4H leukodystrophy: a brain MRI scoring system

Suzanne Vrij-van den Bos MD[1]*, Janna A. Hol MD[1]*, Roberta La Piana MD[2,3]*, Inga Harting MD[4], Adeline Vanderver MD[5], Frederik Barkhof MD PhD[6,7], Ferdy Cayami MD[1,8], Wessel N. van Wieringen PhD[9], Petra J. W. Pouwels PhD,[10,11] Marjo S. van der Knaap MD PhD[1,11,12], Geneviève Bernard MD MSc[3,13,14]#, Nicole I. Wolf MD PhD[1,11]#+

1 Department of Child Neurology, VU University Medical Center, Amsterdam, The Netherlands

2 Laboratory of Neurogenetics of Motion and Department of Neuroradiology, Montreal Neurological Institute and Hospital, McGill University, Montreal, Canada

3 Departments of Neurology and Neurosurgery, and Pediatrics, McGill University, Montreal, Canada

4 Department of Neuroradiology, University Hospital Heidelberg, Germany

5 Department of Neurology, Children's National Medical Center, Washington DC and George Washington University School of Medicine, Washington DC, United States

6 Department of Radiology, VU University Medical Center, Amsterdam, The Netherlands

7 Institute of Neurology & Health Care Engineering, University College London, London, UK

8 Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands

9 Department of Clinical Epidemiology and Biostatistics, VU University Medical Center & Department of Mathematics, VU University, Amsterdam, The Netherlands

10 Department of Physics and Medical Technology, VU University Medical Center

11 Amsterdam Neuroscience Amsterdam, The Netherlands

12 Department of Functional Genomics, VU University, Amsterdam, The Netherlands

13 Department of Medical Genetics, Montreal Children's Hospital, McGill University Health Center, Montreal, Canada

14 Child Health and Human Development Program, Research Institute of the McGill University Health Center, Montreal, Canada

* Contributed equally to this work (first authors)

# Contributed equally to this work (senior authors)

[+] *Corresponding author*

Dr Nicole I Wolf

Dept. of Child Neurology

VU University Medical Center

Postbox 7057

1007 MB Amsterdam

The Netherlands

Phone: +31-20-4444879

Fax: +31-20-444

e-mail: n.wolf@vumc.nl

ABSTRACT

4H leukodystrophy (4H) is an autosomal recessive hypomyelinating white matter (WM) disorder with neurological, dental and endocrine abnormalities. The aim of this study was to develop and validate an MRI scoring system for 4H. A scoring system (0-54) was developed to quantify hypomyelination and atrophy of different brain regions. Pons diameter and bicaudate ratio were included as measures of cerebral and brainstem atrophy, and reference values were determined using controls. Five independent raters completed the scoring system in 40 brain MRI scans collected from 36 patients with genetically proven 4H. Interrater reliability (IRR) and correlations between MRI scores, age, gross motor function, gender and mutated gene were assessed. IRR for total MRI severity was found to be excellent (ICC 0.87, 95% CI 0.80-0.92), but varied between different items with some (e.g. myelination of the cerebellar white matter) showing poor IRR. Atrophy increased with age, in contrast to hypomyelination scores. MRI scores (global, hypomyelination, and atrophy scores) significantly correlated with clinical handicap ($p<0.01$ for all three items) and differed between the different genotypes. Our 4H MRI scoring system reliably quantifies hypomyelination and atrophy in patients with 4H, and MRI scores reflect clinical disease severity.

KEYWORDS

4H, leukodystrophy, MRI, hypomyelination

INTRODUCTION

Hypomyelinating disorders constitute a heterogeneous group of different disease entities, defined by a permanent and significant lack of central nervous system myelin.[1] One of the striking features of these conditions is the large variation in severity: some patients have virtually no neurological signs and normal cognition, others are severely affected early on and wheelchair-bound, usually with less profoundly affected cognitive function. A recent study aimed at correlating gross motor function (GMF) with quantitative white matter MRI parameters and showed that, in a heterogeneous group of patients with hypomyelination, it is mainly the lack of myelin which co-determines motor handicap.[2]

Short of quantitative information, it is desirable to grade imaging severity tailored to a single hypomyelinating entity by assessing the degree of myelin deficit and atrophy using conventional T1- and T2-weighted images, and to correlate these scores with GMF. This would be helpful for retrospective studies, as these images are available for most patients, but also for routine clinical use, providing useful information for future therapeutic trials. Such scoring systems have proven to be a valuable tool for other leukodystrophies such as X-linked adrenoleukodystrophy (X-ALD) and metachromatic leukodystrophy (MLD).[3,4] Recently, a scoring system for Pelizaeus-Merzbacher disease (PMD), the most common hypomyelinating disorder, was published, but was not available when we were embarking on this project.[5]

We aimed to develop a scoring system for a hypomyelinating disorder, 4H leukodystrophy (4H), the second most common hypomyelinating entity[6] characterized by hypomyelination, hypodontia and hypogonadotropic hypogonadism.[7–9] Recessive mutations in either *POLR3A*, *POLR3B* or *POLR1C*, encoding subunits of RNA polymerase III and I, cause almost all cases with the clinical diagnosis of 4H, with mutations in *POLR3B* explaining the majority of cases in Central and Western Europe and mutations in *POLR3C* being the least frequent.[10–14] Clinical variation is wide, ranging from mild, even subclinical ataxia to severe motor

handicap, and the course is, after a stable phase ranging from 2-3 years to more than 10 years, slowly progressive in most. In addition to hypomyelination, MRI findings comprise early cerebellar atrophy, especially in cases with mutations in *POLR3B*, and late, milder supratentorial brain atrophy including thinning of the corpus callosum.[15] Some white matter (WM) structures are better myelinated: the optic radiation, the ventrolateral thalamus, the pyramidal tract in the posterior limb of the internal capsule (PLIC), the dentate nucleus and the medial lemniscus in the brain stem.[16] We present a novel MRI rating scale that reflects these findings and determined its interrater reliability and clinical validity.

PATIENTS AND METHODS

Patient selection

Brain MRI scans were retrospectively collected from patients with genetically proven 4H leukodystrophy from the databases of VU Medical Center in Amsterdam, The Netherlands, and Montreal Children's Hospital in Montreal, Canada. Scans had been performed at different scanners (1.5 and 3T) following different protocols at different centers and were included when at least axial T2- and T1-weighted images and a sagittal series were available. Only patients with MRI scans acquired at age 24 months or later were included because normal myelination is not completed before this age.

Controls

A total of 86 healthy subjects (45 male, 41 female) with a median age of 6.5 years (range 0-29) served as a control group. We focused on younger ages as there were no good normal values for BCR and brainstem diameter available, in contrast to adults. Reasons for MRI were headache, mild epilepsy without focal neurological abnormalities or mild developmental delay. All were evaluated by (pediatric) neurologists and radiologists, and no abnormal neurological or MRI findings were identified.

MRI scoring

We developed a scoring system to assess hypomyelination and atrophy of different brain regions (table 1). Hypomyelination was graded separately on axial T1- and T2-weighted images, depending on the signal intensity in relation to the caudate nucelues. Atrophy was visually scored for selected structures (cerebellum, corpus callosum; figure 1) and quantified using the bicaudate ratio (BCR), defined as the smallest intercaudate distance divided by the transverse width of the inner table of the skull at the same level. These distances were measured on T2-weighted axial images with the smallest intercaudate distance (figure 2C-D). Brainstem atrophy was quantified by measuring maximal anterior-posterior diameter at the level of the mid-pons on sagittal images (figure 2E). Overall, patients could receive a score between 0-44 for hypomyelination and 0-10 for atrophy, adding up to a total score of 0-54, with higher scores corresponding to more severe abnormalities.

Two neuroradiologists (FB and IH) and three pediatric neurologists (RLP, AV and GB), experienced in WM diseases, were asked to individually complete the scoring system for all patients. They were blinded to patient identifiers including age (the scores for the two age-dependent items, BCR and pons diameter, were deduced from the measurements the scorers provided). When three or more raters agreed on a score, this score was used for further analysis on clinical validity. Items for which less than three raters agreed on a score, were sent for a second round of scoring to reach consensus. Interrater reliability (IRR) was assessed for all individual MRI scoring items and measurements as well as for overall MRI severity scores, separately for the first and second round. Total MRI severity, hypomyelination and atrophy scores were analyzed including their association to factors such as age, gender and mutated gene. For both pons diameter and BCR, age-related reference ranges were determined using scans of controls with normal MRI.

Correlation to gross motor function

Patients were retrospectively classified according to the gross motor function classification system (GMFCS)[17] based on clinical examination or clinical notes from the year in which the MRI study was performed. Retrospective estimation of gross motor function scores was previously shown to be reliable.[18]

Statistical analysis

Data were analyzed by SPSS, version 20.0 for Windows (SPSS Inc., Chicago, IL, USA), GraphPad Prism 7 (GraphPad Software, San Diego, CA, USA) and R Software (R Foundation for statistical computing, Vienna, Austria). For determining reference ranges, normal values of BCR and pons diameter were plotted and where appropriate, modelled by a (monotone) exponential curve fitted by means of nonlinear least squares. Subsequently, reference ranges were determined using population means and standard errors.

IRR was expressed using the single measure intraclass correlation coefficient (ICC), which was analyzed by means of a two-way mixed model based on absolute agreement (for separate items) or consistency (for total scores). Cutoffs for ICC were derived from Cicchetti et al. 1994, with ICC values less than 0.40 being considered poor, values between 0.40 and 0.59 fair, values between 0.60 and 0.74 good and values between 0.75 and 1.00 excellent.[19]

Pearson (r) and Spearman (ρ) correlations were computed to determine associations between age, GMF and MRI variables including total MRI score, hypomyelination and atrophy scores. Subgroup differences for gender, gene mutation and age groups were tested using a general linear model, analysis of variance (ANOVA) and Mann-Whitney test and associated p values. P values of 0.05 or less were considered significant.

RESULTS

Patients and controls

A total of 40 MRI brain scans (32 on a 1.5T scanner, 8 on a 3T scanner) were collected from 36 (19 female, 17 male) patients with genetically proven 4H with a median age of 12 years (range 2-39 years). Six patients (4 female, 2 male) had mutations in *POLR3A*, the majority, 27, in *POLR3B* (13 female, 14 male) and 3 in *POLR1C* (2 female, 1 male). Severity of clinical disease varied widely with GMFCS scores ranging from I (n=7) to V (n=2).

Figure 2 shows the evolution of BCR and pons diameter in controls and patients. There was no significant difference between males and females. For grading brainstem and supratentorial atrophy in 4H, we pragmatically designed age groups using normal values (figure 2; tables 2 A and B).

In 4H patients, pons diameter was significantly smaller with a mean diameter of 18.38 mm (standard deviation (SD) 1.98 mm) versus 20.86 mm (SD 1.89 mm) in controls ($p<0.01$). Similarly, BCR was found to be higher in patients with a mean BCR of 0.094 (SD 0.026) versus 0.071 (SD 0.015) in controls ($p<0.01$). Although this difference was already present in younger patients, it further increased with age, indicating progressive volume loss.

Scoring disagreement and interrater reliability

In the first round a total of 1080 scores were collected (27 items scored for 40 MRI scans). For only 48 (4.4%) scores, less than three raters agreed. In one single patient, the maximum number of items with disagreement was 3. After the second assessment of the 48 items with diverging scores, only 3 (0.3%) from the total of 1080 scores remained without agreement of at least 3 raters. We refrained from imposing a consensus score and accepted that no consensus could be reached for these 3 items.

In the first round, IRR was found to be excellent for all sub-scores, with ICC values for consistency of 0.84 (95% CI 0.76-0.90) for hypomyelination, 0.88 (95% CI 0.82-0.93) for atrophy and 0.87 (95% CI 0.80-0.92) for total MRI score. Overall, IRR was lower for items scored on T1-weighted scans, with some items showing a poor IRR (ICC<0.40). These items

included the periventricular frontoparietal border area, parieto-occipital WM (both subcortical and periventricular), the anterior limb of the internal capsule (ALIC) and the cerebellar WM. In contrast, all items scored on T2-weighted images showed at least a fair IRR (ICC>0.40). There was little disagreement on atrophy scores including those scores derived from pons diameter and BCR.

In the second round including the 48 scores with disagreement, results were largely similar. One of the difficult items to score was myelination of the cerebellar WM, especially in patients with severe cerebellar atrophy. We therefore considered omitting this region from the final score. Analyzing results without this item did not influence overall outcome (not shown), therefore we decided not to omit it from the scoring, but, to facilitate future use, to slightly modify the system by rating white matter signal of the middle cerebellar peduncles instead.

MRI severity, hypomyelination and atrophy scores

Median scores were 31/54 (range 8-45) for total MRI score, 23/44 (range 7-39) for hypomyelination and 7/10 (range 0-10) for atrophy. Scores tended to be higher in females than in males (median 34 versus 26 respectively, p=0.024). Total MRI scores did not significantly increase with age figure 3A). When looking at hypomyelination and atrophy scores separately, only atrophy scores were significantly correlated with age (figure 3B-C). Global MRI and hypomyelination scores were significantly higher for patients with mutations in *POLR3A* than for patients with mutations in *POLR3B* while atrophy scores were in the same range (figure 4).

MRI scores and motor disability

Patients with more limitations in GMF were found to have higher total MRI scores (figure 3D). However, a large variability of MRI scores was seen with all GMFCS levels, interestingly most pronounced in patients with mild disability. A similar distribution was seen

for hypomyelination and atrophy scores when analyzed separately, with increasing scores for more advanced clinical disease (figure 3E-F).

MRI scores at follow-up

Follow-up MRI examinations were available for 4 patients with a time interval ranging from 1 to 8 years between scans. Atrophy scores either remained stable (n=1) or slightly worsened (n=3). Hypomyelination scores, and therefore total MRI severity scores, improved in three of them. In those patients who showed improvement of hypomyelination scores, GMFCS remained stable (n=2) or worsened slightly (n=1).

DISCUSSION

The aim of this study was to develop and validate a MRI scoring system for 4H leukodystrophy, assessing hypomyelination and atrophy in this hypomyelinating disease. We decided to score both hypomyelination and atrophy, as cerebellar atrophy is a known early feature of 4H, and as supratentorial atrophy slowly progresses with age. The authors of the recently published score for PMD independently also combined atrophy and myelination scores. They included FLAIR in addition to T1W and T2W images, but omitted measurements to quantify atrophy.[5]

IRR was found to be excellent for hypomyelination, atrophy and global MRI severity scores when assessed using ICC statistics. Nevertheless, in the first round IRR varied between different items of the scoring system with some items showing poor IRR, particularly those scored on T1-weighted images. Apparently, these items were more difficult to score or more sensitive to rater subjectivity. More precise instructions provided to the raters in the second round resulted in better agreement. As some atrophy items were assessed quantitatively (BCR[20], pons diameter[21]), scoring of atrophy may be more objective than in the established scores for X-ALD or MLD.[3,4] BCR was first described as a measure of atrophy in the 1980s

and used for a variety of diseases including multiple sclerosis and Alzheimer's disease.[22,23] Recently, one group used this parameter to measure brain atrophy in PMD.[24] In our patients, brainstem atrophy sets in much earlier than supratentorial atrophy, possibly at least in part due to the early cerebellar atrophy.

All MRI severity scores – hypomyelination, atrophy and global scores – correlated with gross motor function, supporting the hypothesis that not only degree of hypomyelination, but also atrophy co-determines handicap. A recent study on MRI changes in PMD reached similar conclusions, albeit less pronounced for myelination than for atrophy.[5] Also in PMD, another group reported correlation of brain atrophy with disability, though without assessing myelination.[24] In multiple sclerosis, the prototype of acquired white matter disease, brain atrophy is at least as important for clinical, especially cognitive impairment as white matter lesions.[25, 26] Atrophy, in contrast to hypomyelination scores, clearly increased with age, suggesting age-dependent neuroaxonal loss rather than loss of myelin, although more long-term observations are needed to confirm this finding in larger patient groups. Interestingly, in 4H patients with mild clinical disease, MRI scores differed more widely than in patients with more severe disease, confirming our clinical experience.

In females, MRI scores tended to be higher than in male patients; this might be at least in part due to the fact that females were overrepresented in the group of patients with mutations in *POLR3A*. That in the four patients with follow-up scans there were some changes including improved myelination scores despite stable or slightly deteriorated GMF, is difficult to interpret, due to the small number of follow-up MRIs. Larger studies are needed to study evolution of MRI changes and clinical symptoms over time.

We recently described several patients with atypical forms of 4H, without clear myelin deficit and variable neurological signs and symptoms.[27] Those patients were not included in this study as we focused on typical patients with hypomyelination, but it will be interesting to

know whether significantly more myelin in these atypical cases also means less clinical handicap once enough cases will become available.

We were also interested in possible differences between patients with mutations in different genes. Although our group consisted mainly of patients with mutations in *POLR3B,* patients with mutated *POLR3A* had higher mean MRI severity scores than the other two groups, consistent with the greater clinical severity with *POLR3A* mutations shown in our study on a large number of patients with 4H leukodystrophy.[15]

One might wonder why, in the times of quantitative imaging and volumetry we chose a comparably simple approach for this scoring system. It has several benefits: the sequences we used are part of standard protocols and therefore usually available, they can be compared among diverse scanners (with their slightly different protocols), and there is no need for specific control groups. Quantitative methods are more objective than a rater-dependent approach and often depend on specific scanners, making it laborious or even impossible to compare results between different centers or with changing MRI systems in even the same center. A robust, simple method as ours may be used in multicenter studies and also can make use of MRIs collected during a longer period, essential for a rare disease as 4H leukodystrophy. These advantages are at the same time the limitations of this study: the use of different scanners with different field strengths and protocols and the retrospective setup may have affected results. That we did find correlations between MRI changes and clinical severity despite these odds and that these correlations are comparable to other recent MRI studies on PMD, is encouraging and strengthens our interpretation.

In conclusion, this study shows that the proposed MRI scoring system reliably quantifies hypomyelination and atrophy, both correlating with clinical disease severity, in patients with 4H leukodystrophy. It is easy to apply, even retrospectively, might serve as a useful

biomarker for future studies and can be used, with small adaptations, for other hypomyelinating disorders.

ACKNOWLEDGEMENTS

We thank all colleagues referring patients and providing MR images and clinical information. G.B. has received a Research Scholar Junior 1 award from the Fonds de Recherche du Québec en Santé (FRQS). This work was supported by Canadian Institute for Health Research grant MOP-G-287547 to G.B. F.K.C. is a recipient of the Directorate of Higher Education Overseas Scholarship–Dikti Scholarship, Ministry of National Education, Republic of Indonesia. F.B. is supported by the NIHR Biomedical Research Centre at UCLH.

REFERENCES

1. Pouwels PJ, Vanderver A, Bernard G, et al. Hypomyelinating leukodystrophies: translational research progress and prospects. Ann Neurol 2014;76:5–19

2. Steenweg ME, Wolf NI, Wieringen WN, Barkhof F, van der Knaap MS, Pouwels PJ. Quantitative MRI in hypomyelinating disorders. Neurology 2016;87:752–758

3. Loes DJ, Hite S, Moser H, et al. Adrenoleukodystrophy: a scoring method for brain MR observations. Am J Neuroradiol 1994;15:1761–1766

4. Eichler F, Grodd W, Grant E, et al. Metachromatic leukodystrophy: a scoring system for brain MR imaging observations. Am J Neuroradiol 2009;30:1893–1897

5. Sarret C, Lemaire J, Tonduti D, et al. Time-course of myelination and atrophy on cerebral imaging in 35 patients with PLP1-related disorders. Dev Med Child Neurol 2016; 58:706–713

6. Cayami FK, La Piana R, Spaendonk RM van, et al. *POLR3A* and *POLR3B* mutations in unclassified hypomyelination. Neuropediatrics 2015;46:221–228

7. Wolf NI, Harting I, Boltshauser E, et al. Leukoencephalopathy with ataxia, hypodontia, and hypomyelination. Neurology 2005;64:1461–1464

8. Timmons M, Tsokos M, Asab MA, et al. Peripheral and central hypomyelination with hypogonadotropic hypogonadism and hypodontia. Neurology 2006;67:2066–2069

9. Wolf NI, Harting I, Innes AM, et al. Ataxia, delayed dentition and hypomyelination: a novel leukoencephalopathy. Neuropediatrics 2007;38:64–70

10. Bernard G, Chouery E, Putorti M, et al. Mutations of *POLR3A* Encoding a Catalytic Subunit of RNA Polymerase Pol III Cause a Recessive Hypomyelinating Leukodystrophy. Am J Hum Gen 2011;89:415–423

11. Saitsu H, Osaka H, Sasaki M, et al. Mutations in *POLR3A* and *POLR3B* Encoding RNA Polymerase III Subunits Cause an Autosomal-Recessive Hypomyelinating Leukoencephalopathy. Am J Hum Gen 2011;89:644–651

12. Daoud H, Tétreault M, Gibson W, et al. Mutations in *POLR3A* and *POLR3B* are a major cause of hypomyelinating leukodystrophies with or without dental abnormalities and/or hypogonadotropic hypogonadism. J Med Gen 2013;50:194–197

13. Thiffault I, Wolf NI, Forget D, et al. Recessive mutations in *POLR1C* cause a leukodystrophy by impairing biogenesis of RNA polymerase III. Nat Commun 2015;6:7623

14. Tétreault M, Choquet K, Orcesi S, et al. Recessive Mutations in *POLR3B*, Encoding the Second Largest Subunit of Pol III, Cause a Rare Hypomyelinating Leukodystrophy. Am J Hum Gen 2011;89:652–655

15. Wolf NI, Vanderver A, Spaendonk RM van, et al. Clinical spectrum of 4H leukodystrophy caused by *POLR3A* and *POLR3B* mutations. Neurology 2014;83:1898–1905

16. Steenweg ME, Vanderver A, Blaser S, et al. Magnetic resonance imaging pattern recognition in hypomyelinating disorders. Brain 2010;133:2971–2982

17. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Dev Med Child Neurol 1997;39:214–223

18. Mayson TA, Ward V, Davies KR, et al. Reliability of retrospective assignment of gross motor function classification system scores. Dev Neurorehabil 2013; 16:207–209

19. Cicchetti DV. Multiple comparison methods: establishing guidelines for their valid application in neuropsychological research. J Clin Exp Neuropsychol 1994;16:155–161

20. Butzkueven H, Kolbe SC, Jolley DJ, et al. Validation of linear cerebral atrophy markers in multiple sclerosis. J Clin Neurosci 2008;15:130-137

21. Raininko R, Autti T, Vanhanen SL, Ylikoski A, Erkinjuntti T, Santavuori P. The normal brain stem from infancy to old age. A morphometric MRI study. Neuroradiology 1994;36:364-368

22. Caon C, Zvartau-Hind M, Ching W, Lisak RP, Tselis AC, Khan OA. Intercaudate nucleus ratio as a linear measure of brain atrophy in multiple sclerosis. Neurology 2003;60:323–325

23. Brickman A, Honig L, Scarmeas N, et al. Measuring cerebral atrophy and white matter hyperintensity burden to predict the rate of cognitive decline in Alzheimer disease. Arch Neurol 2008;65:1202–1208

24. Laukka J, Stanley J, Garbern J, et al. Neuroradiologic correlates of clinical disability and progression in the X-Linked leukodystrophy Pelizaeus-Merzbacher disease. J Neurol Sci 2013;335:75–81

25. Bermel RA, Bakshi R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. Lancet Neurol 2006;5:158–170

26. Roosendaal SD, Bendfeldt K, Vrenken H, et al. Grey matter volume in a large cohort of MS patients: relation to MRI parameters and disability. Mult Scler. 2011;17:1098–1106

27. La Piana R, Cayami F, Tran L, et al. Diffuse hypomyelination is not obligate for POLR3-related disorders. Neurology 2016;86:1622–1626

FIGURE LEGENDS

Figure 1: Corpus callosum and cerebellar atrophy.

In (A), a normal corpus callosum is shown with a score of 0 points, in (B) a slightly thinned corpus callosum (score 1 point) and in (C), a severey thinned corpus callosum (score 2 points). (D) depicts a normal cerebellar hemisphere (0 points), (E) a mildly atrophic cerebellar hemisphere (1 point) and (F) a severely atrophic cerebellar hemisphere (2 points). (G) displays a normal cerebellar vermis (0 points), (H) a mildly atrophic one (1 point) and (I) a severely atrophic cerebellar vermis (2 points).

Figure 2: Bicaudate ratio and pons diameter in controls and patients

(A) shows controls (black dots) with the (monotone) exponential curve (red) fitting best for this parameter (BCR = b0 + b1*exp(-month / ϑ) with b0 = 6.475216e-02, b1 = 1.052684e-03 and ϑ = -1.149132e+02; striped line: -1 standard deviation, dotted line: -2 standard deviations) and patients. Mainly patients > 16 years show supratentorial volume loss. (B) displays evolution of pons diameter in controls (black dots), again with the (monotone) exponential curve (red) fitting best for this parameter (pons diameter = b0 + b1*exp(-month / ϑ) with b0 = 23.903129, b1 = -6.064256 and ϑ = 132.567216; striped line: -1 standard deviation, dotted line: -2 standard deviations), and patients, showing early brainstem atrophy in 4H patients. (C) and (D) are examples of measuring BCR in two patients, one with a normal result of 0.079 at age 13 years (C), one with clear atrophy and a BCR of 0.155 at age 33 years (D). (E) shows a sagittal T2W image of a young patient with normal pons diameter of 19.5 mm at age 2 years. Red diamonds: patients with mutations in *POLR3A*, blue triangles: patients with mutations in *POLR3B*, green squares: patients with mutations in *POLR1C*.

Figure 3: MRI scores in relation to age and gross motor function.

Global MRI score (A) and hypomyelination score (B) do not show significant correlation with age, whereas atrophy is significantly correlated with age (C). All three scores significantly correlate with gross motor function (D-F). Red diamonds: patients with mutations in *POLR3A*, blue triangles: patients with mutations in *POLR3B*, green squares: patients with mutations in *POLR1C*.

Figure 4: MRI scores in relation to mutated gene

Score results (mean and standard error of the mean) depicted for the three patient groups (patients with mutations in *POLR3A*, *POLR3B* and *POLR1C*, respectively). Global (A) and hypomyelination (B) scores significantly differ between patients with *POLR3A* and *POLR3B* mutations (Mann-Whitney test, * $p < 0.05$; *** $p < 0.001$). Atrophy scores do not show significant differences (C). We did not include the three patients with *POLR1C* mutations in this analysis as this latter group was very small. Red diamonds: patients with mutations in *POLR3A*, blue triangles: patients with mutations in *POLR3B*, green squares: patients with mutations in *POLR1C*.

TABLES

Table 1: Brain MRI scoring system for 4H leukodystrophy

Brain area – white matter		**T1W**	**T2W**	**T1W + T2W**
Frontal white matter				
	Subcortical	0-2	0-2	0-4
	Periventricular	0-2	0-2	0-4
Frontoparietal border area				
	Subcortical	0-2	0-2	0-4
	Periventricular	0-2	0-2	0-4
Parieto-occipital white matter				
	Subcortical	0-2	0-2	0-4
	Periventricular	0-2	0-2	0-4
Internal capsule				
	ALIC	0-2	0-2	0-4
	Pyramidal tracts in PLIC	0-2	0-2	0-4
Corpus callosum				
	Genu	0-2	0-2	0-4
	Splenium	0-2	0-2	0-4
Middle cerebellar peduncles*		0-2	0-2	0-4
Total white matter score		***0-22***	***0-22***	***0-44***
Brain area - atrophy				
Supratentorial (bicaudate ratio)				0-2
Corpus callosum				0-2
Brainstem (pons diameter)				0-2
Cerebellar vermis				0-2
Cerebellar hemispheres				0-2
Atrophy score				***0-10***
Total score (white matter and atrophy)				***0-54***

White matter scores: T1W images: 0 = hyperintense, 1 = isointense, 2 = hypointense in relation to caudate nucleus. T2W images: 0 = hypointense, 1 = isointense, 2 = hyperintense in relation to caudate nucleus.

Atrophy scores: bicaudate ratio and pons diameter see tables 2 A-B. Corpus callosum: 0 = normal, 1 = slightly thinned, 2 = severely thinned. Cerebellar vermis and cerebellar hemispheres: 0 = normal, 1 = mild atrophy with slightly enlarged interfolial spaces, 2 = severe atrophy with clearly enlarged interfolial spaces.

*middle cerebellar peduncles: this item replaces cerebellar white matter in the original scoring system.

Table 2A: Scoring brainstem atrophy based on pons diameter

Score	2-3 years	4-5 years	6-7 years	8-9 years	10-12 years	13-14 years	15-19 years	20-22 years	≥ 23 years
0	≥18	≥19	≥19	≥20	≥20	≥21	≥21	≥22	≥22
1	17.99-16.01	18.99-16.01	18.99-17.01	19.99-17.01	19.99-18.01	20.99-18.01	20.99-19.01	21.99-19.01	21.99-20.01
2	≤16	≤16	≤17	≤17	≤18	≤18	≤19	≤19	≤20
	Pons diameter (mm)								

Table 2B: Scoring supratentorial atrophy based on bicaudate ratio

Score	2-9 years	10-18 years	19-23 years	≥ 24 years
0	<0.085	<0.090	<0.095	<0.100
1	0.085-0.100	0.090-0.105	0.095-0.110	0.100-0.115
2	>0.100	>0.105	>0.110	>0.115
	BCR			