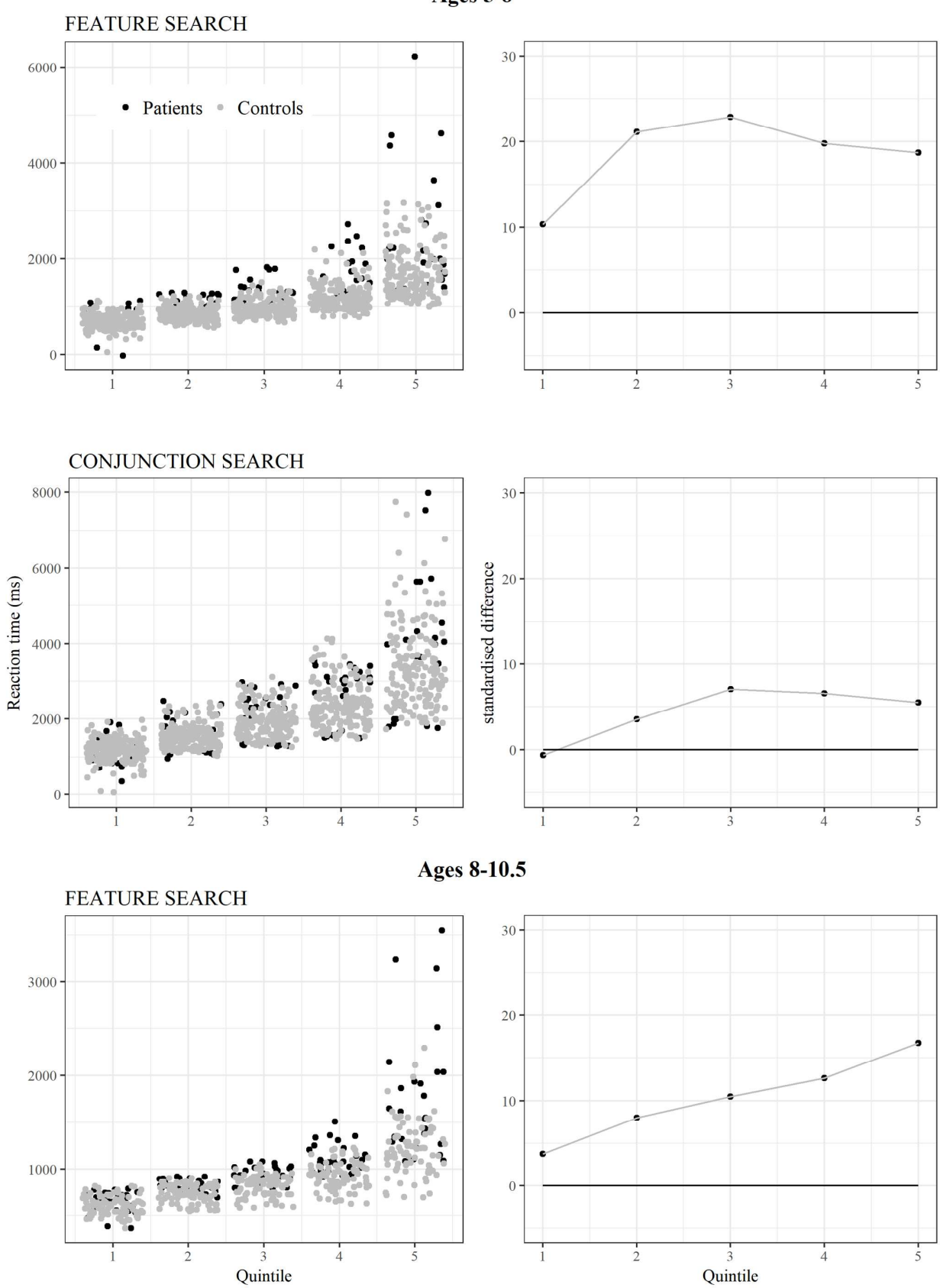


Figure 3. Overall mean RT and search efficiencies (display size slopes) in a visual search task for MPS-IVa patients (filled circles) and controls (open circles). Feature search results are in the left panels and conjunction search on the right. The developmental trajectory for controls are is plotted as a solid line with 95% prediction intervals (dashed lines). The red dotted line in the efficiency plots represents efficient search (no increase in reaction time for larger displays).


MPS-IVa
Ages 5-8
FEATURE SEARCH
Patients
Controls
CONJUNCTION SEARCH
Reaction time (ms)
standardised difference
Ages 8-10.5
FEATURE SEARCH
Quintile
Quintile

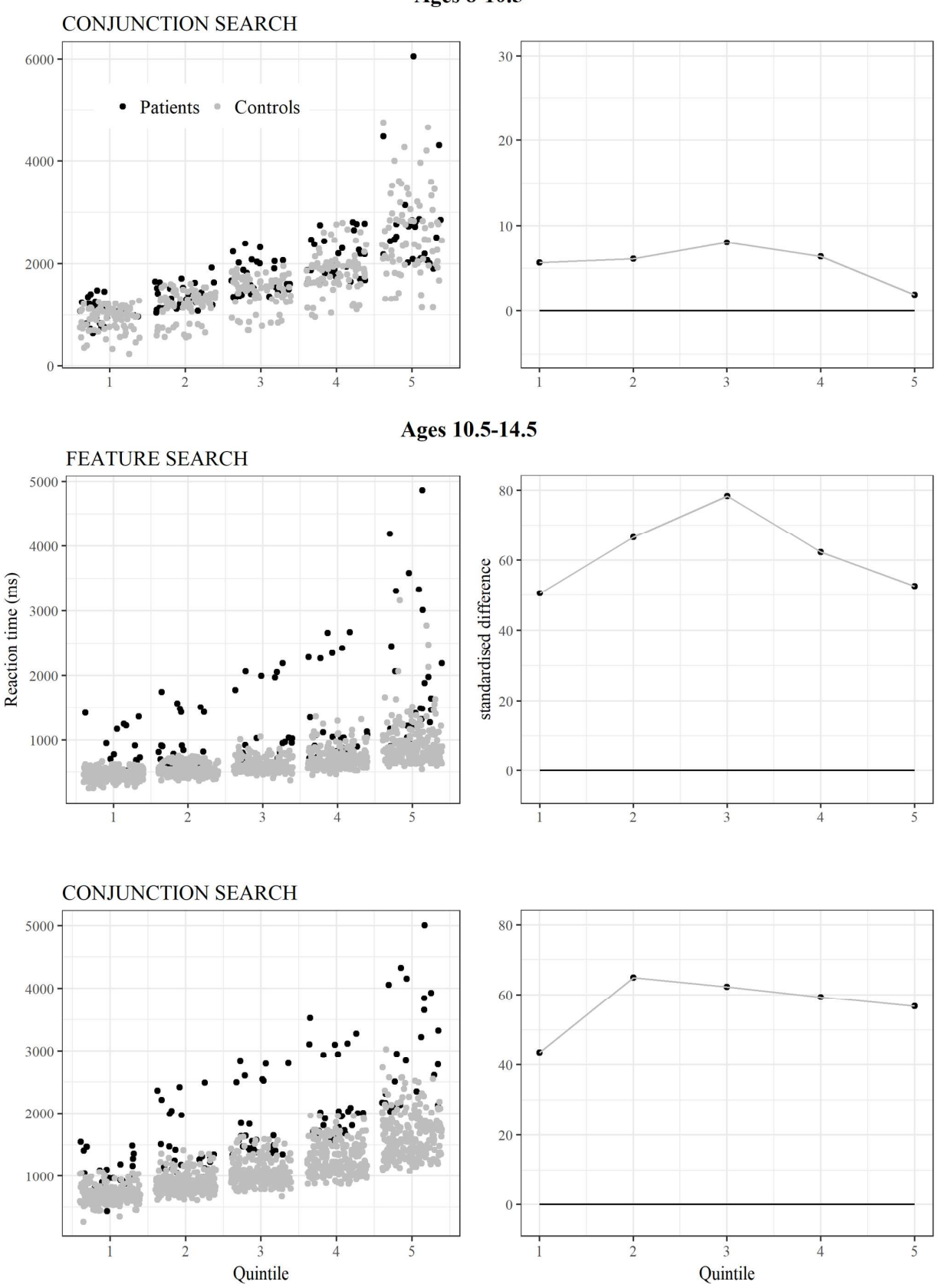


Figure 4. MPS-IVa and Control reaction times from feature and conjunction search. Reaction times are displayed in quintile bins along with the difference between quintile means for patients and controls (control value = 0). Differences between means were standardised by the standard error from the control bin. Note change in scale for final two difference plots.

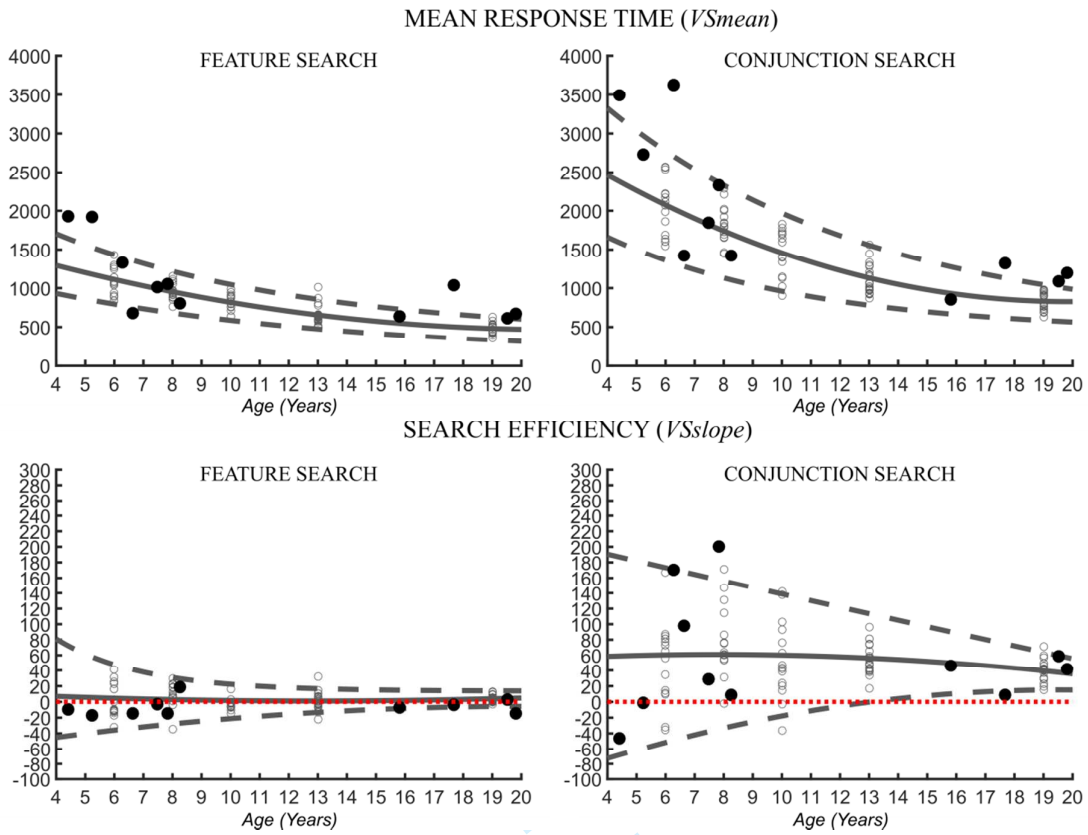
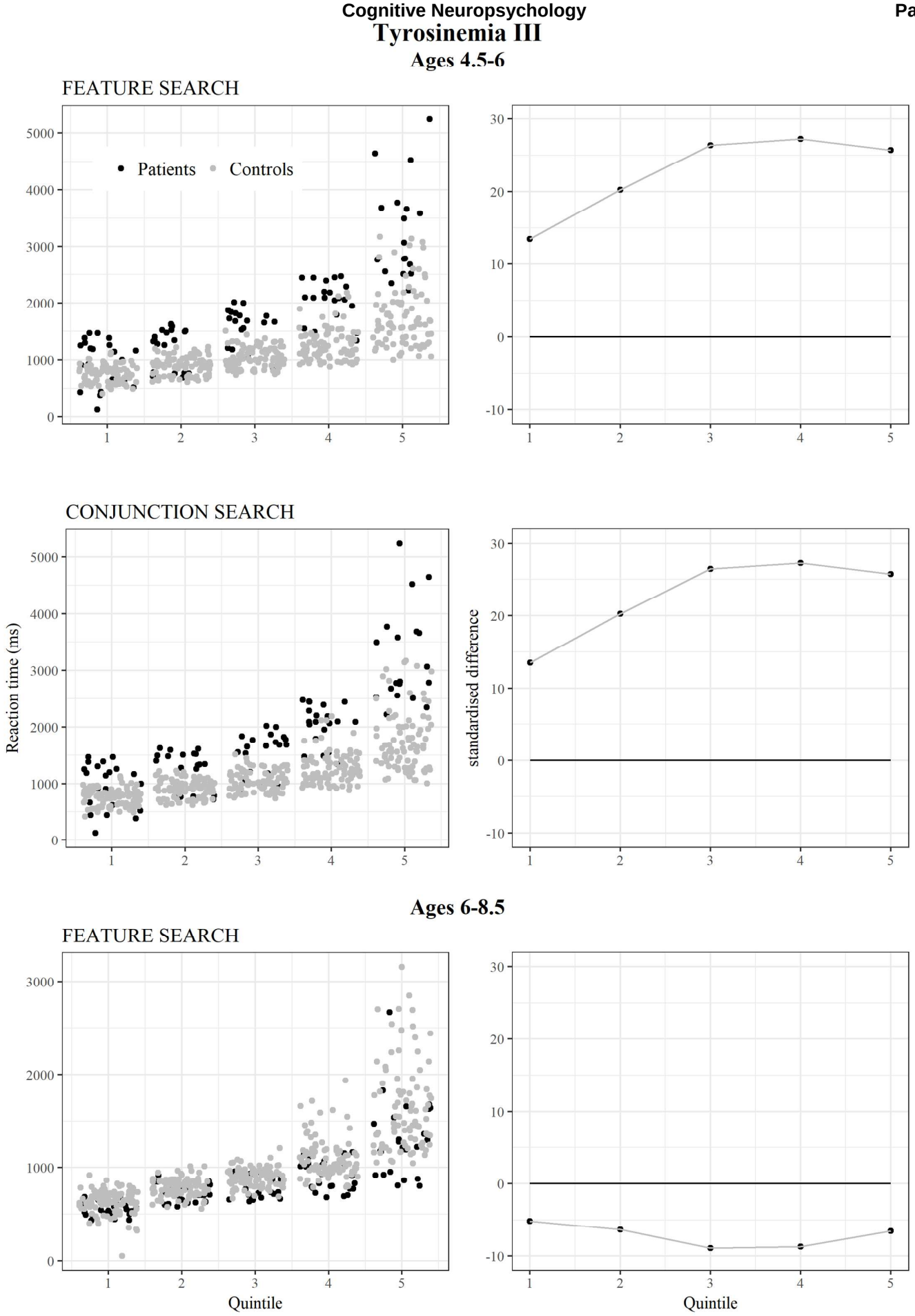


Figure 5. Overall mean RT and search efficiencies (display size slopes) in a visual search task for T3 patients (filled circles) and controls (open circles). Feature search results are in the left panels and conjunction search on the right. The developmental trajectory for controls are is plotted as a solid line with 95% prediction intervals (dashed lines). The red dotted line in the efficiency plots represents efficient search (no increase in reaction time for larger displays).


Tyrosinemia III
Ages 4.5-6
FEATURE SEARCH
Patients
Controls
CONJUNCTION SEARCH
Reaction time (ms)
standardised difference
Ages 6-8.5
FEATURE SEARCH
Quintile
Quintile

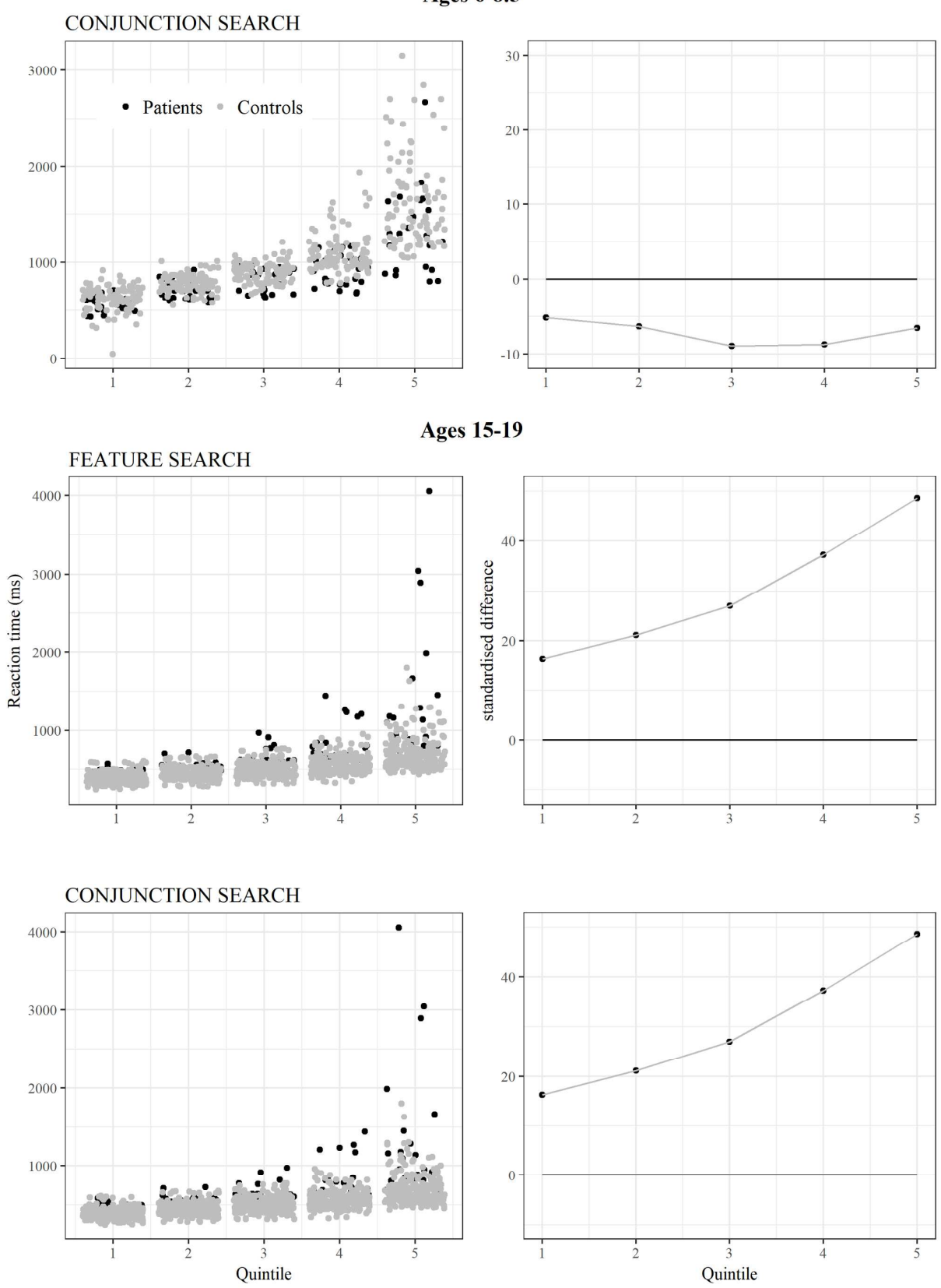


Figure 6. Tyrosinemia III and Control reaction times from feature and conjunction search. Reaction times are displayed in quintile bins along with the difference between quintile means for patients and controls (control value = 0). Differences between means were standardised by the standard error from the control bin. Note change in scale for final two difference plots.

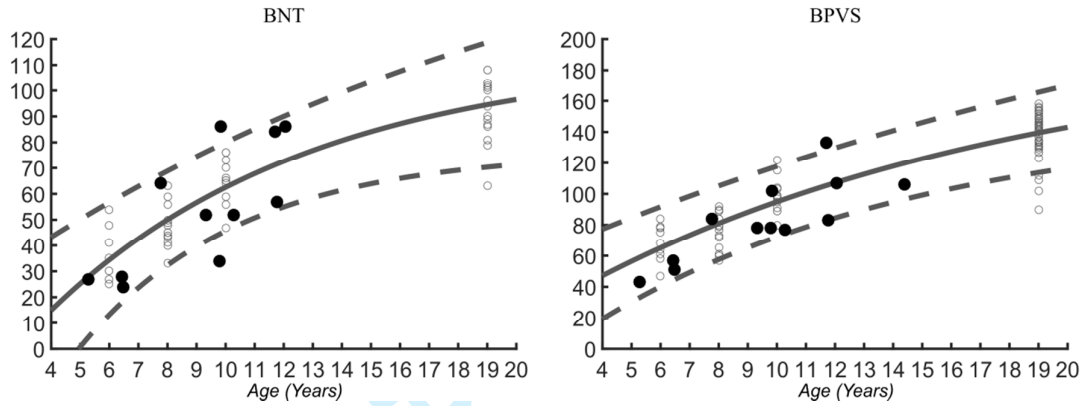


Figure 7. Scores for MPS-IVa patients and controls on the Boston Naming Test (BNT; production) and British Picture Vocabulary Scale (BPVS; comprehension). Patients are filled circles and typically-developing controls are open circles. Typically developing trajectories (solid line) and 95% prediction intervals are included (dashed lines).

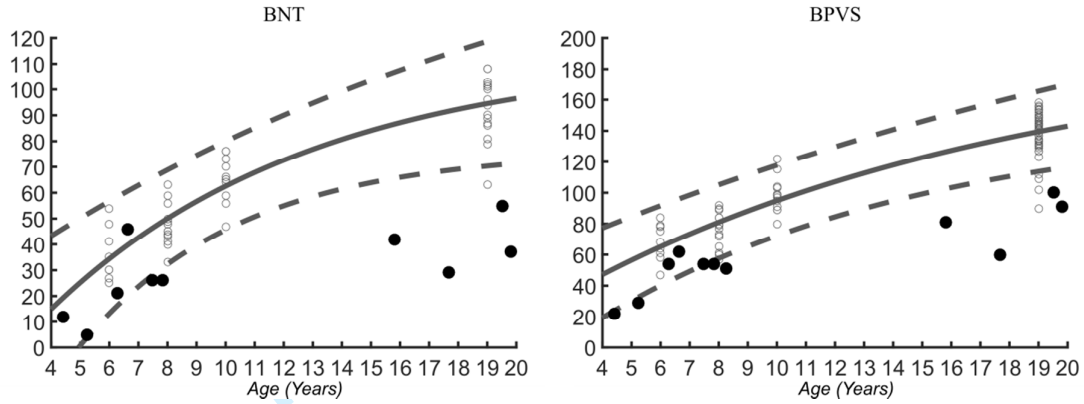


Figure 8. Scores for Tyrosenimia III patients and controls on the Boston Naming Test (BNT; production) and British Picture Vocabulary Scale (BPVS; comprehension). Patients are filled circles and typically-developing controls are open circles. Typically developing trajectories (solid line) and 95% prediction intervals are included (dashed lines).

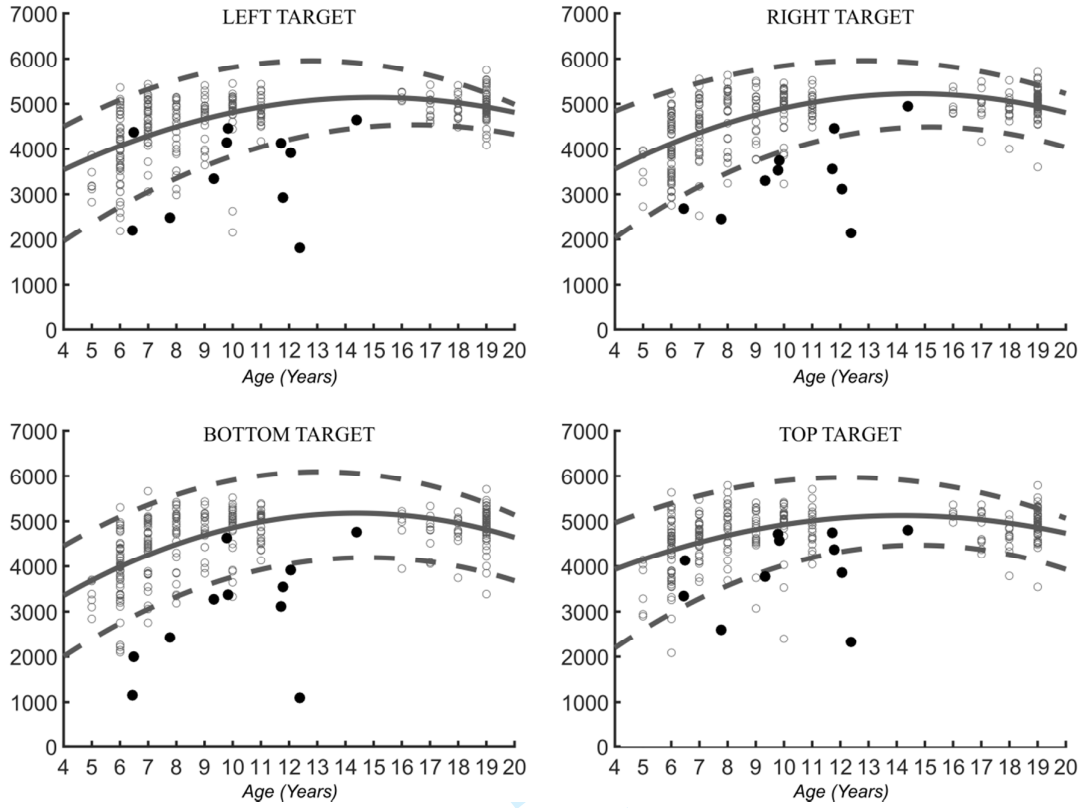


Figure 9. Fixation dwell time for MPS-IVa patients (filled circles) and TD controls (open circles). Typically developing trajectories (solid line) and 95% prediction intervals (dotted lines) are included.

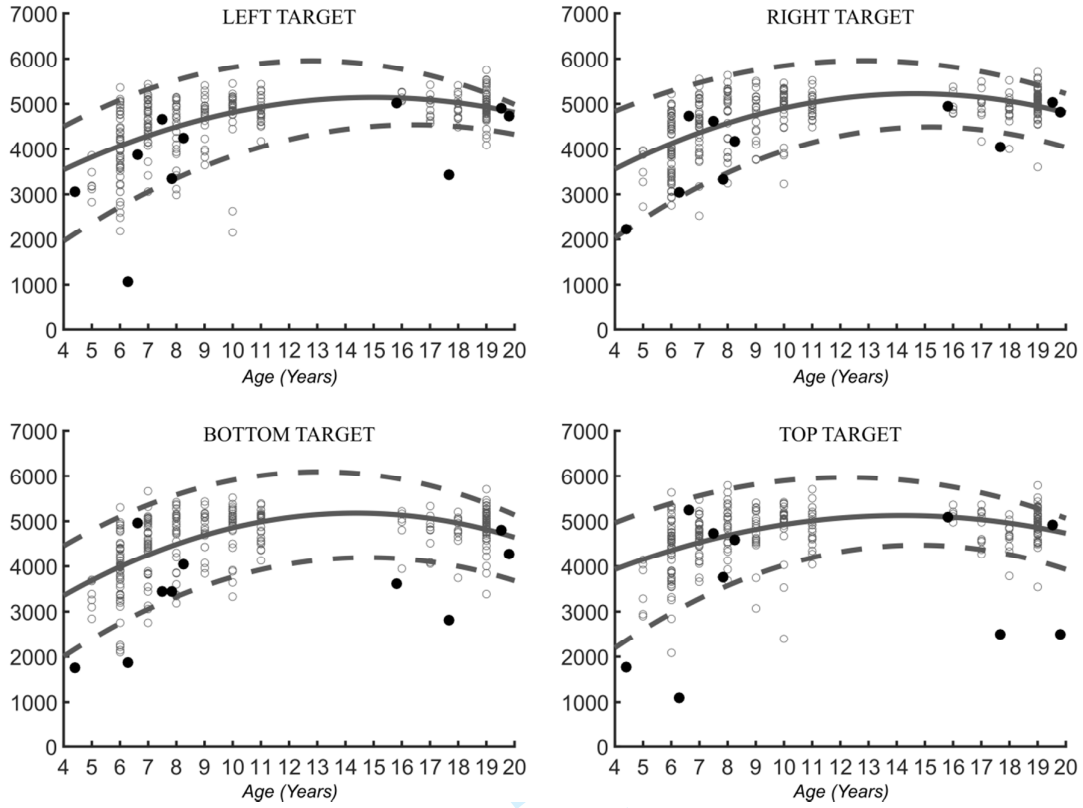


Figure 10. Fixation dwell time for Tyrosinemia III patients (filled circles) and TD controls (open circles). Typically developing trajectories (solid line) and 95% prediction intervals (dotted lines) are included.

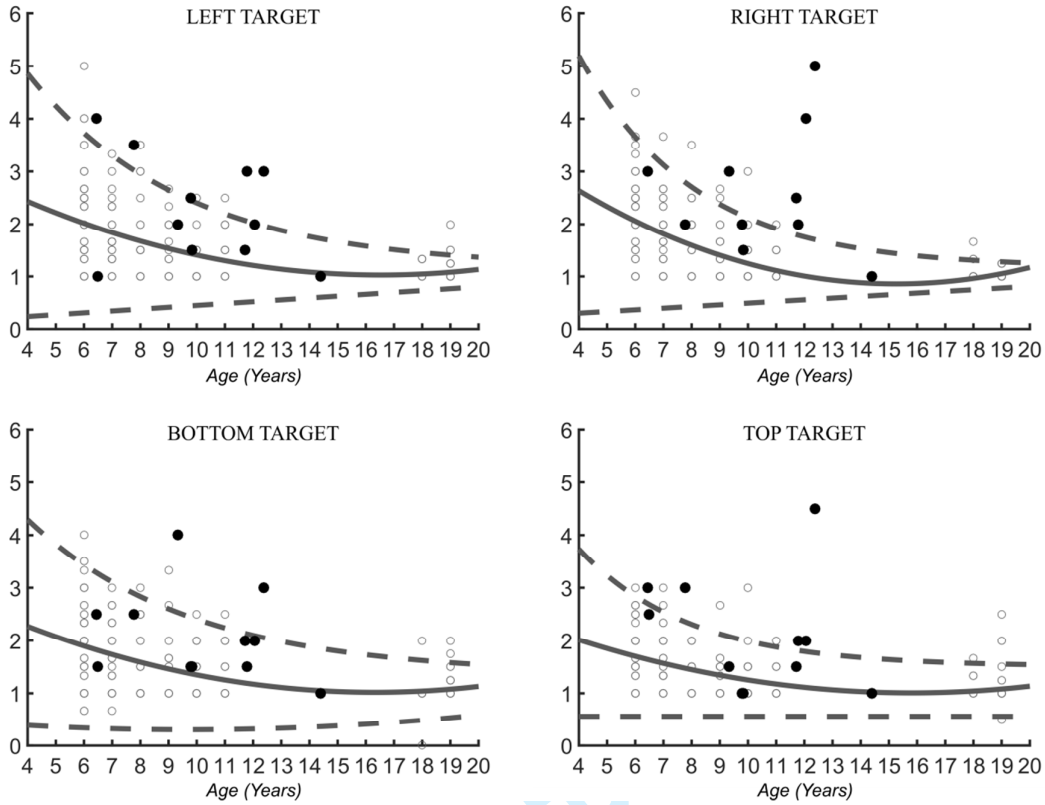


Figure 11. Number of intrusive saccades for MPS-IVa patients (filled circles) and TD controls (open circles). Typically developing trajectories (solid line) and 95% prediction intervals (dotted lines) are included.

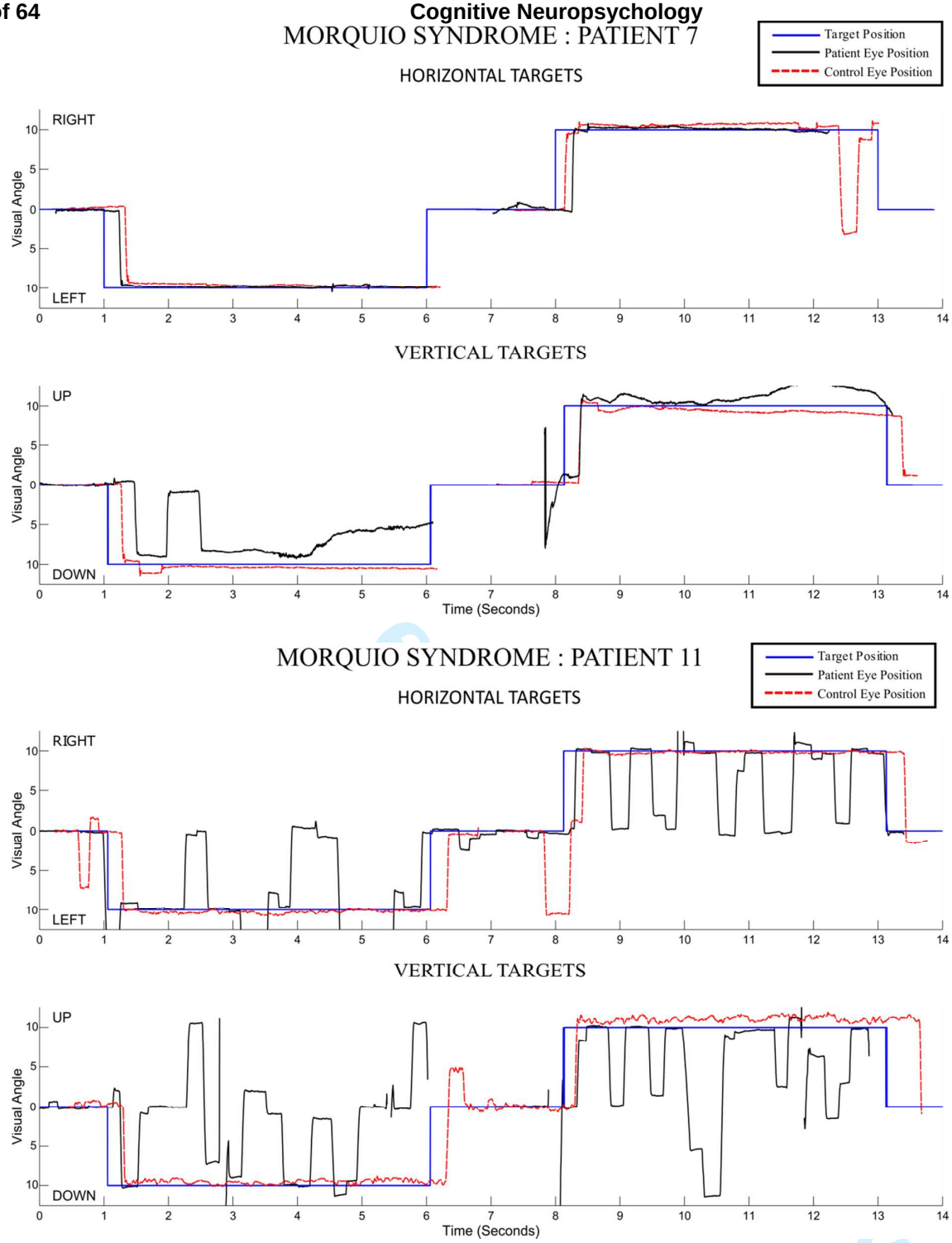


Figure 12. Fixation task saccades over time for MPS-IVa patients 7 and 11. The visual stimulus (blue line) is presented along with eye position of the patients (black line) and age-matched TD controls (red dashed line). Both patients showed fixation duration deficits buts only patient 11 displayed an elevated intrusive saccade count. Patient 7 disengages before the end of the trial.

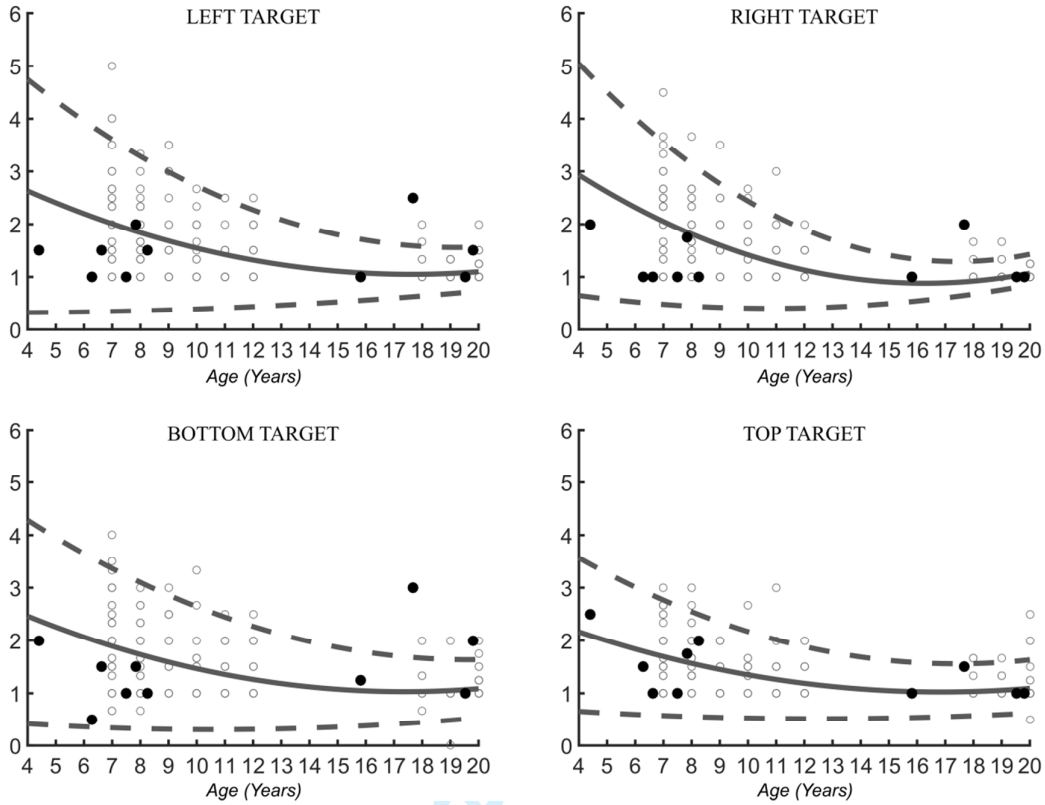


Figure 13. Number of intrusive saccades for Tyrosinemia III patients (filled circles) and TD controls (open circles). Typically developing trajectories (solid line) and 95% prediction intervals (dotted lines) are included.

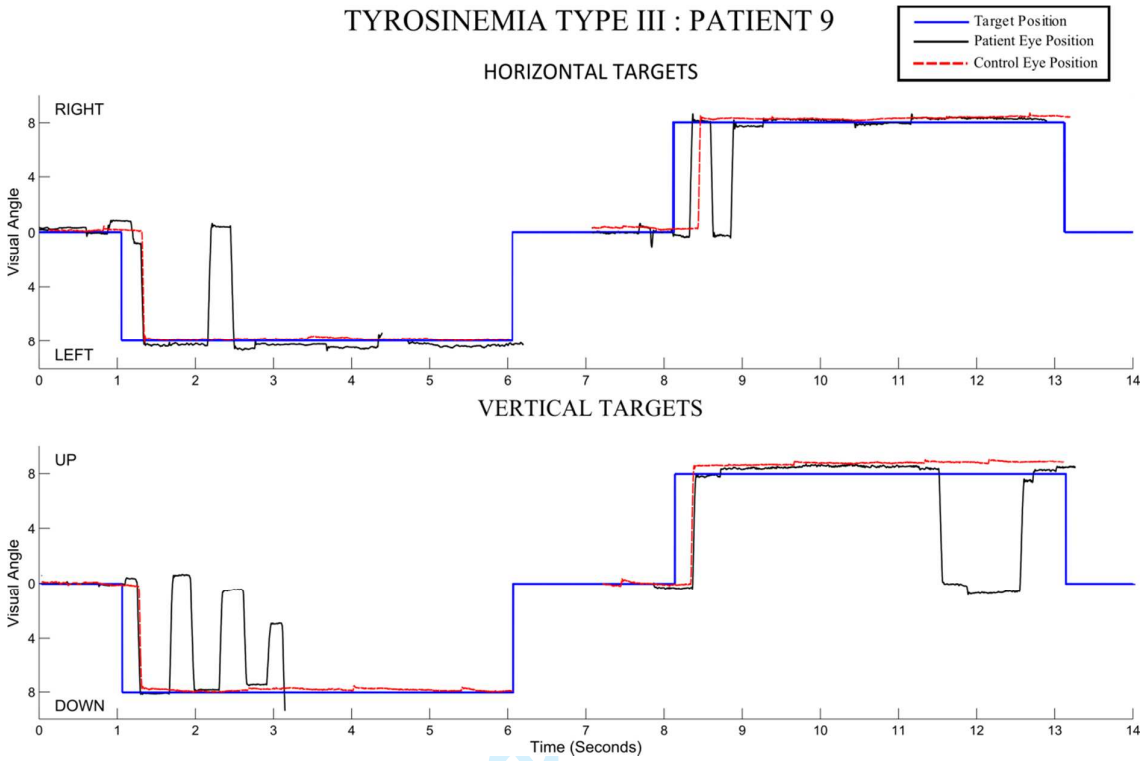


Figure 14. Fixation task saccades over time for T3 patient 9. The visual stimulus (blue line) is presented along with eye position of the patient (black line) and age-matched controls (dashed red line).

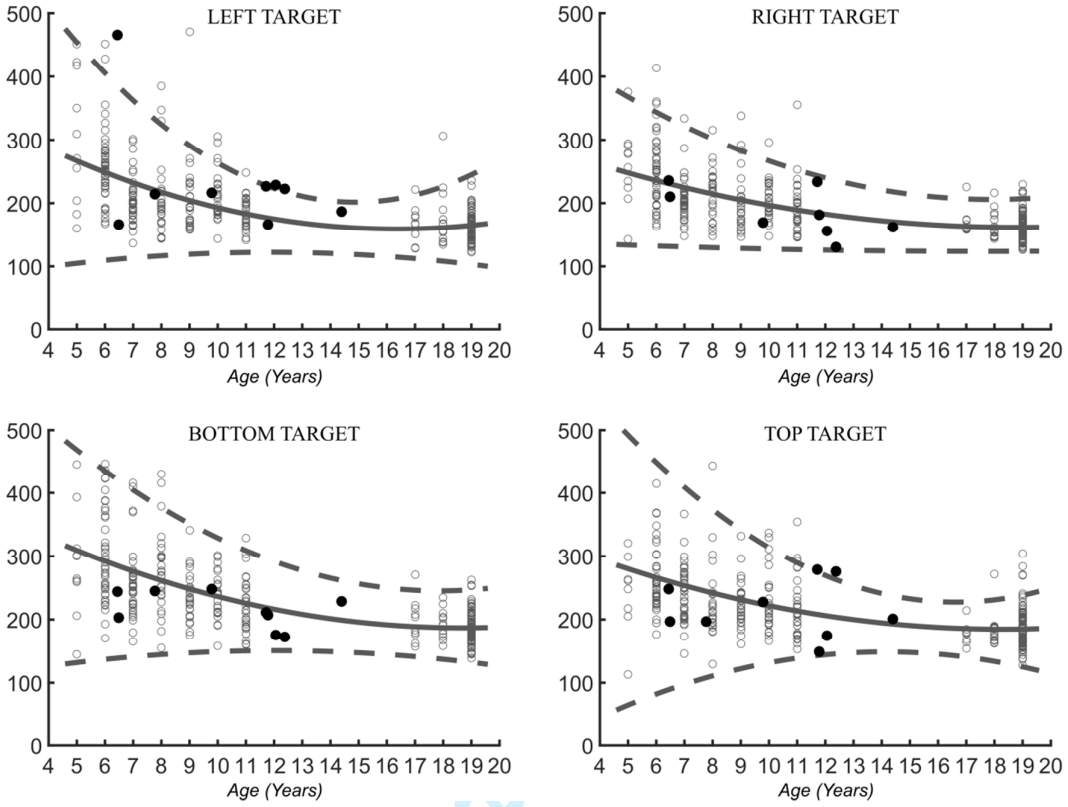


Figure 15. Average saccadic onset time (*SaccOnset*) for MPS-IVa patients (black dots) and TD controls (outlined grey dots). Horizontal target locations are shown in the top row of graphs and vertical target locations are shown in the bottom row of graphs. The typical developmental trajectory (grey solid line) is expressed as a quadratic function. 95% confidence limits (grey dashed-line) are presented.

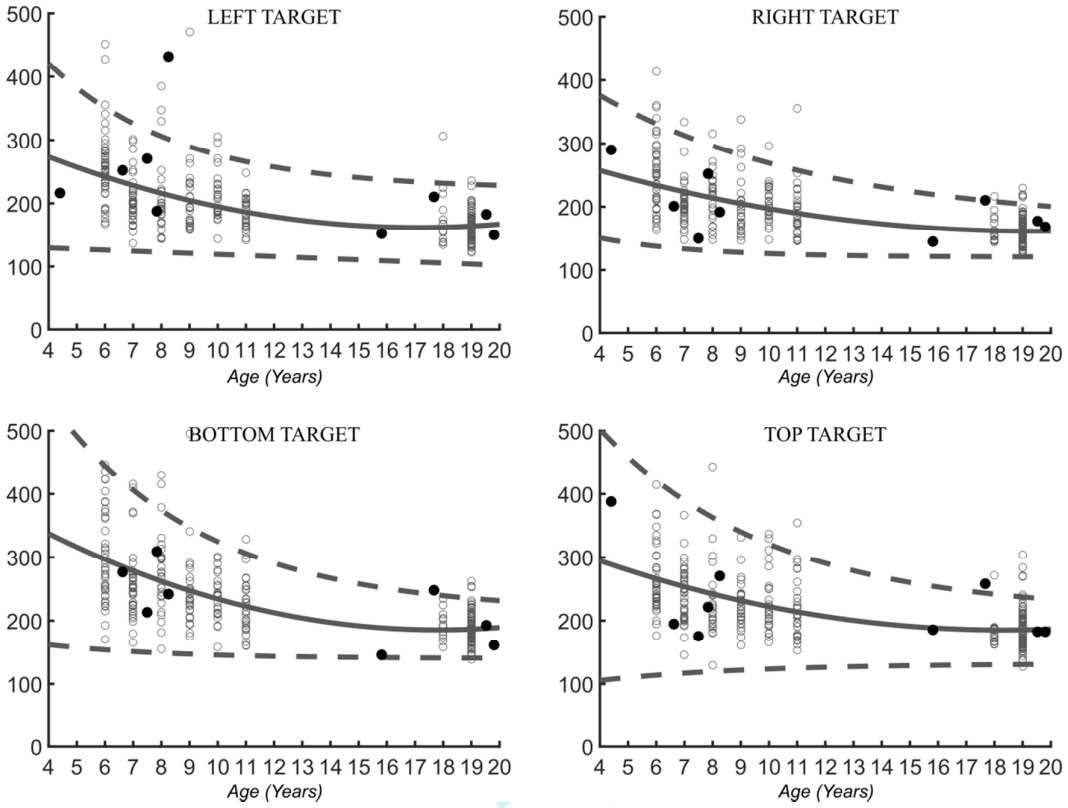


Figure 16. Average saccadic onset time (*SaccOnset*) for T3 patients (black dots) and TD controls (outlined grey dots). Horizontal target locations are shown in the top row of graphs and vertical target locations are shown in the bottom row of graphs. The typical developmental trajectory (grey solid line) is expressed as a quadratic function. 95% confidence limits (grey dashed-line) are presented.

Appendices

Appendix 1. MPS-IVa and Tyrosinemia III patient demographics

Group	PID	Gender	CA (Years)	BPVS MA (Years)	Linguistic Profile
MPS-Iva	1	M	5.27	4.02	BL – English / Pashto
MPS-Iva	2	F	6.44	5.00	BL – English / Pashto
MPS-IVa	3	F	6.48	5.07	BL – Pashto / English
MPS-IVa	4	M	7.77	8.02	BL – English / Pashto
MPS-IVa	5	F	9.33	8.07	BL – Pashto / English
MPS-IVa	6	F	9.79	10.08	ML – English
MPS-IVa	7	M	9.84	7.07	BL – Manderin / English
MPS-IVa	8	M	10.28	7.06	BL – Pashto / English
MPS-IVa	9	M	11.71	15.10	ML - English
MPS-IVa	10	M	11.78	7.10	ML - English
MPS-IVa	11	F	12.05	11.04	BL – English / Pashto
MPS-IVa	12	M	14.39	11.03	BL – English / Pashto
T3	1	F	4.42	5.32	BL – English / Pashto
T3	2	M	5.23	3.08	BL – Pashto / English
T3	3	F	6.28	5.32	BL – English / Punjabi
T3	4	M	6.62	9	ML - English
T3	5	F	7.48	5.92	BL – English / Pashto
T3	6	F	7.83	5.00	BL – Pashto / English
T3	7	M	8.25	2.92	BL – English / Pashto
T3	8	M	15.81	5.32	BL – English / Pashto
T3	9	M	17.68	7.92	BL – Pashto / English
T3	10	F	19.52	6.08	BL – English / Punjabi
T3	11	F	19.81	10.32	BL – English / Punjabi

Note: ML, Monolingual, BL, Bilingual

Appendix 2. MPS-IVa model results

Model	AIC	ΔAIC	Akaike Weight
Simple RT			
Condition X Age + Age X Group	-458.7	0.0	0.23
Condition X Age + Group	-458.7	0.0	0.22
Condition X Age	-458.5	0.2	0.20
Condition X Age + Age X Group + Condition X Group	-455.8	2.8	0.05
Condition X Age + Condition X Group	-455.8	2.9	0.05
Age X Group	-455.4	3.3	0.04
Feature search			
Age + Group	3162.8	0.0	0.36
Group X Age	3164.2	1.3	0.18
Display Size² + Age + Group	3164.5	1.7	0.15
Display Size² X Age + Group X Age + Display Size² X Group	3169.5	6.7	0.01
Conjunction search			
Display Size X Age + Group X Age	5539.5	0.0	0.19
Display Size + Age X Group	5539.5	0.0	0.19
Display Size X Age + Group	5540.2	0.8	0.13
Display Size + Age + Group	5540.3	0.8	0.13
Display Size X Age + Group X Age + Display Size X Group	5541.3	1.8	0.08
Display Size X Age X Group	5543.1	3.6	0.03
Boston Naming Test			
Age	560.5	0.0	1.00
Age X Group	829.3	268.8	0.00
British Picture Vocabulary Scale			
Age X Group	829.3	0.0	0.53
Age + Group	830.0	0.7	0.38
Age	833.0	3.6	0.09
Group	1004.8	175.5	0.00
Fixation dwell time			
Condition X Age + Age X Group + Condition X Group	16713.0	0.0	0.52
Condition X Age X Group	16714.1	1.2	0.29
Condition X Age + Condition X Group	16715.0	2.1	0.18
Condition X Age + Age X Group	16721.7	8.8	0.01
Intrusive saccades			
Condition X Age + Group	1734.9	0.0	0.29
Condition X Age + Condition X Group	1735.1	0.3	0.26
Condition X Age + Age X Group	1736.0	1.1	0.17
Condition X Age + Age X Group + Condition X Group	1736.1	1.2	0.16
Condition X Age X Group	1738.2	3.3	0.06
Saccade onset			
Condition X Age + Condition X Group	-977.0	0.0	0.58

Condition X Age + Age X Group + Condition X Group	-975.0	2.0	0.21
Condition X Age	-972.8	4.2	0.07
Saccade velocity			
Condition X Age	15446.6	0.0	0.63
Condition X Age + Group	15448.5	2.0	0.24
Condition X Age + Age X Group	15450.2	3.7	0.10

Appendix 3. Tyrosinemia III model results.

Model	AIC	ΔAIC	Akaike Weight
Simple RT			
Condition X Age X Group	-422.1	0.00	0.45
Age + Group	-420.1	1.97	0.17
Age + Group + Condition	-419.5	2.60	0.12
Age X Group	-418.5	3.63	0.07
Feature search			
Group X Age	3232.2	0.0	0.29
Group X Age + Display Size X Group	3233.1	0.9	0.18
Display Size + Age X Group	3233.2	1.1	0.17
Display Size X Age + Group X Age + Display Size X Group	3233.7	1.6	0.13
Display Size X Age + Group X Age	3233.8	1.6	0.13
Display Size X Age X Group	3235.4	3.2	0.06
Conjunction search			
Display Size + Age + Group	5565.7	0.00	0.18
Display Size + Age X Group	5565.8	0.07	0.17
Display Size X Age + Group	5566.5	0.81	0.12
Display Size X Age + Group X Age	5566.6	0.87	0.11
Display Size X Group + Age	5567.2	1.47	0.09
Group X Age + Display Size X Group	5567.2	1.52	0.08
Display Size + Age	5568.0	2.29	0.06
Display Size X Age + Group X Age + Display Size X Group	5568.0	2.31	0.06
Display Size X Age X Group	5568.7	2.97	0.04
Display Size X Age	5568.8	3.08	0.04
Boston Naming Test			
Age X Group	532.6	0.0	0.93
Age + Group	537.8	5.2	0.07
Age	566.4	33.8	0.00
Group	593.9	61.4	0.00
British Picture Vocabulary Scale			
Age X Group	807.3	0.0	0.93
Age + Group	812.5	5.2	0.07
Age	876.6	69.3	0.00
Group	975.7	168.4	0.00
Fixation dwell time			
Condition X Age + Condition X Group	16009.9	0.0	0.34
Condition X Age X Group	16010.2	0.3	0.29
Condition X Age + Group	16011.2	1.3	0.18
Condition X Age + Age X Group + Condition X Group	16011.9	2.0	0.13
Condition X Age + Age X Group	16013.2	3.3	0.07

Intrusive saccades			
Condition X Age + Age X Group	1645.5	0.0	0.29
Condition X Age	1645.6	0.0	0.28
Condition X Age X Group	1646.9	1.4	0.15
Condition X Age + Group	1647.4	1.9	0.11
Age + Condition	1649.2	3.7	0.05
Saccade onset			
Condition X Age	-975.4	0.0	0.51
Condition X Age + Group	-974.1	1.3	0.26
Condition X Age + Age X Group	-972.1	3.3	0.10
Saccade velocity			
Condition X Age + Condition X Group	15455.0	0.0	0.36
Condition X Age	15455.8	0.8	0.24
Condition X Age + Group	15456.2	1.2	0.19
Condition X Age + Age X Group + Condition X Group	15457.0	2.0	0.13
Condition X Age + Age X Group	15458.2	3.2	0.07