Seminar paper

Cognitive and brain science

Iron, Lipid & MRI correlations

written by Shirly Eliezer,

Mezer lab, ELSC

Table of contents

Abstract

Quantitative magnetic resonance imaging (qMRI) provides biophysical parametric measurements such as T1, T2, T2*, MT, MTV. These parameters carry information about the local microstructural environment of the protons (such as myelin in the brain). In this research, we will focus on T1. T1, longitudinal relaxation time, the time constant which determines the rate at which excited protons return to equilibrium. The literature indicates that lipids have a strong effect on the contrast of brain qMRI maps. The iron content and water fraction (WF) of cellular compartments are also known to influence the qMRI parameters. There have been a few quantitative attempts to find a relationship between lipid content and T1. Additionally, it is unclear how much iron content contributes to T1 tissue contrast. In this research, we aim to find the best model that describes the relationship between iron and lipid content to the qMRI parameters, while focusing on the longitudinal relaxation rate parameter, R1 (1/T1).

Introduction

Advances in the field of magnetic resonance imaging (MRI) have led to the development of quantitative MRI (qMRI). qMRI provides biophysical parametric measurements that are useful in the investigation and diagnosis of normal and abnormal biological processes, such as aging, also for neurological diseases such as Parkinson's, cancerous lesions, for normal physiological processes such as the formation and treatment of myelin, respiratory processes and more (Callaghan, et al., 2014; Yeatman, Wandell & Mezer, 2016; Gracien et al., 2017; Moffat et al. 2005; Hamstra et al. 2005).

qMRI is aimed at the direct measurement of the physical tissue properties.
The tissue can be characterized by two different relaxation times - T1 and T2. T1 (longitudinal relaxation time) is the time constant that measures the time at which excited protons return to equilibrium. In this study we will focus on the longitudinal relaxation rate (R1=1/T1). It is a measure of the rate taken for spinning protons to realign with the external magnetic field. (Neurology, 2016).

qMRI enables the creation of parametric maps which allows a reliable comparison of brain structure across different time points and different MRI scanners, making it possible to assess normal brain development, as well as pathological conditions.

The human brain is comprised mainly of water (70–80%), proteins (8–11%) and lipids (5 15%) (Samorajski & Rolsten, 1973; Shtangel & Mezer, 2020). The distribution of these

molecules varies between brain regions, across lifespan, and in different pathological states. Lipids are known to strongly affect the contrast of brain qMRI maps (Shtangel & Mezer, 2020).

Iron is an important metal involved in various physiological processes, such as ATP generation and DNA replication (Chang, 2019; Mills et al., 2010; Qian & Ke, 2019). Particularly, iron is essential for a variety of neurological processes (McCarthy & Kosman, 2015; Rouault, 2013). Iron transport in the brain is effectuated by several pathways; namely, transferrin-dependent iron transport, non-transferrin bound iron (NTBI) mobilization, uptake and export by and from neurons, oligodendrocytes, astrocytes, and microglia (Hohnholt & Dringen, 2013; Roy Sarkar & Dutta, 2019). Furthermore, Ferritin is the main iron storage protein, conformed by two types of subunits, H type (heavy) and L type (light), which co-assemble into a supramolecular spherical-shaped protein (Chang, 2019). The iron content and water fraction (WF) of cellular compartments are thought to influence qMRI parameters (Stüber et al., 2014).

There have been a few quantitative attempts to find a relationship between lipid content and T1 (Bot et al., 2004; Mottershead et al., 2003; Schmierer et al., 2004).
Due to restricts in conventional imaging techniques in directly measuring the concentration of the lipids in the human brain, the relationship between T1 and lipids remains vogue. Additionally, it is unclear how much iron content contributes to T1 tissue contrast (Gelman et al., 2001). The literature regarding iron and T1 contrast is controversial: data showing a clear relationship between brain T1 and iron concentration (Ogg and Steen, 1998; Vymazal et al., 1995) contradict other publications showing no significant correlation of T1 and iron (Steen et al., 2000).

The main objective of this work was to deduce a model representing the effect of lipids and iron on qMRI parameters, specifically R1. In the research we started with a linear model to evaluate the relationship between iron, lipid and R1. During the results and predictions validation, the accuracy wasn't satisfied as I will show later in this paper, and complicated models were created.

Methods

Data collection

This study describes a phantom system designed to assess the contribution of various membrane lipids and iron forms to qMRI parameters using a clinical human scanner.
Three lipid types have been chosen: phosphatidylcholine (PC), a mixture of phosphatidylcholine-sphingomyelin (PC_SM), and a mixture of phosphatidylcholine sphingomyelin (PC_SM), and a mixture of phosphatidylcholine-cholesterol

(PC_Choles). In addition, 3 iron types have been chosen for the experiment: free iron Fe^{+2}, Ferritin, Transferrin.

The phantom systems are composed of several different samples in a glass box; each contains approximately 29 test tubes. In each box we can control the lipid type and iron type and their concentrations. Each test tube in each box contains a different concentration of lipid and a different concentration of iron, as seen in fig. 1.

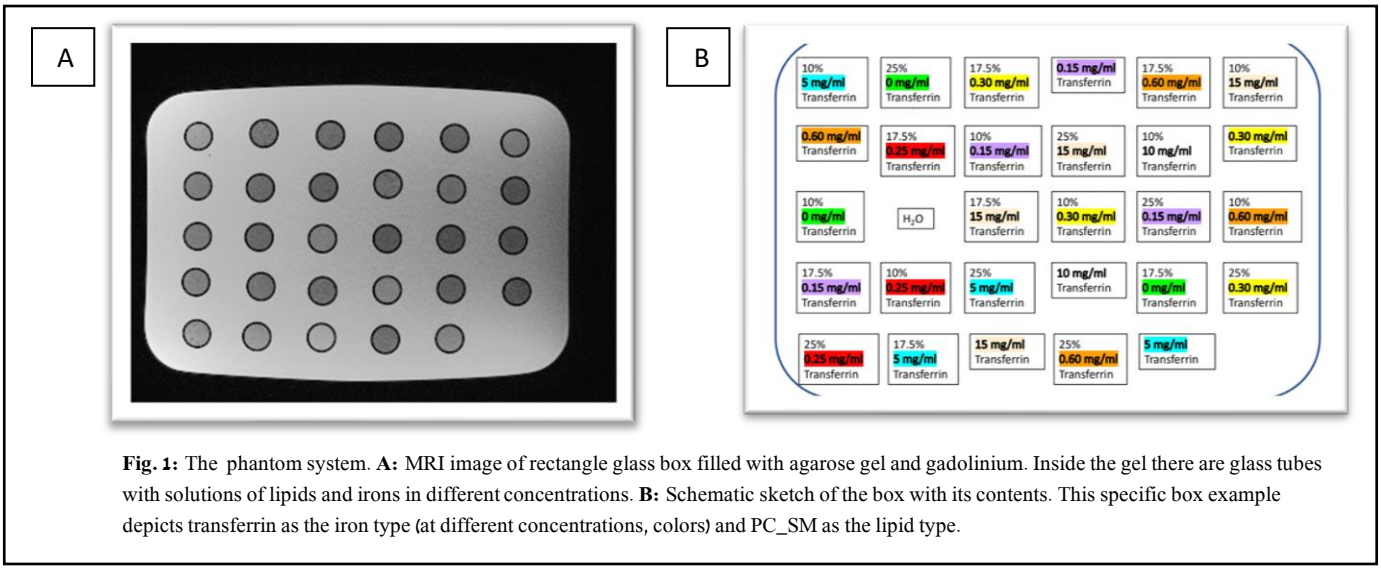

Fig. 1: The phantom system. **A:** MRI image of rectangle glass box filled with agarose gel and gadolinium. Inside the gel there are glass tubes with solutions of lipids and irons in different concentrations. **B:** Schematic sketch of the box with its contents. This specific box example depicts transferrin as the iron type (at different concentrations, colors) and PC_SM as the lipid type.

Each box was scanned in 3 T MRI machine, and qMRI parameters were established. The entire data of the phantom system is represented in a csv file as seen in fig. 2., The table contains all the samples information (lipid type, iron type, lipid concentration, and iron concentration), and the qMRI parameters as received from qMRI analyses.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	[Protein](mg/ml)	iron [mg/ml]	lipid [%]	MTV exp	R1 [1/sec]	R2s [1/sec]	MTV [fraction]	R2 [1/sec]	MT [p.u.]	type	estimation [mg	iron type	Lipid type	ExpNum	Fe] sigma [mg/ml	Date	Exp. Folder nam
2	0	0	10	0.1	0.367098478	0.003427238	0.089390428	2.96297588	0.00391594	PC+Chol+Fe2	0.0000	Fe2	PC_Choles	1	0.0000	10.12.19	PC_Chol_Fe_1(
3	0	0.15	17.5	0.175	1.281386236	0.00457829	0.14862217	4.78468786	0.006907897	PC+Chol+Fe2	0.1500	Fe2	PC_Choles	1	0.1500		
4	0	0.6	25	0.25	2.65315135	0.013634661	0.240180643	8.23043163	0.007324512	PC+Chol+Fe2	0.6000	Fe2	PC_Choles	1	0.6000		
5	0	0.05	0	0	0.329832368	-9.01334E-05	-0.013546626	1.76211518	-0.000342001	Fe2	0.0500	Fe2	PC_Choles	1	0.0500		
6	0	0.3	17.5	0.175	1.834075168	0.00699874	0.126721888	5.61798578	0.005469975	PC+Chol+Fe2	0.3000	Fe2	PC_Choles	1	0.3000		
7	0	0.3	0	0	0.633609147	0.001351296	-0.000888862	3.0534457	-6.25996E-05	Fe2	0.3000	Fe2	PC_Choles	1	0.3000		
8	0	0	25	0.25	0.592734978	0.010961047	0.275886846	4.72813326	0.012138631	PC+Chol+Fe2	0.0000	Fe2	PC_Choles	1	0.0000		
9	0	0.25	0	0	0.493848493	0.000376406	-0.010080424	2.45399863	0.000359091	Fe2	0.2500	Fe2	PC_Choles	1	0.2500		
10	0	0.05	17.5	0.175	0.750882823	0.004305815	0.169796811	3.60360366	0.007378913	PC+Chol+Fe2	0.0500	Fe2	PC_Choles	1	0.0500		
11	0	0.3	10	0.1	2.180984699	0.006359395	0.130476372	6.51465274	0.004760396	PC+Chol+Fe2	0.3000	Fe2	PC_Choles	1	0.3000		
12	0	0.25	25	0.25	1.658634983	0.00876514	0.23682036	6.0422848	0.010123455	PC+Chol+Fe2	0.2500	Fe2	PC_Choles	1	0.2500		
13	0	0.6	17.5	0.175	2.768887169	0.010160625	0.167370712	8.06453547	0.006439343	PC+Chol+Fe2	0.6000	Fe2	PC_Choles	1	0.6000		
14	0	0.15	10	0.1	1.147384627	0.004041637	0.105632023	4.33839451	0.004343224	PC+Chol+Fe2	0.1500	Fe2	PC_Choles	1	0.1500		
15	0	1	0	0	1.494054506	0.004055194	0.024929104	5.36193461	0.001527849	Fe2	1.0000	Fe2	PC_Choles	1	1.0000		
16	0	0.05	25	0.25	0.842075043	0.008849045	0.242251104	4.77326997	0.012483551	PC+Chol+Fe2	0.0500	Fe2	PC_Choles	1	0.0500		
17	0	0	10	0.1	0.382233151	0.003213436	0.077758948	2.16216242	0.004182086	PC+Chol+Fe2	0.0000	Fe2	PC_Choles	1	0.0000		
18	0	0.05	17.5	0.175	0.762224151	0.005649447	0.181561389	4.00000044	0.007826045	PC+Chol+Fe2	0.0500	Fe2	PC_Choles	1	0.0500		
19	0	0.15	25	0.25	1.054876318	0.010700769	0.244766773	5.6022421	0.010445322	PC+Chol+Fe2	0.1500	Fe2	PC_Choles	1	0.1500		
20	0	0	17.5	0.175	0.475479431	0.007354696	0.182084621	3.05342234	0.008516913	PC+Chol+Fe2	0.0000	Fe2	PC_Choles	1	0.0000		
21	0	0.6	10	0.1	3.181948591	0.010960248	0.130255312	8.88887102	0.005301018	PC+Chol+Fe2	0.6000	Fe2	PC_Choles	1	0.6000		
22	0	0.25	17.5	0.175	1.488582835	0.006967736	0.155875101	4.66200339	0.006717807	PC+Chol+Fe2	0.2500	Fe2	PC_Choles	1	0.2500		
23	0	0.8	0	0	1.240634721	0.003136427	-0.01177233	4.64037191	0.00062741	Fe2	0.8000	Fe2	PC_Choles	1	0.8000		
24	0	0.15	25	0.25	1.078673729	0.009811247	0.241549351	5.08906071	0.009822066	PC+Chol+Fe2	0.1500	Fe2	PC_Choles	1	0.1500		
25	0	0.6	0	0	1.035233518	0.003508214	-0.012988652	4.35729995	0.000658026	Fe2	0.6000	Fe2	PC_Choles	1	0.6000		
26	0	0.05	10	0.1	0.617838319	0.003592084	0.081873383	3.00750283	0.003701782	PC+Chol+Fe2	0.0500	Fe2	PC_Choles	1	0.0500		
27	0	0.15	0	0	0.424632422	0.000650831	0.035031049	2.20994179	-0.000287742	Fe2	0.1500	Fe2	PC_Choles	1	0.1500		
28	0	0.25	10	0.1	1.509560911	0.004632622	0.096799733	4.88997429	0.004299738	PC+Chol+Fe2	0.2500	Fe2	PC_Choles	1	0.2500		
29	0	0.3	25	0.25	1.970620144	0.012155837	0.285428587	6.60060459	0.010578382	PC+Chol+Fe2	0.3000	Fe2	PC_Choles	1	0.3000		
30	0	0	25	0.25	0.338275552	0.007270488	0.22703061	4.59770414	0.000697194	PC+Fe2	0.0000	Fe2	PC	2	0.0000	30.10.19	PC_Fe_301019
31	0	0.3	0	0	0.605951816	0.000902777	0.012496483	2.87769071	0.000267144	Fe2	0.3000	Fe2	PC	2	0.3000		
32	0	0.25	25	0.25	1.091587123	0.010066123	0.226835081	5.30504451	0.001023545	PC+Fe2	0.2500	Fe2	PC	2	0.2500		
33	0	0	10	0.1	0.308433589	0.004230988	0.10190572	2.43902641	-2.59339E-05	PC+Fe2	0.0000	Fe2	PC	2	0.0000		
34	0	0.05	25	0.25	0.483213056	0.007377667	0.193518585	3.46620423	0.000427813	PC+Fe2	0.0500	Fe2	PC	2	0.0500		
35	0	0.25	10	0.1	1.060574208	0.005699485	0.076094729	5.24935809	0.000862248	PC+Fe2	0.2500	Fe2	PC	2	0.2500		
36	0	0.05	0	0	0.332805677	-0.001274485	-0.003531639	1.82648487	-0.000185845	Fe2	0.0500	Fe2	PC	2	0.0500		
37	0	0.15	10	0.1	0.705169857	0.003815084	0.110850459	3.91389348	0.000913524	PC+Fe2	0.1500	Fe2	PC	2	0.1500		
38	0	0	17.5	0.175	0.337929585	0.008873406	0.207902827	3.6764707	0.000103426	PC+Fe2	0.0000	Fe2	PC	2	0.0000		
39	0	0.15	25	0.25	0.841071732	0.008540202	0.211134403	4.61893798	0.000825125	PC+Fe2	0.1500	Fe2	PC	2	0.1500		

Fig. 2: The entire data after MRI scans. Represented in csv worksheet, 10 experiments (boxes) were scanned. In total, 3 iron types; Fe^{+2} (represented here as Fe^{+2} for simplicity), Ferritin, Transferrin, and 3 lipid types; PC_SM, PC, and PC_Choles, 4 lipid concentrations; 0, 10, 17.5, 25 and several iron concentrations according to iron type.

Pre-processing of the data

Out of 10 experiments (273 samples), 2 were disqualified. One experiment was disqualified due to inaccurate amounts of substances (lipid concentration) and the other was disqualified due to the fact it contains sample of only protein that was irrelevant for the current research.

Outliers

To detect outliers, Z-score was calculated. Z-score is a statistical measure represent the number of standard deviations away from the mean that a certain data point is. The equation of Z-score calculated as follows: $Z = \frac{x-\mu}{\sigma}$ where μ=mean of the R1 values X; σ=Standard deviation of the R1; X= R1. Z-score values greater than or less than + 3 or − 3, respectively, are considered outliers (Misra et al., 2019; Tabachnick & Fidell, 2013).

Visualization of the outliers was created by using a box plot. A boxplot is another convenient approach to identifying univariate outliers (Field & Miles, 2010; Tukey, 1977) which graphically depicting groups of numerical data through their quartiles. It has lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles. As seen in fig. 3, outliers can be detected outside the upper quartiles.

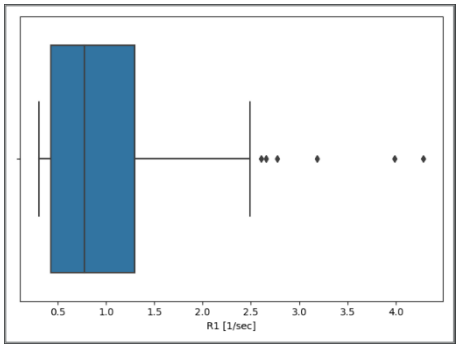

Fig. 3: visualization of outliers using Box Plot.

After calculating the Z-score for the R1 values, values with distance a of 3 standard deviations and more from the mean, were disqualified (3 samples).

After performing pre-processing methods on the data, 163 samples remain.

Relationship between the variables.

To assess the relationship between the independent variables and the dependent variable, Pearson's correlation coefficient was calculated. The correlation coefficient is a measure of linear correlation between two sets of data (Correlation Coefficient, 2021; Pearson, 1895). The Pearson correlation coefficient was calculated by dividing the covariance of the two variables, with the products of their standard deviations.

$P = \frac{COV(X,Y)}{\sigma x * \sigma Y}$. The covariance is calculated as follows: $COV(X,Y) = \frac{E[(X-\mu x)(Y-\mu Y)]}{\sigma x * \sigma Y}$, where μ, σ defined above.

Linear Regression Models

Simple linear regression models were created to examine the ability to predict R1 values by iron and lipid concentrations. Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (Freedman, 2009).

In this research, the response is the R1 (1/sec) values, which are derived from T1 results (R1 = 1/T1), and the explanatory variables are the lipid and iron concentrations represented as []. As mentioned earlier, the research focuses on 3 different iron types: Fe^{+2}, Ferritin, and Transferrin. In order to address the iron concentration which bounds to the proteins (Ferritin, Transferrin), we estimated the number of bound iron and calculated its concentration in the known amount of protein used.

In the first stage, simple and multiple linear regression were calculated. The models equations were:

1. R1 = a*[iron form] + b
2. R1 = a*[lipid type] + b
3. R1 = a*[iron form] + b*[lipid type] + c

The term Iron type represents the iron form; i.e., Fe^{+2}, transferrin, or ferritin. The term lipid type represents the lipid solution; i.e., PC, PC_SM, or PC_cholest. And the terms [Iron form] and [lipid type] represent the concentrations of the iron form and lipid type, respectively.

In the second stage, multiple linear regression with interaction model was created.

4. R1 = a*[iron form] + b*[lipid type] + c*[iron form]*[lipid type] + d.

The third stage included linear regression with categorical variables; iron types and lipid types, to understand how the different types affect the relationship between the substance's concentration and the target.

5. R1 = a*[lipid type]*lipid type + b * [iron form] * iron type + c.

To properly use the categorical variables, encoded with dummy variables was established. In this stage, a column of every iron type and lipid type was added to the data frame. The code '1' presented in the column represents the iron type used in each sample, as well as the lipid type (Fig. 4).

Lipid type_PC_SM	Lipid type_PC_Cholest	Lipid type_PC	Iron type_Transferrin	Iron type_Ferritin	Iron type_Fe2	[%] lipid	sigma [mg/ml] [Fe]	
0	1	0	0	0	1	10.0	0.000000	0
0	1	0	0	0	1	17.5	0.150000	1
0	1	0	0	0	1	25.0	0.600000	2
0	1	0	0	0	1	17.5	0.300000	4

Fig. 4: Dummy variables coding. 3 columns of iron types exchange the original column of the iron type, and also for the lipid type. The number 1 was written in a specific column to fit the types used in the sample. In this example, we can detect 4 samples with iron type Fe^{+2} and lipid type PC_Cholest.

For convenience, we use the equation R1 = a*[lipid type]*lipid type + b*[iron form]*iron type + c when in fact in the current technique, each iron type and each lipid type demonstrated a different coefficient.

When addressing to the different types of iron and lipids, the current model (model 5) performs the interaction between iron concentrations and iron type, and lipid concentrations to lipid type as follows; each type matches to different coefficient (slope). The full equation was written as: R1 = a*[iron form]*iron_type_Fe2 + b*[iron form]* iron _type_Ferritin + c*[iron form]* iron _type_Trans + d*[lipid type]*lipid_type_PC_SM + e*[lipid type]*lipid_type_PC_Cholest + f*[lipid type]*lipid_type_PC + g.
The data promises that in each sample, exactly one iron type is used, as well as the lipid type. So, for each combination of iron type and lipid type, we will get a different equation, based on the difference between the coefficient's types. In this way, when the sample involved Fe^{+2} as iron type and PC_Cholest as lipid type, for example, the regression equation is R1 = a*[Fe] + e*[lipid type] *lipid_type_PC_Cholest + g, after assigning 1 to iron_type_Fe2 and to lipid_type_PC_Cholest and 0 to iron_type_Ferritin, iron_type_Trans, lipid_type_PC_SM, and lipid_type_PC in the equation above.

To evaluate the goodness of fit of a model, R^2 coefficient of determination was calculated (Renaud et al., 2010). R^2 is a statistical measure of how well the regression predictions approximate the real data points. An R^2 of 1 indicates that the regression predictions perfectly fit the data. $R^2 = \frac{TSS - RSS}{TSS}$ where RSS is the sum of squares of residuals, a measure of the discrepancy between the data and an estimation model and TSS is the total sum of squares - sum over all squared differences between the observations and their overall mean.

Furthermore, the mean absolute error was calculated. The mean absolute error is a way of comparing forecasts with their eventual outcomes (Willmott et al., 2005). The calculation is as follows $MAE = \frac{\sum_{k=1}^{n} |y_i - \hat{y}|}{n}$ where n is the total number of the samples, y_i is specific value from the measured samples and $\hat{y}$ is the predicted value.

Cross validation

Cross-validation is a data resampling method to assess the generalization ability of predictive models and to prevent overfitting (Hastie et al., 2008; Duda et al., 2001).

A central question in supervised learning concerns the accuracy of the resulting model. Overfitting is the case where a model is perfectly adapted to the data set at hand but then unable to generalize well to new, unseen data (Berrar et al., 2013). In this research, leave one out cross-validation technique was obtained. the available learning set is partitioned into n disjoin. The model is trained using $n - 1$ subsets, which, together, represent the training set. Then, the model is applied to the remaining subset, which is denoted as the validation set, and the performance is measured. This procedure is repeated until each of the n subsets has served as validation set. The average of the n performance measurements on the n validation sets is the cross-validated performance. The test error in leave one out cross validati*on* is approximately an unbiased estimate of the true prediction error (Hastie et al., 2008).

Results

Pearson correlation coefficients was calculated for R1 target variable. It can be determined that correlation coefficient value smaller than 0.5 doesn't indicate a strong linear relation (Mukaka, 2012; Benesty et al., 2009). The results indicated low positive correlation between R1 and iron concentrations ($r(163) = 0.49$, $p < 0.05$), and negligible correlation between R1 to lipid concentrations ($r(163)=0.086$, $p > 0.2$).

Accordingly, simple Linear regression models were created (model 1, model 2). Both models yielded low R^2 values, indicating models that does not properly explain the dependent variable by the independent variable (Hamilton, Ghert & Simpson, 2015), as seen in fig. 5.

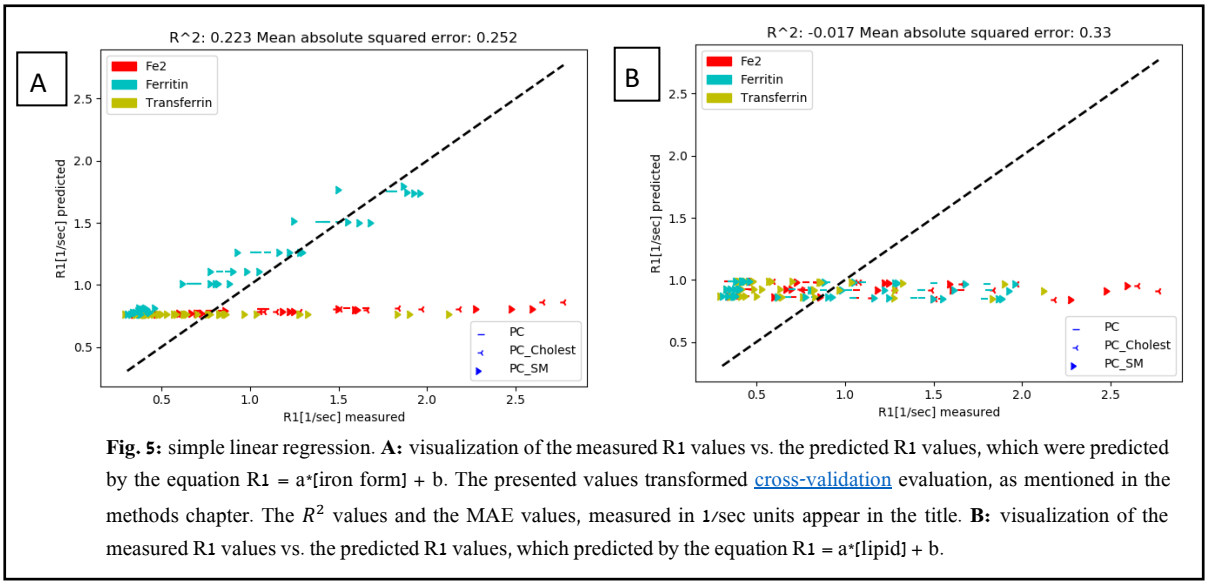

Fig. 5: simple linear regression. **A:** visualization of the measured R1 values vs. the predicted R1 values, which were predicted by the equation R1 = a*[iron form] + b. The presented values transformed cross-validation evaluation, as mentioned in the methods chapter. The R^2 values and the MAE values, measured in 1/sec units appear in the title. **B:** visualization of the measured R1 values vs. the predicted R1 values, which predicted by the equation R1 = a*[lipid] + b.

Multiple linear regression (model 3) and multiple linear regression with interaction (model 4) were performed to evaluate the abilities of iron concentrations and lipids concentrations to predict R1 values as can be seen in Fig. 6.

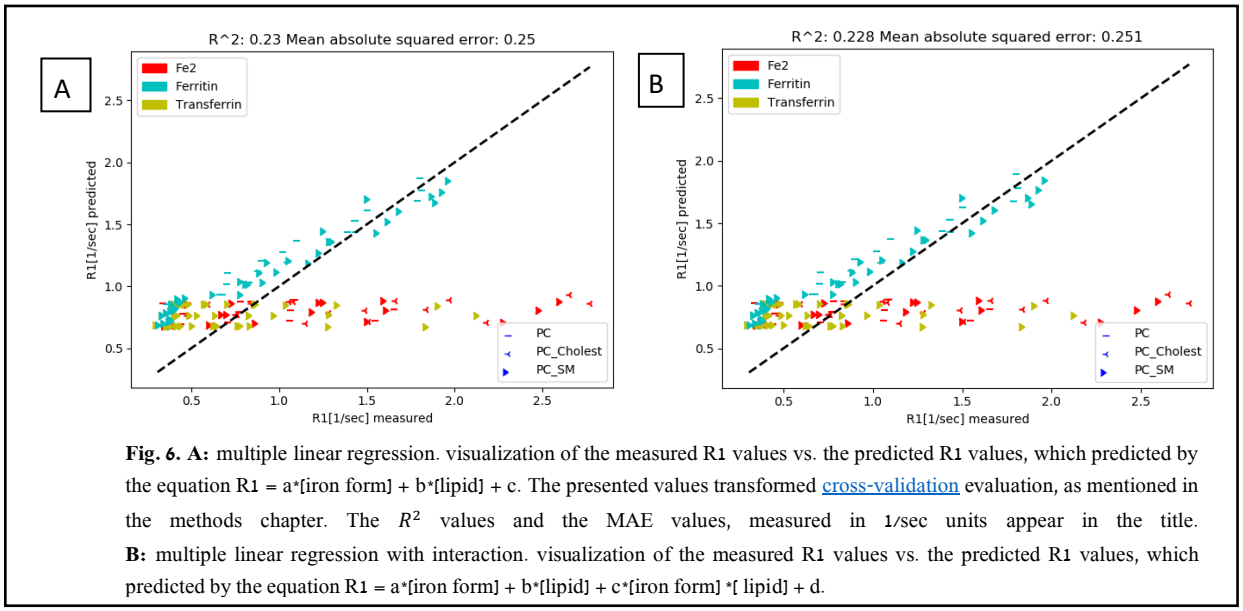

Fig. 6. A: multiple linear regression. visualization of the measured R1 values vs. the predicted R1 values, which predicted by the equation R1 = a*[iron form] + b*[lipid] + c. The presented values transformed cross-validation evaluation, as mentioned in the methods chapter. The R^2 values and the MAE values, measured in 1/sec units appear in the title. **B:** multiple linear regression with interaction. visualization of the measured R1 values vs. the predicted R1 values, which predicted by the equation R1 = a*[iron form] + b*[lipid] + c*[iron form] *[lipid] + d.

It can be identified for the first time in fig.6 that groups of samples with the same iron type, and the same lipid type, have common slope, and the clustering of the iron types is more noticeable than of the lipid types. We hypothesized that there is an effect of the iron types and the lipid types on the way the substances concentration effects R1 values. The data analysis revealed the relationship between the different types of iron and types of lipids. That is, depending on the **type** of iron and the **type** of lipid, there is effect of the iron and lipid concentrations on the target. Using seaborn library, a scatter plot created to was created to visualize the effect of iron types and lipid types on the substances concentration in the goal of predicting R1 values. In Fig. 7 one can clearly detect clusters of common types of lipids and irons.

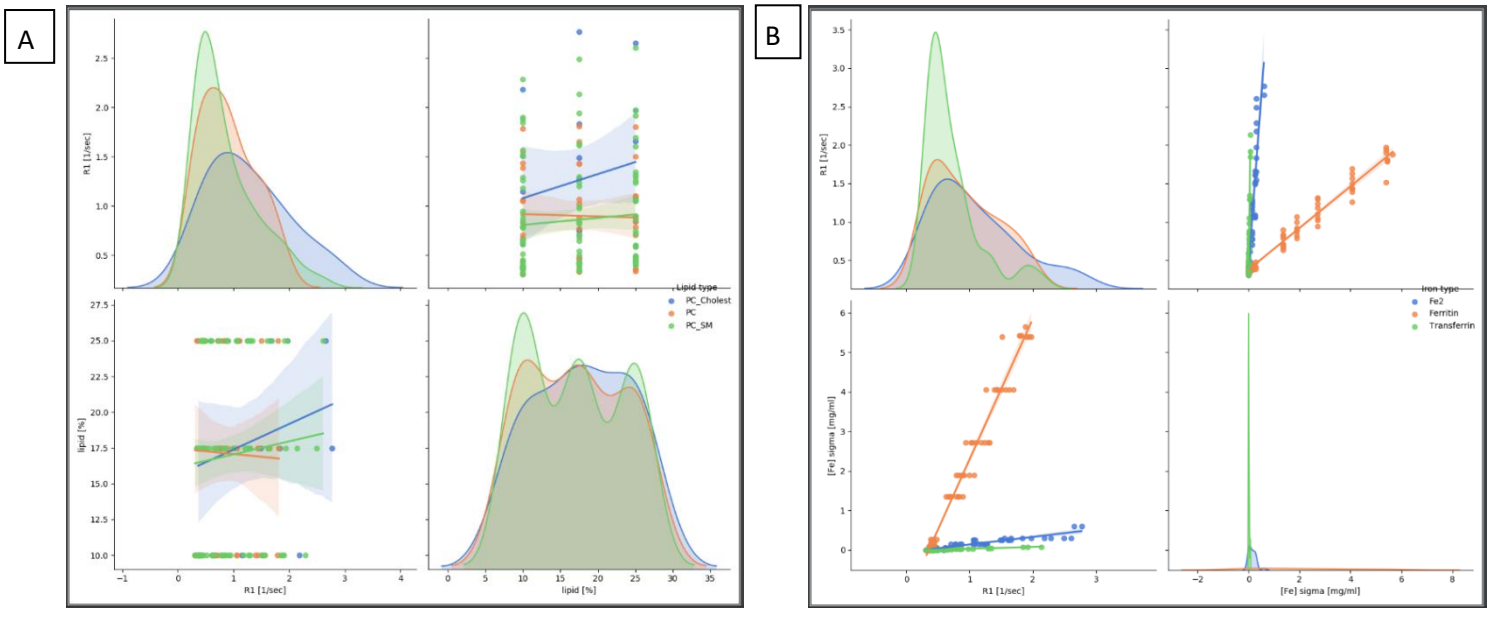

Fig. 7: relations between lipid to R1 and iron to R1 with the effect of lipids and iron types. **A:** pair plot depicting the relationship between lipid concentrations and R1 when common behaviors can be identified for all samples from the same type of lipid. **B:** pair plot depicting the relationship between iron concentrations and R1 when common behaviors can be identified for all samples from the same type of iron.

Due to the last results, linear regression involves interaction between categorical variables (iron types and lipid types) and continues variables (iron and lipid concentrations) was performed (model 5).

In agreement with the hypothesis, it was found that when referring to the effect of the types of iron and lipids on the amounts of iron and lipids, respectively, R1 predicted values approximate the real data points very well ($R^2 = 0.926, MAE = 0.024$). It can also be derived from the results that the type of iron contributed more to explain the data than the lipid type, as can be seen in Fig. 8B.

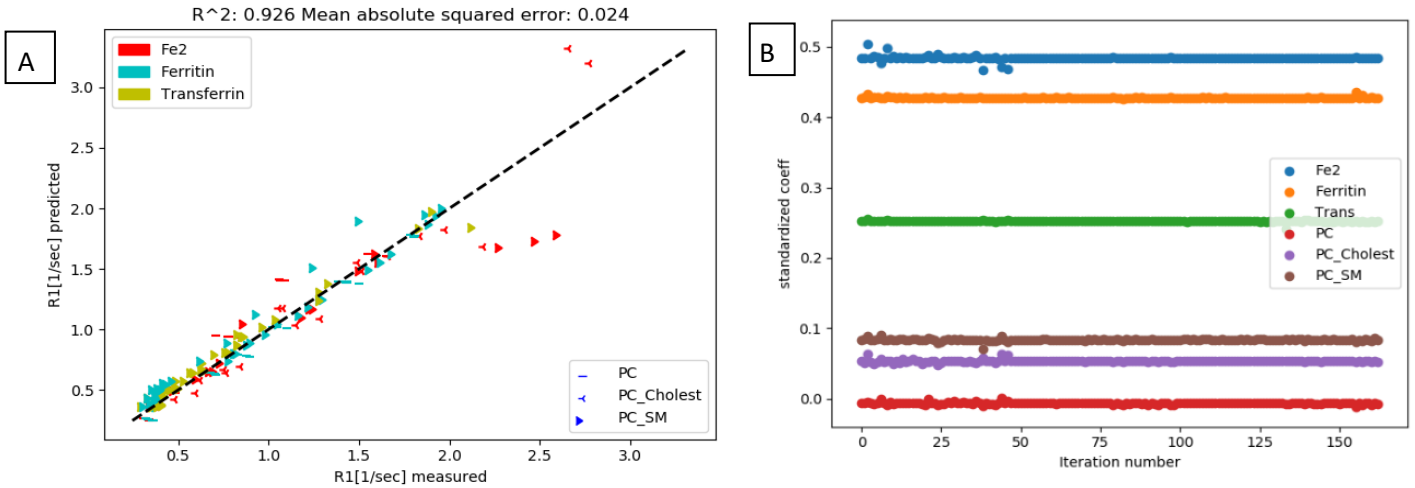

Fig. 8: A: linear regression involves interaction with categorical variables. Visualization of the measured R1 values towards the predicted R1 values, which predicted by the equation R1 = a*[iron form]*iron type + b*[lipid type]*lipid type + c. The presented values transformed cross-validation evaluation, as mentioned in the methods chapter. The R^2 values appearing in the title, and the MAE values, measured in 1/sec units. **B:** As mentioned, each type of iron and each type of lipid paired with another coefficient. The data was standardized using Z-score technique to be able to compare the different coefficients. The graph represents the coefficient of each type during each one of the cross-validation iterations. It can be identified that iron types coefficients are higher and have more variance between them, then lipid types coefficients. Due to that it can be derived that iron types explain the data in a better way then lipid types. The iron form Fe^{+2} has the most effect on the target, due to its highest coefficient value.

Further results

As mentioned at the beginning of this paper, qMRI parameters T2, T2*, MT and MTV were calculated as well. Attempts to find the best fit model that allows to predict each one of these parameters based on iron and lipids information were made. Similar to the process described in this work regarding R1, data analyses and models' creation were made for each parameter. The table below presents the best models chosen by the highest R^2 value and the lowest MAE value.

qMRI parameter	Model	R^2	MAE
R2(1/T2)	a*[Fe] + b*[lipid] + c*[Fe]*[lipid] + d	0.969	10.704
R2*(1/T2*)	a*[Fe] + b*[lipid] + c*[Fe]*[lipid] + d	0.97	0.001
MT	a*[Fe]*iron type + b*[lipid]*lipid type + c	0.727	0.0001
MTV	a*[Fe] + b*[lipid] + c	0.916	0.001

Discussion

In this research, the goal was to find the best fit model that allows predicting R1 values based on lipid and iron types and their concentrations. One positive outcome that can be derived from this experiment, is the possibility of inferring diseases and defects in the human brain, depending on the results of R1. Further studies can deduct from this study the relationship between R1 and iron and lipid concentrations, as well as iron and lipid types. Given that relationship, detecting abnormal values of iron and lipid concentrations, can help in diagnostics neurological symptoms and diseases.

In addition to the findings in the literature (Rooney et al., 2007), in this experiment, it was found that the types of iron and lipids also influence R1 values prediction, and not just their concentrations. As seen in Fig. 8 B, the iron types have more effect than the lipid type, due to their higher coefficient (Cornell & Berger, 1987).

There are some limitations to this research. Out of 10 experiments that have been used in this research, not all the combinations of the lipid types and iron types have been tested. The combinations of Ferritin with PC_Cholest, and Transferrin with PC, and PC_Cholest were not included in the phantom system. Due to that fact, it is possible that the results supporting the larger meaning of the iron type, won't be demonstrated in future similar experiments. In other words, the fact that there is much more variance due

to iron vs. lipid makes the comparison between them harder. Second, the concentrations presented in this research are not correlated exactly to the human brain iron and lipid concentrations (Yeatman et al., 2014). Due to that fact, it would not be accurate to draw conclusions about the human brain based on the results of this study.

Further experiments can aim to model the relation between iron and lipid to other qMRI parameters. With the full models connecting iron and lipids to qMRI parameters, it will be possible to turn the model and find the right equation for predicting iron and lipid in the human brain. These results can lead to innovation in the world of science and medicine.

The code of the research can be found in the following address: https://github.com/ShirlyEliezer/iron_lipids. In order to download the code to your computer, clone to the desire folder with the command: git clone https://github.com/ShirlyEliezer/iron_lipids

References

1. Callaghan, M. F. et al. Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. Neurobiol. Aging 35, 1862–1872 (2014).
2. "Correlation Coefficient: Simple Definition, Formula, Easy Steps". Statistics How To, 2021.
3. Yeatman, J. D., Wandell, B. A. & Mezer, A. A. Lifespan maturation and degeneration of human brain white matter. Nat. Commun. 5, 4932 (2014).
4. Pearson, Karl (20 June 1895). "Notes on regression and inheritance in the case of two parents". Proceedings of the Royal Society of London. 58: 240–242
5. Cox, S. R. et al. Ageing and brain white matter structure in 3,513 UK Biobank participants. Nat. Commun. 7, 13629 (2016).
6. David A. Freedman *(2009). Statistical Models: Theory and Practice*. Cambridge University Press. *p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression e* right hand side, each with its own slope coefficient
7. Lorio, S. et al. Disentangling in vivo the effects of iron content and atrophy on the ageing human brain. Neuroimage 103, 280–289 (2014).
8. *Renaud, O., & Victoria-Feser, M. P. (2010). A robust coefficient of determination for regression. Journal of Statistical Planning and Inference, 140(7), 1852-1862.*
9. Gracien, R.-M. et al. Evaluation of brain ageing: a quantitative longitudinal MRI study over 7 years. Eur. Radiol. 27, 1568–1576 (2017).
10. *Willmott, Cort J.; Matsuura, Kenji (December 19, 2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance". Climate Research. 30: 79–82. doi:10.3354/cr030079.*
11. Draganski, B. et al. Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). Neuroimage 55, 1423–1434 (2011).
12. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, 2nd edition, Springer, New York/Berlin/Heidelberg, 2008.
13. Tardif, C. L. et al. Investigation of the confounding effects of vasculature and metabolism on computational anatomy studies. Neuroimage 149, 233–243(2017).
14. R. Duda, P. Hart, D. Stork, Pattern Classification, John Wiley & Sons, 2001.
15. Carey, D. et al. Quantitative MRI provides markers of intra-, inter-regional, and age-related differences in young adult cortical microstructure. Neuroimage 182, 429–440 (2017).
16. D. Berrar, W. Dubitzky, Overfitting, in: W. Dubitzky, O. Wolkenhauer, K.-H. Cho, H. Yokota (Eds.), Encyclopedia of Systems Biology, Springer, 2013, pp. 1617–1619.
17. Bot, J., Blezer, E., Kamphorst,W., Nijeholt, G., Ader, H.J., Castelijns, J.A., Nicolay, K., Bergers, E., Ravid, R., Polman, C., Barkhof, F., 2004. The spinal cord in multiple sclerosis: relationship of high-spatial-resolution quantitative MR imaging findings to histopathologic results. Radiology 233 (2), 531–540
18. Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. Malawi medical journal, 24(3), 69-71.
19. Misra, S., Osogba, O., & Powers, M. (2019). Unsupervised outlier detection techniques for well logs and geophysical data. Machine Learning for Subsurface Characterization, 1.
20. Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In Noise reduction in speech processing (pp. 1-4). Springer, Berlin, Heidelberg.
21. Tabachnick, B. G., Fidell, L. S. (2013) Using multivariate statistics, 6th ed. Boston, MA: Pearson.
22. Hamilton, D. F., Ghert, M., & Simpson, A. H. R. W. (2015). Interpreting regression models in clinical outcome studies.
23. Field, A. P., Miles, J. (2010) Discovering statistics using SAS: (and sex and drugs and rock n roll), Thousand Oaks, CA: SAGE.
24. Cornell, J. A., & Berger, R. D. (1987). Factors that influence the value of the coefficient of determination in simple linear and nonlinear regression models. Phytopathology, 77(1), 63-70.
25. Tukey, J. W. (1977) Exploratory data analysis, Reading, MA: Addison-Wesley.
26. Yeatman, J. D. et al. Lifespan maturation and degeneration of human brain white matter. Nat. Commun. 5:4932 doi: 10.1038/ncomms5932 (2014).
27. Shtangel O, Mezer AA. A phantom system for assessing the effects of membrane lipids on water proton relaxation. NMR in Biomedicine. 2020; 33: e4209. https://doi.org/10.1002/nbm.4209
28. Mottershead, J.P., Schmierer, K., Clemence, M., Thornton, J.S., Scaravilli, F., Barker, G.J.,Tofts, P.S., Newcombe, J., Cuzner, M.L., Ordidge, R.J., McDonald, W.I., Miller, D.H.,2003. High field MRI correlates of myelin content and axonal density in multiple sclerosis — a post-mortem study of the spinal cord. J. Neurol. 250 (11), 1293–1301
29. Moffat BA, Chenevert TL, Lawrence TS, Meyer CR, Johnson TD, Dong Q, Tsien C, Mukherji S, Quint DJ, Gebarski SS, Robertson PL, Junck LR, Rehemtulla A, Ross BD Proc Natl Acad Sci U S A. 2005 Apr 12; 102(15): 5524-9.
30. Hamstra DA, Chenevert TL, Moffat BA, Johnson TD, Meyer CR, Mukherji SK, Quint DJ, Gebarski SS, Fan X, Tsien CI, Lawrence TS, Junck L, Rehemtulla A, Ross BD Proc Natl Acad Sci U S A. 2005 Nov 15; 102(46): 16759-64.
31. Schmierer, K., Scaravilli, F., Altmann, D.R., Barker, G.J., Miller, D.H., 2004. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. Ann. Neurol. 56(3), 407–415.
32. Stüber C, Morawski M, Schäfer A, et al. Myelin and iron concentration in the human brain: a quantitative study of MRI contrast. Neuroimage. 2014; 93 (P1): 95-106. https://doi.org/10.1016/j.neuroimage.2014.02.026
33. Samorajski T, Rolsten C. Age and regional differences in the chemical composition of brains of mice, monkeys and humans. Prog Brain Res. 1973; 40(C): 253-265. https://doi.org/10.1016/S0079-6123(08)60692-5