COMP0118 - Coursework 3: Analysis of MRI T2 Relaxometry

Francesco Seracini
University College London
francesco.seracini.24@ucl.ac.uk

Abstract

Prematurity is still the leading cause of death worldwide for children under the age of five [1, 2]. Even though the survival rate of extremely preterm infants (EPT; born before 26 weeks) is increasing thanks to advances in medicine, the proportion of these children who develop disabilities later in life is not showing a decreasing trend [1, 2, 3]. To address this issue, several studies have used MRI and, in particular, T2 relaxometry, to image the brains of preterm infants in order to understand its causes and explore possible solutions. This study investigates mono and multi-compartment models for estimating T2 decay characteristics, highlighting the limitations and challenges of this imaging approach. Subsequently, these methods are applied to a cohort of EPT subjects and compared with term-born individuals. We examined potential correlations and analyzed specific brain regions known for high myelin density to assess whether they are less myelinated in EPT subjects.

1 Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive and versatile technique that allows for the exploration of brain anatomy and microstructure. The contrast in these images is generated by the intrinsic magnetic properties of the tissue, such as the T2 relaxation rate, which relates to the spin-spin relaxation and provides valuable information about tissue composition. In fact, the signal intensity at different echo-times (TE) and in different brain regions should follow an exponential decay, with this effect being more pronounced in tissues with shorter T2 values, such as white matter (WM) and gray matter (GM), that darken faster than tissues with longer T2, like cerebrospinal fluid (CSF), as shown in Figure 1. The different T2 values can be used to investigate both normal tissue and tissue pathology in neurodegeneration and neurodevelopment [4, 5].The differences between these tissues are given by how they bind water and, in particular, the water inside the myelin sheaths is very restricted, resulting in a much shorter T2 than the other tissues.

This is particularly important for studying the causes of different long-term morbidities that still affect a large proportion of people born extremely preterm (EP; $22 - 26$ weeks). In fact, the brain has a rapid growth in the period of postnatal development after EP birth [6] and myelination could be affected during this time, potentially altering the correct development of brain function. This is why it could be very beneficial to use T2 to determine the myelin water fraction across the brain, which has been proven to have a strong correlation with myelin [5, 6, 7, 8].

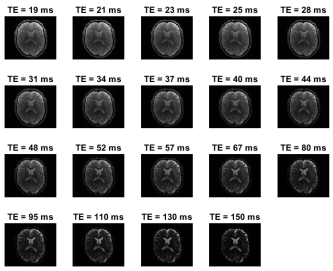


Fig. 1: Example of signal intensity decay over increasing echo-times, demonstrating how differences in T2 relaxation times affect the contrast between WM, GM and CSF.

However, several problems arise with this approach, mainly due to the imperfections of MRI analysis and machines. In fact, even though, theoretically the signal should follow a monotonic and mono-exponential decay, many measurements don't present this trend, as shown in Figure 2, because of the presence of many acquisition artifacts that can distort the signal and compromise the pre-

cision of T2 estimation. Noise, motion, magnetic field inhomogeneities and partial volume effects contribute to deviations from the ideal exponential model.

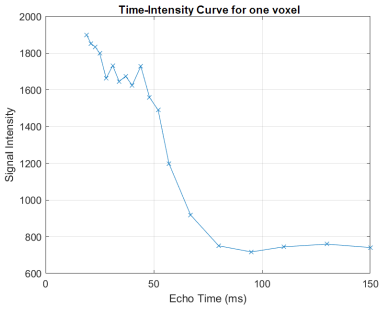


Fig. 2: Time-intensity decay curve of the T2-weighted signal for a representative voxel located at [42, 25, 42] in the first subject.

In this work we first investigate the factors that may affect the signal and the model fitting, analyzing the results of models with increasing levels of complexity. Subsequently, we apply these methods to study how the estimated parameters vary between subjects in different parts of the brain, with the aim of identifying potential correlations between the estimated parameters and subject-specific features such as gestation weeks, gender, height, weight, and IQ.

2 Methods

2.1 First Dataset

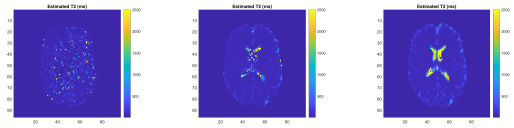
For the first part of our analysis we used six datasets from different subjects. Each subject's dataset included the main images, which consist of multi-echo T2-weighted scans, acquired at 19 different TEs, a boolean mask that indicates if each voxel is part of the brain or the background, a multi-class brain segmentation for the T2 image data, a complete and a simplified version of multi-class brain parcellation.

2.2 T2 Relaxometry

We started by considering a single-compartment model where signal decays exponentially and can be described by the equation:

$$S(t) = S_0 e^{-t/\mathrm{T2}} \tag{1}$$

To estimate the S0 and T2 parameters, we developed various functions using different algorithms and compared the obtained results. First of all, we used a simple two-point algorithm that utilizes only 2 images acquired at different TEs. The limitations of this approach are evident, and the results are heavily dependent on the choice of the 2 images. When the TEs are too similar, the T2 map becomes more imprecise and doesn't reflect the expected brain structure, as shown in Figure 3.



(a) (b) (c)

Fig. 3: Comparison of T2 maps generated using the two-point method with different echo time (TE) pairs: (a) TEs 1 and 2, (b) TEs 1 and 10, and (c) TEs 1 and 19.

After that, we implemented several algorithms that take more than 2 images and for each of them we computed the mean error, calculated as the average across all the voxels in the brain of the sum of square differences, obtained using the images for all the TEs, and the computational time needed. The results are summarized in Table 1.

Model	Error	Time [s]
Linear Least Squares	1.1016×10^4	1.6653
Weighted Least Squares	1.0203×10^4	3.5794
Non-Negative Least Squares	6.4392×10^3	5.3702
Non-Linear Least Squares	5.5086×10^3	62.023

TABLE 1: Comparison of average fitting errors and execution times for algorithms modeling a single-compartment T2 decay.

The minimum error is obtained by the Non-Linear Least Squares even though its computational time is by far the highest. Depending on time and resources we could consider it the best option, or alternatively choose the Non-Negative Least Squares that has a higher error but a much lower computational time. On the other hand, both the Linear Least Squares and the Weighted Least Squares perform poorly and are not recommended.

2.3 Two-compartments T2 Relaxometry

The one-compartment model doesn't reflect accurately the structure of the brain because it as-

sumes that each voxel contains a uniform tissue type. In reality, however, this is an excessive simplification of the reality [6, 8] because many voxels contain a mix of components, such as myelin water and intra/extracellular water, each with different T2 values. This result in a multi-compartment model described by the following equation:

$$S(\text{TE}, \{T2\}) = S_0 \sum_i v_i e^{-\text{TE}/T2} \qquad (2)$$

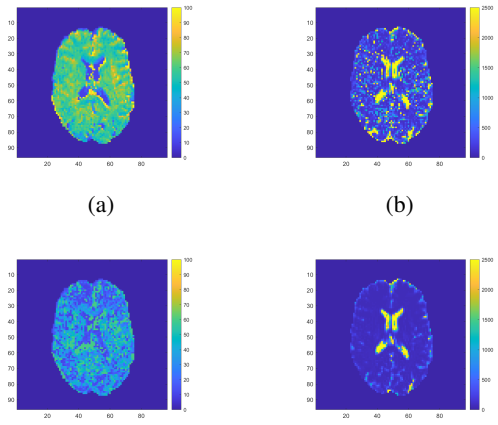
subject to $v_i \geq 0 \; \forall i$ and $\sum v_i = 1$, [1, 7, 8]

We started by exploring a simpler two-compartment model with different algorithms. We tried the Non-Negative Least Squares and the Non-Linear Least Squares with two approaches, estimating the T2 values and fixing them, focusing only on estimating S0 and the volume fraction v_1. Theoretically, literature suggests using three compartments with T2 of 10-50 ms for myelin water, $70-90$ ms for white and gray matter and 2 s for CSF [5, 6, 8, 9]. Using only two compartments we had to decide how to separate these values. For the NLLS algorithm, while estimating the T2 values, we opted to use a constraint of $10-110$ ms for the first compartment and $110-2500$ ms for the second compartment in order to separate CSF, which has a much higher T2 value, from the other parts of the brain. For the other two models, instead, we used fixed T2 values of 60 ms and 2000 ms for the two compartments. The average errors are reported in Table 2. These results are heavily influenced by the choice of the constraints or the initialization of the fixed T2 values. In fact, we could have chosen to separate myelin water from the other brain tissues, regrouping WM, GM and CSF, obtaining very different T2 maps, as shown in Figure 4.

Model	Error
NLLS with fixed T2 values	5.1954×10^3
NLLS with estimated T2 values	3.8787×10^3
NNLS	5.1954×10^3

TABLE 2: Comparison of average fitting errors for algorithms modeling a two-compartment T2 decay.

Using the segmentations in the datasets it is possible to estimate the mean values and the confidence intervals for each parameter. However, the results obtained are significantly different from the expected values. In fact, even though the T2 maps seem overall correct, there are many voxels that do not reflect the underlying biological structure. For example, using the NLLS with varying T2 values we obtained for the CSF an average volume fraction v_1 of 0.43, while in theory it is expected to be close to 0. The discrepancy is likely due to the use of only two compartments, which do not allow us to correctly estimate the tissue composition, leading to numerical inaccuracies.



(a) (b) (c) (d)

Fig. 4: T2 maps obtained using the NLLS model with two compartments under different constraint settings. Subfigures (a) and (b) use compartmental T2 ranges of [10–110] ms and [110–2500] ms, respectively, while (c) and (d) apply ranges of [10–60] ms and [60–2500] ms.

Comparing the results in Table 1 and in Table 2 we can conclude that the two-compartment models perform better than the one-compartment models. In particular, the Non-Linear Least Squares algorithm applied to the two-compartment model, estimating the T2 values, obtain the minimum mean error. We then decided to use the Parametric-Bootstrap Procedure with 1000 iterations to investigate the parameters precision for the NLLS algorithm for several voxels and the results for a single voxel are displayed in Table 3. The confidence intervals show plausible parameter estimates and narrow confidence ranges, indicating a good fit and model adequacy.

Parameter	95% range
$S0$	[1180.8560, 1180.9226]
$v1$	[0.74534, 0.74548]
$T2_1$	[42.8861, 42.8980]
$T2_2$	[2480.3068, 2499.9999]

TABLE 3: 95% confidence intervals computed via Parametric Bootstrap for the two-compartment NLLS model at voxel (45, 45, 28).

We have seen that the model with the lowest residual error so far is the two-compartment NLLS. However, this model requires estimating four parameters ($S0, v1, T2_1, T2_2$), while NNLS needs to estimate only $S0$ and $v1$, since it relies on fixed T2 values. To investigate the ability to fit multiple parameters we used the Akaike Information Criterion using the corrected version (AICc) to account for the limited number of measurements. We chose the middle slice and calculated the AICc value for both models, selecting for each voxel the model with the lowest AICc. The resulting map is shown in Figure 5. We observe that the two models perform similarly in terms of the number of voxels where each achieves the lowest AICc. For this reason, we chose to proceed with the NLLS model, as it yields lower residual errors and computational time is not a critical concern in our analysis.

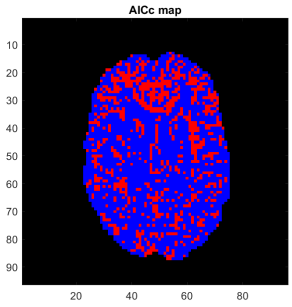


Fig. 5: AICc-based model selection map for slice 28. Blue voxels indicate regions where the NNLS model outperforms NLLS (lower AICc), while red voxels indicate the opposite.

To support this choice we calculated the theoretical upper bound of the residual error. In fact, assuming Gaussian noise, the residual error per voxel should follow the relation error $\leq N\sigma^2$, where $N = 19$ is the number of measurements (TEs). To estimate the noise standard deviation we selected a ROI of $[1-20, 1-20]$ that, according to the provided mask, contains only background voxels and where the signal is assumed to consist solely of noise. We calculated the average standard deviation across TEs and obtained $\sigma = 1.6486 \times 10^1$, resulting in a maximum error of 5.1639×10^3. This value is higher than the error obtained with NLLS but lower than that obtained using the NNLS algorithm.

2.4 Multi-component T2 Relaxometry

We extended the previous models to multiple compartments and, in particular, we implemented the NLLS algorithm for 3 and 4 compartments and the NNLS for 3 and 10 compartments. Their respective errors are presented in Table 4.

Model	Error
NLLS for 3 comp model	3.8162×10^3
NLLS for 4 comp model	3.7126×10^3
NNLS for 3 comp model	4.6905×10^3
NNLS for 10 comp model	3.7956×10^3

TABLE 4: Comparison of average fitting errors for algorithms modeling a multi-compartment T2 decay.

It is evident that increasing the number of compartments leads to a lower residual error. However, the incremental improvement diminishes as more compartments are added. Using more compartments could be useful to distinguish different brain structures. In a model with three compartments we can separate effectively myelin water, white and gray matter, and CSF [5, 6, 8, 9]. Adding even more compartments could allow us to better analyze different brain tissues, considering that different WM and GM structures, such as Internal Capsule, Splenium, Frontal White and others, decay at different rates [8]. Furthermore, different brain pathologies can produce unique decay curves [8]. However, the number of compartments cannot be increased indefinitely, as it is constrained by the number of available measurements, which in this case is 19. To select the most appropriate model for the remainder of our analysis, we referred to previous studies that successfully employed a three-compartment approach. This choice aligns with the principle of Occam's Razor, favoring simpler models that avoid unnecessary complexity. For the fitting algorithm, we selected NLLS due to its lower residual error compared to alternative methods.

We used parameter priors to try to speed up the fit and improve the accuracy. For each voxel, we assigned the corresponding segmentation values to the three volume fractions, normalizing them to ensure their sum equals one. Results from the middle slice indicate a modest improvement: without priors, the average error was 4.4516×10^3 with a computational time of 7.0259×10^1, whereas using priors reduced the error to 4.4430×10^3 and the computational time to 6.7801×10^1. The improvements are limited, likely due to inaccuracies in the segmentation and numerical imprecision in the optimization process, but promising. Better results could be achieved by studying in detail the structure of the segmentation and by using different priors for each voxel.

2.5 Second Dataset

For this part of the project we analyzed the imaging data obtained from 143 subjects at 19 years of age; 89 are EPT while 54 are full-term(FT). The numbers 90, 107, 115 and 129, all EPT, are excluded because their data are acquired at different TEs than the others adolescents, leaving us with 85 EPT subjects. The T2-weighted images were obtained at 10 echo-times, $TE = 13, 16, 20, 25, 30, 40, 50, 85, 100, 150ms$. For each subject we also have a brain mask, a segmentation, a white matter parcellation that shows the different lobes and a parcellation that highlights the corpus callosum and the internal capsule. Moreover, we have information about height, weight, gender and IQ for most of the subjects.

3 Results

3.1 Overall Brain

We applied the previously described NLLS algorithm to a three-compartment model for all subjects, saving the results in the corresponding parameter maps. We then computed the mean values of the different parameters across the whole brain and compared the results between EPT and FT subjects, also distinguishing between male and female subgroups. In particular, for the shortest T2 component, associated with myelin water and representing the main focus of our study, we obtained the following values: EPT = 2.1769×10^1, FT = 2.2726×10^1, M-EPT = 2.1278×10^1, F-EPT = 2.2130×10^1, M-FT = 2.3394×10^1, F-FT = 2.2267×10^1. We observe that all EPT values are lower than the corresponding FT ones.

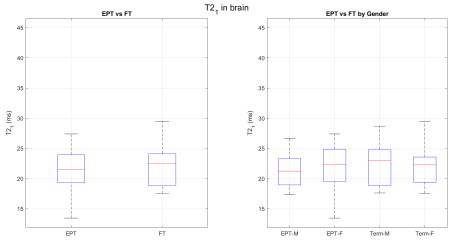


Fig. 6: Comparison between average $T2_1$ values across every brain veoxel for EPT and FT subjects.

However, as shown in Figure 6 and confirmed by the p-value ($p = 0.096$) obtained by applying the built-in `ttest2` function in MATLAB, the difference is not statistically significant. In this test, and in the next ones, we set the null hypothesis $H_0 : \mu_{\text{EPT}} \geq \mu_{\text{FT}}$, that the mean of $T2_1$ for EPT subjects is greater than or equal to that of FT subjects. For the other parameters we obtained similar results; the differences between EPT and FT are low and not statistically significant. We also computed the correlations between the different demographic data and all the parameters. Table 5 shows that there is no strong correlation between any of the parameters and predictors.

3.2 White Matter

We also focused on the white matter regions, where myelin water is more abundant [6]. We started by analyzing the entirety of WM segmentation obtaining for $T2_1$ EPT = 2.4432×10^1, FT = 2.5552×10^1 with p-value ($p = 0.095$). We repeated the procedure described in the previous section, trying to find correlations between height, weight, gestation weeks, gender and the estimated parameters. The results are showed in Table 6 and seem to indicate that there is no strong correlation between these two categories. We then used the parcellation provided to compute the parameters average for the different lobs of the brain : Frontal, Temporal, Parietal, Occipital, Cingulate and Insula. Figure 7 shows a similar trend for all the lobs; the $T2_1$ average value for EPT is lower than the one for FT but not in a significant way.

Predictor	Parameter	p-value	$\|r\|$
GestWeeks	$T2_2$	0.0075	0.226
IQ	$S0$	0.0341	0.185
Sex	$S0$	0.0381	0.176
IQ	$T2_2$	0.0482	0.173
Sex	$T2_2$	0.0528	0.165
Height	$T2_3$	0.1957	0.131
GestWeeks	$T2_1$	0.1323	0.128
Height	$T2_2$	0.2323	0.121
Weight	$S0$	0.2750	0.111
Height	$S0$	0.3274	0.099
GestWeeks	$T2_3$	0.3021	0.088
IQ	$T2_3$	0.4956	0.060
GestWeeks	$S0$	0.4884	0.059
Weight	$T2_2$	0.6907	0.040
IQ	$T2_1$	0.6901	0.035
Weight	$T2_1$	0.7653	0.030
Sex	$T2_3$	0.7462	0.028
Weight	$T2_3$	0.8182	0.023
Sex	$T2_1$	0.8779	0.013
Height	$T2_1$	0.9683	0.004

TABLE 5: Correlation between global brain parameters and demographic predictors.

Predictor	Parameter	p-value	$\|r\|$
Sex	$T2_2$	0.0244	0.191
Height	$T2_2$	0.0730	0.181
Height	$T2_3$	0.0766	0.179
IQ	$S0$	0.0652	0.162
IQ	$T2_2$	0.0933	0.147
Sex	$S0$	0.0973	0.141
GestWeeks	$T2_2$	0.0988	0.141
GestWeeks	$T2_1$	0.1593	0.120
Sex	$T2_3$	0.2821	0.092
Height	$S0$	0.4019	0.085
Weight	$S0$	0.4955	0.069
Weight	$T2_2$	0.5107	0.067
IQ	$T2_1$	0.6432	0.041
Weight	$T2_3$	0.7808	0.028
GestWeeks	$S0$	0.7659	0.025
Sex	$T2_1$	0.7693	0.025
Height	$T2_1$	0.8404	0.020
IQ	$T2_3$	0.8908	0.012
GestWeeks	$T2_3$	0.9099	0.010
Weight	$T2_1$	0.9497	0.006

TABLE 6: Correlation between WM parameters and demographic predictors.

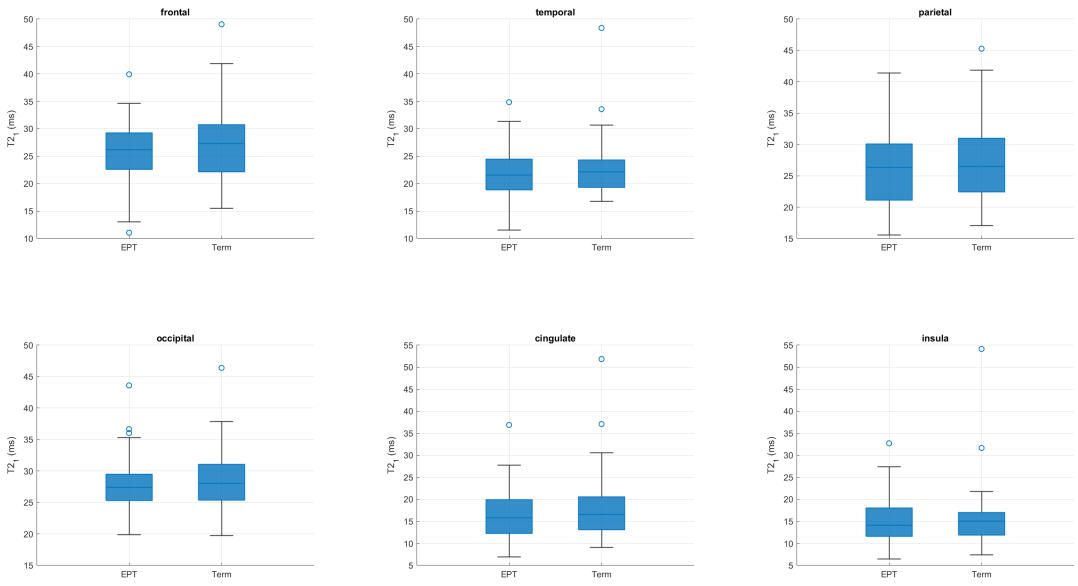


Fig. 7: Comparison between average $T2_1$ values across WM lobs for EPT and FT subjects: (a) Frontal, (b) Temporal, (c) Parietal, (d) Occipital, (e) Cingulate, (f) Insula

3.3 Corpus Callosum and Internal Capsule

We performed a detailed analysis of the Corpus Callosum and the Internal Capsule, two brain structures rich in myelin water that are crucial for understanding the long-term effects of prematurity in extremely preterm (EPT) individuals [2]. For the Corpus Callosum the results are displayed in Table 7.

Parameter	EPT (mean ± std)	Term (mean ± std)
S0	2227.89 (312.65)	2241.53 (299.47)
$T2_1$	18.89 (4.75)	16.85 (5.40)
$T2_2$	71.42 (3.70)	69.14 (2.58)
$T2_3$	2129.16 (48.68)	2124.84 (75.16)

TABLE 7: Mean and standard deviation of parameters in the corpus callosum for EPT and FT subjects.

The average values for the two cohorts are very similar and fall within each other's standard deviation ranges. This observation is supported by the p-values, which do not provide sufficient evidence to suggest a significant difference between the groups.

Table 8 presents the same results for Internal Capsule.

Parameter	EPT (mean ± std)	Term (mean ± std)
S0	2212.11 (364.59)	2276.48 (359.98)
$T2_1$	15.89 (4.09)	16.60 (6.84)
$T2_2$	67.89 (3.14)	67.61 (2.38)
$T2_3$	2122.60 (46.09)	2134.60 (106.88)

TABLE 8: Mean and standard deviation of parameters in the internal capsule for EPT and FT subjects.

4 Discussion

In the first part of our analysis, we confirmed the hypothesis that most voxels contain more than one compartment, demonstrating that the multi-compartment model outperforms the single-compartment model. We also highlighted issues related to the MRI scanning procedure and discussed how parameter priors can be a helpful tool to overcome these challenges. More advanced techniques could be used to estimate more accurate priors for volume fractions, particularly to achieve clearer segmentation of myelin water fractions within each voxel, which would allow us to improve the speed and quality of the fitting process.

In the second part of our study, we investigated whether EPT subjects exhibit less myelinated brain regions compared to FT subjects. To this end, we focused on analyzing differences in $T2_1$ values. A summary of the results is presented in Table 9. We did not find any statistically significant differences between EPT and FT subjects across the regions analyzed. However, we observed lower average $T2_1$ values in EPT adolescents for all segmentations except for the corpus callosum. This suggests that further studies may be necessary before drawing definitive conclusions. We also attempted to identify correlations between the estimated parameters and demographic variables such as height, weight, sex, and IQ. No strong correlations were observed, nor was there any evidence suggesting clear differences between male and female subjects.

Several methodological factors may help explain these results. During the fitting process, some voxels produced warnings indicating that the system matrix was close to singularity. This may have led to unstable or poorly scaled parameter estimates in those voxels, compromising the reliability of the results. In addition, we observed some values that were clearly out of range, as seen in the outliers present in the boxplots shown in Figure 7, which further supports the possibility of numerical instability or fitting inaccuracies in specific regions. Furthermore, the assumption of Gaussian noise could be wrong and different models, such as Rician noise, could allow us to obtain better results.

Moreover, as observed in the first part of the analysis, the model struggles to accurately distinguish between tissue compartments in boundary voxels, where partial volume effects are more evident. This likely limits the ability to isolate the myelin signal precisely, particularly in regions with complex anatomical boundaries.

Improvements such as more robust segmentation, better prior estimation, and a more stable fitting routine could increase the likelihood of detecting significant group differences in future studies.

Region	EPT (mean ± std)	Term (mean ± std)	p-value
brain	21.77 (3.30)	22.73 (4.66)	9.63e-02
wm	24.43 (4.09)	25.55 (5.32)	9.53e-02
frontal	25.86 (4.94)	27.39 (6.12)	6.27e-02
temporal	21.79 (3.92)	22.72 (4.96)	1.24e-01
parietal	26.15 (5.44)	27.06 (6.28)	1.93e-01
occipital	27.57 (3.96)	28.37 (4.71)	1.54e-01
cingulate	16.46 (5.27)	17.94 (7.55)	1.05e-01
insula	15.25 (4.64)	15.66 (6.68)	3.45e-01
corpus callosum	18.89 (4.75)	16.85 (5.40)	9.87e-01
internal capsule	15.89 (4.09)	16.60 (6.84)	2.49e-01

TABLE 9: Comparison of $T2_1$ values across brain regions for EPT and FT subjects. Values are reported as mean (standard deviation).

References

[1] W. H. Organization, N. . C. H. The Partnership for Maternal, and S. the Children, "Born too soon: the global action report on preterm birth," 2012.

[2] B. Laureano, H. Irzan, S. Ourselin, N. Marlow, and A. Melbourne, "Myelination of preterm brain networks at adolescence," 2024, school of Biomedical Engineering & Imaging Sciences, King's College London, UK; Dept. of Medical Physics and Biomedical Engineering, University College London, UK; Institute for Women's Health, University College London, UK.

[3] H. C. Glass, A. T. Costarino, S. A. Stayer, C. Brett, F. Cladis, and P. J. Davis, "Outcomes for extremely premature infants," *Anesthesia and Analgesia*, vol. 120, no. 6, p. 1337, 2015.

[4] C. e. a. Hagmann, "T2 at mr imaging is an objective quantitative measure of cerebral white matter signal intensity abnormality in preterm infants at term-equivalent age," *Radiology*, 2009, 252, 209–217.

[5] C. e. a. Laule, "Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology," *Multiple Sclerosis*, 2006, 12, 747–757.

[6] E. Dingwall, J. A. de Zwart, J. H. Duyn, and A. Melbourne, "Insights into brain microstructure with t2 distribution mri," *NeuroImage*, vol. 256, p. 119271, 2022.

[7] C. Laule, P. Kozlowski, E. Leung, D. Li, A. Mackay, and G. Moore, "Myelin water imaging of multiple sclerosis at 7 t: correlations with histopathology," *Neuroimage* 2008;40(4):1575–80.

[8] A. MacKay, C. Laule, I. Vavasour, T. Bjarnason, S. Kolind, and B. Mädler, "Insights into brain microstructure from the t2 distribution," *Magn Reson Imaging* 2006;24(4):515–25.

[9] T. Prasloski, B. Mädler, Q.-S. Xiang, A. MacKay, and C. Jones, "Applications of stimulated echo correction to multicomponent t2 analysis," *Magn Reson Med* 2012;67(6):1803–14.