

Can MRI measure myelin? Systematic review, qualitative assessment, and meta-analysis of studies validating microstructural imaging with myelin histology

Alberto Lazari^a, Ilona Lipp^{b,*}

^a Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, UK

^b Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany



ARTICLE INFO

Keywords:

Myelin
MRI
Validation
Microstructural imaging
Relaxometry
Magnetization transfer
Diffusion
Histology

ABSTRACT

Recent years have seen an increased understanding of the importance of myelination in healthy brain function and neuropsychiatric diseases. Non-invasive microstructural magnetic resonance imaging (MRI) holds the potential to expand and translate these insights to basic and clinical human research, but the sensitivity and specificity of different MR markers to myelination is a subject of debate. To consolidate current knowledge on the topic, we perform a systematic review and meta-analysis of studies that validate microstructural imaging by combining it with myelin histology. We find meta-analytic evidence for correlations between various myelin histology metrics and markers from different MRI modalities, including fractional anisotropy, radial diffusivity, macromolecular pool, magnetization transfer ratio, susceptibility and longitudinal relaxation rate, but not mean diffusivity. Meta-analytic correlation effect sizes range widely, between $R^2 = 0.26$ and $R^2 = 0.82$. However, formal comparisons between MRI-based myelin markers are limited by methodological variability, inconsistent reporting and potential for publication bias, thus preventing the establishment of a single most sensitive strategy to measure myelin with MRI. To facilitate further progress, we provide a detailed characterisation of the evaluated studies as an online resource. We also share a set of 12 recommendations for future studies validating putative MR-based myelin markers and deploying them *in vivo* in humans.

1. Introduction

Myelin is crucial for healthy brain function. Early studies found that myelin provides insulation and facilitates electrical conduction in neural circuits (Basser, 2004; Goldman and Albus, 1968; Rushton, 1951; Waxman, 1980). More recently, a host of observations have emphasised a wider set of roles for myelination, from enabling high frequency conduction (Saab et al., 2016) to providing trophic support for axons (Fünfschilling et al., 2012; Jensen and Yong, 2016; Lee et al., 2012). Changes in myelination occur as part of normal brain development (Gibson and Peterson, 1991; Ziegler et al., 2019) and aging (Hill et al., 2018; Peters, 2009). Moreover, recent studies have highlighted the possibility that myelination may undergo subtle increases and decreases also in adulthood, in response to neuronal activity or experience (Sampaio-Baptista et al., 2013; Sinclair et al., 2017), and that these changes may be crucial for learning and memory formation (Mckenzie et al., 2014; Pan et al., 2020; Steadman et al., 2020).

When myelin is damaged or lost (demyelination), neural communication is affected. Demyelination is a hallmark of many neurolog-

ical diseases, such as multiple sclerosis, and affects white matter as well as grey matter (GM) (Lucchinetti et al., 2011). The possibility of inducing remyelination, either through pharmacological treatments (Stankoff et al., 2016) or behavioural interventions through experience-dependent myelin plasticity (Purger et al., 2016), is an emerging focus in clinical trials. It is also known that adverse events like social isolation can impact brain myelination (Liu et al., 2012; Makinodan et al., 2012) and myelination is involved in neuropsychiatric disorders including autism (Zikopoulos and Barbas, 2010) and schizophrenia (Stedehouder and Kushner, 2017). Therefore, robust *in vivo* markers to assess myelination could enable diagnosis and treatment monitoring for a wide range of conditions, as well as more detailed study of healthy brain function in humans.

MRI is the most frequently used tool to study brain structure and function *in vivo*, since it is noninvasive and widely available. However, capturing information about myelin with MRI can be challenging due to its relatively low resolution. Myelination has several microscopic features that likely influence the overall myelin content in a given MR voxel - from nodal length to internodal length to myelin thickness

* Corresponding author.

E-mail address: lippi@cbs.mpg.de (I. Lipp).

<https://doi.org/10.1016/j.neuroimage.2021.117744>

Received 25 September 2020; Received in revised form 5 January 2021; Accepted 9 January 2021

Available online 30 January 2021

1053-8119/Crown Copyright © 2021 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

(Almeida and Lyons, 2017; Zatorre et al., 2012). These features are in the range of micrometers, while the resolution of images acquired using a conventional MRI scanner lies in the millimetre range. Therefore, microstructural parameters can only be estimated as summary measures, often of total myelin content, and inferred from information obtained in comparatively large tissue volumes, e.g. cubic mm-sized voxels. This is possible because many aspects of MR physics are influenced by myelination. Macromolecules in the myelin sheath - such as lipids and proteins - influence relaxation rates (as measured by the longitudinal relaxation rate R_1 , the transverse relaxation rate R_2 , the effective transverse relaxation rate R_2^* , and myelin water fraction MWF) (MacKay et al., 2009), as well as magnetization transfer (MT; as quantified by the magnetization transfer ratio MTR, and the macromolecular pool size MP) (Grossman et al., 1994; Henkelman et al., 1993; Wolff and Balaban, 1989). The structure of the myelin sheath hinders local water diffusion (as measured by diffusion-weighted imaging (DWI) Beaulieu (2002, 2014)). Moreover, myelin has diamagnetic properties, which influence the local magnetic field strength (Möller et al., 2019; Weiskopf et al., 2015). This affects the effective transverse relaxation rate (R_2^*), and can be more directly measured by quantitative susceptibility mapping (QSM). In fact, most MR markers are likely sensitive to multiple biophysical mechanisms and consequently also to various microstructural features of the tissue - for example, relaxation rates may also be influenced by axon diameter and axon packing (Lynn et al., 2020). The biophysical mechanisms underlying myelin-sensitive MR-modalities have been reviewed extensively elsewhere (Does, 2018; Edwards et al., 2018; Möller et al., 2019; Novikov et al., 2019).

In recent years a growing number of MR techniques have been successfully applied to study brain microstructure, and biophysical multi-compartment models have been developed to explicitly attempt to capture and quantify myelin-specific signals (Alonso-Ortiz et al., 2015; Campbell et al., 2017; Heath et al., 2018; MacKay et al., 2009; Mezer et al., 2013; Piredda et al., 2020). Currently, it is not clear which of these MR markers is the most biologically accurate non-invasive measure for myelin, and how the measures differ in their sensitivity to specific aspects of myelination. While theoretical modelling can help in designing novel MR methods (Veraart et al., 2019), any model of MR signals will have intrinsic limitations given the complexity of brain tissue, including simplifying assumptions about the tissue geometry and the MR physics. For example, magnetization exchange between the modelled compartments is often assumed to be much slower than the MR measurements (Barta et al., 2015; Does, 2018; Levesque and Pike, 2009) and myelin water is often assumed to be the only driver of fast decay (Cohen-Adad, 2014). In light of these limitations, validation is a necessary step towards making MRI-based methods biologically interpretable, and applicable to clinical and basic research (Barros et al., 2019; Cohen-Adad, 2018).

Histological validation studies have been conducted for a variety of microstructural MRI metrics. Validity is often considered as the extent of agreement between a measured parameter and an underlying biological parameter of interest. For this reason, the gold standard for validation studies aiming to assess the accuracy of MR markers is to compare MRI and underlying myelin content within the same tissue. These studies perform MRI scanning in animals or in humans post-mortem, process the tissue histologically to obtain a ground-truth measure of myelination in the tissue, and then test whether variance in the microstructural MRI metric is driven by variance in myelination as assessed with histology. Neuroimaging studies conducted *in vivo* often use individual validation studies of this kind as evidence to justify employing a specific microstructural MRI measure to study myelination. However, validation studies use varying methodologies and have varying outcomes, making it difficult to assess to what extent such justifications are valid.

To better understand which MRI marker is best suited to measure myelin, we aim to collate evidence from the validation literature on microstructural MRI-metrics for myelin. First, we provide a comprehensive overview of validation studies reporting a correlation between MRI and

histology. Second, we assess qualitatively a range of key methodological details known to influence histological signals (e.g. tissue processing), MR signals (e.g. state of the tissue during scanning) and correlation between the two (e.g. ROI definition method). Third, we perform meta-analyses to investigate how much variance is shared between each microstructural MR marker and histological myelin metrics. Fourth, we use insights from our systematic review to highlight the limitations of existing validation work and to develop a list of recommendations for future validation and *in vivo* imaging experiments.

2. Methods

2.1. Systematic review: study selection

A systematic review was conducted using PRISMA guidelines (Shamseer et al., 2015) and incorporating best practices from the AMSTAR 2 checklist for clinical meta-analyses where applicable (e.g. searching across multiple datasets, performing study selection in duplicate) (Shea et al., 2017). Articles on quantitative validation of MR markers for myelin were searched for in Pubmed and Scopus (search date: February 20, 2020), using the following search terms: ((myelin[Title/Abstract]) AND (post-mortem[Title/Abstract] OR postmortem[Title/Abstract] OR histol*[Title/Abstract] OR *ex vivo*[Title/Abstract] OR histochem*[Title/Abstract] OR histopath*[Title/Abstract])) AND (MR*[Title/Abstract] OR magnetic resonance imaging[Title/Abstract]) AND (brain[Title/Abstract] OR spinal cord[Title/Abstract]). In addition, the literature was complemented with 45 articles from the authors literature library, which included references from recent reviews (e.g. Heath et al., 2018; Mancini et al., 2020). These articles were subject to the same full-text screening procedure as the database-derived articles.

We included only peer-reviewed articles. We excluded review articles, phantom-only studies, atlas-based validation studies, MR spectroscopy studies, studies that do not quantify both myelin and MRI, and studies that do not perform quantitative comparisons of myelin and MRI. Additionally, we excluded papers that only used the intensity of single-contrast weighted MRI images, such as T1w - as opposed to calculating metrics by combining several weighted images from multiple contrasts - for correlation with histology. We did not specify exclusion criteria based on the investigated species or pathology or myelin validation method used.

All studies detected in the systematic search were screened for inclusion criteria in their abstracts. As most studies were excluded at this stage, abstract screening was independently run by two investigators (I.L. and A.L.) and discordant decisions on article inclusions were resolved by discussion and consensus-finding between the investigators. To verify the inclusion criteria based on the full-text of each article, a second stage of screening was also performed.

2.2. Systematic review: information extraction

To provide a comprehensive overview of the state of the validation literature for myelin, from each selected paper, we extracted information on the methodology of the validation aspect of the study. We chose methodological factors that may impact on the MRI-based metrics, the histological quantification of myelin or their correlation (also see Barros et al., 2019 on a similar review on iron imaging). The information is reported in Supplementary Tables 1–5 and is also openly available for external use at: <https://lazari.github.io/Myelin-Validation-Systematic-Review/>. Below, we describe which information was selected and how it was reported.

2.2.1. Imaging: modality, parameters and metrics

Studies used one of five imaging modalities: a) **Relaxometry**: acquisitions aimed at estimating relaxation times or relaxation rates (as relaxation rate = $1/\text{relaxation time}$) or myelin water fraction based

on the short relaxation rate of water trapped between myelin sheaths (Mackay et al., 1994), independently of the specific sequence used for such acquisition. Within Relaxometry, the metric MWF refers to the fraction of myelin water estimated through multi-component modelling. b) **DWI**: metrics based on diffusion models (these metrics are known to be influenced by both myelin and non-myelin factors, but are often used as microstructural markers (Beaulieu, 2014; De Santis et al., 2014; Pierpaoli et al., 1996) c) **MTI**: acquisitions applying off-resonance pulses to induce magnetization transfer effects and aiming to quantify them, such as through the measures Macromolecular Proton Fraction (MPF), Magnetization Transfer Ratio (MTR) or f (ratio of the macromolecular water pool over the free water pool); d) **QSM**: metrics based on the quantification of susceptibility based imaging; and e) **Others**: acquisition that do not fall into categories a-d, such as the T1w/T2w ratio (Glasser and Van Essen, 2011).

The quantitative MT models that aim to assess the proportion of macromolecules (assumed to be mostly macromolecules in myelin) in a voxel all use different terms for this parameter. While f denotes the ratio of the macromolecular water pool over the free water pool, the fraction (ratio of the macromolecular water pool divided by the sum of all pools) is denoted in a number of ways. We summarize this metric as macromolecular pool (MP) and indicate in brackets the original term used in the table.

In addition to the metrics quantified, we considered the field strength at which the imaging was done and the resolution of the acquisition protocol. If voxel resolution was not reported, it was deduced from field of view and matrix size. We report all resolutions in mm for comparability across studies, rounded to three significant figures.

2.2.2. Tissue types

We extracted information on what species was used and whether it came from healthy individuals or natural or induced pathologies. We also reported which part of the central nervous system was imaged (e.g. brain vs spinal cord, whole brain vs sections or smaller samples), as validity of MR markers may be varying depending on which area of the nervous system is being considered (fractional anisotropy, for instance, is likely to be more valid in tissue where fiber orientation is more coherent, as it is in white matter).

We additionally extracted information on what anatomical structures or tissue types were used for the statistical analysis specifically. Depending on the level of detail provided by the paper, this was either anatomical regions of interest (ROI) or tissue types (e.g. identified by degree of pathology). We also extracted whether manual or automatic delineation of ROIs was performed. We classified the ROI definition as manual to indicate this with or defined and labeled or designated were used without further mentioning of an automated or semi-automated tool. If atlases were used to segment ROIs, the specific atlas used is reported. If voxelwise analysis statistics were computed, the ROI definition was set to NA (not applicable).

2.2.3. Tissue preparation

As stage and type of tissue fixation both affect MR parameters (Dusek et al., 2019), we extracted information on the state of the tissue during MR imaging. We distinguished between *in vivo*, *in situ*, fresh, fixed. If scanning was not done *in vivo*, we also looked at post-mortem times (defined as time from death to fixation), which are indicative of the autolytic state of the tissue at the time of fixation. Post-mortem time was reported as NA (not applicable) when scanning was done *in vivo* and/or in perfusion fixed tissue. If it is reported with \pm sign, then this indicates mean and standard deviation across samples.

As temperature has an effect on MR relaxation times (Birkel et al., 2016) and diffusion (Dhital et al., 2016), tissue temperature during scanning was extracted. If scanning was performed *in vivo*, body temperature was assumed. Last but not least, we report how tissue was treated for histology (cryosectioning or paraffin embedding), which can affect tissue staining success and intensity (Werner et al., 2000).

2.2.4. Histology methods

Given the diversity of potential histology methods, we extracted information on how myelin was histologically visualised, including staining methods.

Some microstructural features of the brain, such as iron and axonal density, are related to myelin and have similar effects as myelin on MR signals (Möller et al., 2019). We assessed whether iron or axons have also been considered by the studies we reviewed, and report this information in the form of binary columns in the table). This was determined by searching the text for any mention of histological methods that could be used to estimate iron or axons (e.g. iron stains, anti-neurofilament immunohistochemistry or electron microscopy).

We extracted the histology quantification method, which belonged to one of the following categories: a) **Staining fraction**: this category was chosen if the microscopy image was segmented and the fraction / percentage of area stained was quantified, b) **Staining intensity**: this category was chosen if the optical density of one or more colour channels was quantified, c) **Inverse staining intensity**: similar to staining intensity, but instead light transmittance was quantified, leading to a different expected direction of the correlation coefficient between MRI and histology.

We also report the thickness of the sections used for histology (all reported in μm , converted if necessary for the sake of comparison across studies).

2.2.5. Statistics

An inclusion criterion for our screening was that quantitative correlation between MRI and histology must have been performed. Therefore, we extracted the specific type of correlation that was used in each paper - whether average values within a given ROIs were compared between histology and imaging, or whether pixel-wise or voxel-wise correlations were employed.

We then considered what design was used for the main correlation analysis. Validation studies can either focus on whether for the same subject differences in MR metrics across ROIs are reflected in histological metrics (within-subject validation design, also known as spatial correlation) or whether for the same ROI, differences in MR metrics between subjects are reflected in histological metrics (between-subject validation design). In these two types of designs, variability is driven by different factors: in within-subject designs, variability is driven by anatomical and tissue differences in myelin content; whereas in between-subject designs, variability is driven by external factors (such as pathology, age or genotype). Therefore, we distinguished between a) **Between-subject**: correlations where each data point originates from a separate sample/subject; b) **Within-subject**: correlations where all data points originate from the same sample/subject (see Figure 1 for a schematic illustration of the difference between between-subject and within-subject design); c) **Mixed (modelled)** - designs where all ROIs across all subjects were included in the same analysis, taking into account that multiple data points were derived from the same subject. By formally correcting for the between-subject or within-subject variance in the data, these studies effectively report either within-subject or between-subject correlation coefficients. d) **Mixed (not modelled)** - all ROIs across all subjects included in the same analysis, without taking into account that multiple data points were derived from the same subject. In these studies, it is impossible to disentangle whether the correlation coefficient is driven by within- or between-subject variance.

We extracted the sample size used for the correlation analysis. We reported sample size per group (number of subjects in each group, e.g. in control and in disease group), as well as total subjects (total subjects across groups). For studies using multiple ROIs, the effective sample size for the correlation differed from the number of subjects used. Therefore, we also reported the number of ROIs used. In Within-subject studies with varying numbers of ROIs for each individual slice/sample, the range of ROIs for each slice/sample was reported. In Mixed (unmodelled) statis-

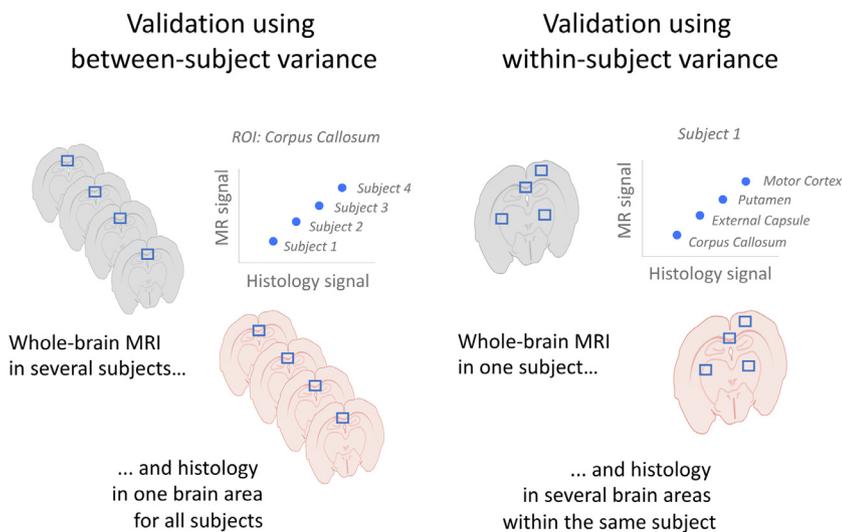


Fig. 1. Mock examples of between-subject vs within-subject design in validation studies. In between-subject validation studies (left), each data point comes from a different subject and correlations are computed separately for each ROI. In within-subject validation studies (right), each data point comes from a different brain region and correlations are computed separately for each subject. In mixed designs, data points from different subjects and different regions are pooled into one data analysis.

tical approaches, often only the total number of ROIs across all subjects is reported in the article, and we reported this figure instead.

The matching of histology and MRI was also considered. No co-registration indicates that ROIs were specified in native space in histology and MRI. NA (not applicable) was used if registration is not feasible, such as for electron microscopy, where the field of view for microscopy is generally only a small fraction of the field of view for MRI. If MRI and histology data were coregistered, we extracted the type of registration (e.g. manual, affine transform, etc.)

For the statistical results, we reported statistical methodology (Pearson correlation, Spearman correlation or linear regression), correlation effect size, significance of the correlation, and regression slope (if applicable, otherwise NA). Statistical methodologies including mixed effect models and multilevel models were all labelled as linear regression. Reported effect sizes included Pearson or Spearman correlation coefficient (directional) or a coefficient of determination R^2 (non-directional). If regression was run but no linear equation details were reported, we indicated this with not reported in the table.

2.3. Meta-analysis: study selection

Meta-analyses aim to estimate effect sizes across multiple studies. However, only comparable metrics and statistical designs can be pooled together. For example, different meta-analyses need to be run for papers reporting Spearman and Pearson correlation coefficients, which can differ in value even when applied to the same data, since Spearman correlation coefficients measure monotonicity and Pearson correlations measure linearity (de Winter et al., 2016). Moreover, studies using within-subject and between-subject designs need to be considered in different analyses. Therefore, we excluded from our meta-analysis studies where correlation coefficient type could not be determined, and studies using a mixed design where either within- and between-subject variance could have determined the outcome. For the remaining studies, we performed meta-analyses for within-subject and between-subject designs separately.

As meta-analyses aim to combine information from multiple studies and subjects, meta-analyses were carried out only for MRI markers that were validated in more than two independent studies. Moreover, only studies with more than three subjects were included, as this was necessary to calculate a standard error for each study's effect size (Schulze, 2005; Schwarzer et al., 2007). For studies with between-subject designs using separate ROIs or separate groups of subjects, the correlation coefficients were averaged (using Fishers r -to- z transformation) before entering the meta-analysis. If correlations with separate histological markers

were reported as results, the results were fed separately into the meta-analysis.

2.4. Meta-analyses: statistical analysis

All meta-analyses were run through the *meta* R package (Schwarzer et al., 2007). The coefficient of determination R^2 was used as a common effect size measure and a fixed-effect model was employed to weigh different studies based on their sample size (Schulze, 2005). For each MR modality, a forest plot was used to summarise studies included, study sample sizes, and effect sizes of individual studies as well as meta-analytic effect size (with 95% confidence intervals). A confidence interval that had a range of exclusively positive values was considered indicative of evidence for a correlation between MRI and histology.

2.5. Meta-analyses: p -value distribution in the literature

Questionable Research Practices, whether performed consciously or unconsciously, can lead to skewed effect sizes in the literature. For instance, as significant studies are more likely to be published, there can be both conscious and unconscious biases leading researchers to report biased p -values (Button et al., 2013). The effects of this pressure to find significance are reflected in the literature: although p -values would be expected to appear randomly in the literature, p -values right under 0.05 are over-represented in the psychology literature (Head et al., 2015). To test whether the validation literature in our meta-analyses is biased, we perform a p -curve analysis through the *dmetar* R package (Harrer et al., 2019; Simonsohn et al., 2014).

2.6. Meta-analyses: quantifying and correcting for publication bias in the literature

Funnel plots are a common tool to estimate publication bias in meta-analyses (Egger et al., 1997). For any given effect, a meta-analysis calculates, one would expect high-sample-size studies to best approximate that effect (visualised by the peak of the funnel). In contrast, low-sample-size studies are expected to provide noisier estimates of the true underlying effect size, leading to more variance in the effect size estimations and thus forming the wide part of the funnel as effect sizes decrease.

While funnel plots can capture the presence of publication bias, they are also sensitive to other effects. For example, an asymmetric funnel plot can arise from poor methodology in small-sample-size studies, heterogeneous true effect sizes across studies, or appear by chance

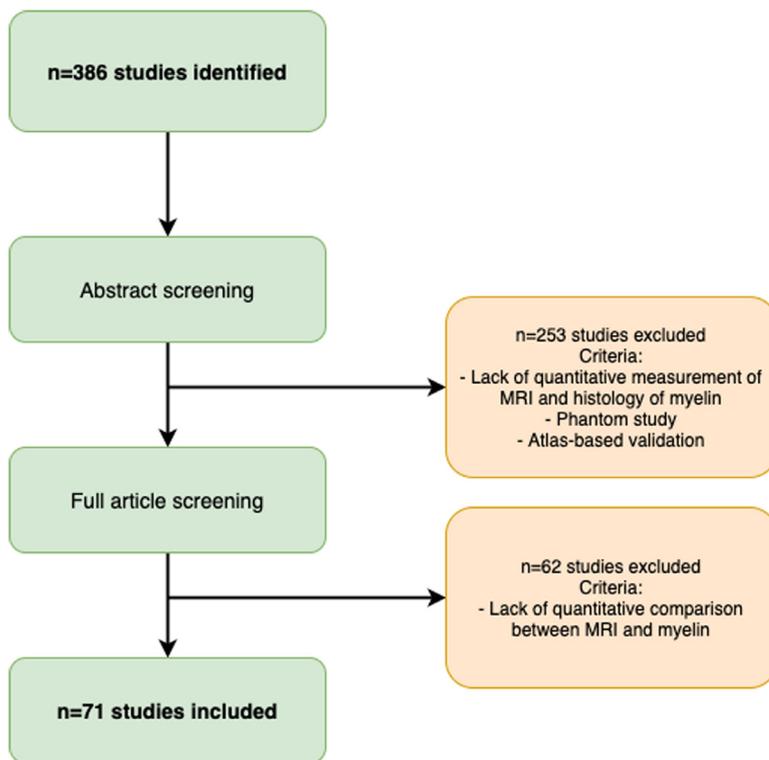


Fig. 2. Flow diagram of study selection.

(Sterne et al., 2011). However, funnel plots have the advantage that by quantifying potential publication bias, they allow correcting for it. To demonstrate the robustness of our results to potential publication bias, we perform trim-and-fill procedures based on Duval and Tweedie (2000) and implemented through the *trimfill* function in the *dmr* R package.

Finally, we also complement funnel plots with Spearman correlations between sample sizes and effect sizes across the studies included in the meta-analysis.

2.7. Data and code availability

All extracted information and quantitative meta-analysis code is openly available at <https://lazaral.github.io/Myelin-Validation-Systematic-Review/>.

3. Results

3.1. Exploring the features of the MR-histology validation literature

Our search yielded 386 unique articles (Fig. 2). Of these articles, 294 were from Pubmed, 46 from Scopus, 45 from our expert library. One additional study was included based on the recommendations of one of the reviewers of this manuscript. Abstract screening led to exclusion of 253 studies, while full-text screening further excluded 62 studies, with a total of 71 remaining studies fitting our inclusion criteria (Table 1). We believe this constitutes the state of the field at the time of writing.

Information from the included studies was extracted and summarised in Supplementary Tables 1–5. To have a better overview of the literature, we provide quantitative summaries of our findings on the literature content (Fig. 3). We find that validation studies have a median sample size of 13 (Fig. 3A), comparable to the median of the most cited fMRI studies during the period of publication included in our meta-analysis (12), but below current median sample size (20) (Szucs and Ioannidis, 2020). The validation literature uses a wide range of histological markers (with 9 studies using more than one marker, Fig. 3F). It also

employs a wide variety of species and field strengths used for experiments (Fig. 3D). Taken together, these factors highlight that the validation literature has considered different field strengths, species, and MRI and histology approaches to quantify myelin, making common results across papers highly generalizable.

The studies we included test the validity of a wide-range of potential myelin markers, reflecting the range of markers used in application studies. For instance, the most commonly validated MR markers are DWI-based (such as FA), with relaxometry and MT-based studies close seconds (Fig. 3F). We also note that many markers appear only once in the literature (more specifically: Orientation Dispersion Index, Neurite Density Index, T1w/T2w, Relaxation Along a Fictitious Field in the rotating frame of rank 4, MTR from inhomogeneous MT, diffusion basis spectrum imaging - based radial diffusivity, relative semi-solid proton fraction, diffusion standard deviation index), suggesting there are many more markers in the literature that lack thorough validation. We do not include these markers in Fig. 3F for simplicity.

Finally, we observed that many studies fail to report some crucial aspect of their methodologies or results. We therefore reported the most commonly missing experimental details (Fig. 3C).

3.2. Meta-analysis

A total of 32 studies were selected for inclusion in a quantitative meta-analysis based on this selection process (24 reporting Pearson only, 7 reporting Spearman only, and 1 reporting both; 22 using between-subject design, 10 using within-subject design). The selected studies are indicated in Table 1 (last column). To ensure comparability between studies included, separate meta-analyses were run for studies using between-subject and within-subject variance (Figs. 4 and 5 respectively). Moreover, studies using Pearson and Spearman coefficients were also pooled separately (meta-analyses are labelled Pearson or Spearman within Fig. 4). The metrics for which the specified minimum number of studies passed our inclusion criteria for the meta-analyses were: FA, RD, MD, susceptibility, MTR and MP for between-subject meta-analyses, and R1, MP and MTR for within-subject meta-analyses. On the included stud-

Table 1

Overview of the assessed validation studies. Acronyms: **DWI:** Diffusion-weighted imaging; **AK:** axial kurtosis; **AI:** Anisotropy Index (tADC/lADC); **DBSI-RD:** diffusion basis spectrum imaging - based radial diffusivity; **DK:** diffusion kurtosis metrics; **DT:** diffusion tensor metrics; **FA:** fractional anisotropy (from diffusion tensor model); **IADC:** longitudinal apparent diffusion coefficient (not modelled with tensor); **MD:** mean diffusivity (from diffusion tensor model); **MK:** mean kurtosis; **RD:** radial / transverse diffusivity (from diffusion tensor model); **RK:** radial kurtosis; **SDI:** diffusion standard deviation index; **tADC:** transverse apparent diffusion coefficient (not modelled with tensor); **Relaxometry:** **MWF:** myelin water fraction; **R1:** longitudinal relaxation rate; **R2*:** effective transverse relaxation rate; **RAFF4:** Relaxation Along a Fictitious Field in the rotating frame of rank 4; **T1:** longitudinal relaxation time; **T2:** transverse relaxation time; **T2*:** effective transverse relaxation time; **MT:** magnetisation transfer; **BPF:** bound pool fraction; **F:** pool size ratio; **Fb:** macromolecular proton fraction; **ih-MTR:** MTR from inhomogeneous MT; **MOb:** fraction of magnetization that resides in the semi-solid pool and undergoes MT exchange; **MP:** macromolecular pool; **MPF:** macromolecular proton fraction; **MTR:** magnetisation transfer ratio; **PSR:** Macromolecular-to-free-water pool-size-ratio; **STE-MT:** MTR based on short echo time imaging; **T1sat:** T1 of saturated pool; **UTE-MTR:** MTR based on ultrashort echo time imaging; **QSM:** quantitative susceptibility mapping; **Others:** **rSPF:** relative semi-solid proton fraction from a 3D ultrashort echo time (UTE) sequence within an appropriate water suppression condition; **T1w/T2w:** ratio of image intensity in a T1-weighted vs T2-weighted acquisition. **AMG:** Autometallographic myelin stain; **LFB:** Luxol fast blue stain; **MA:** fraction of myelinated axons; **MAB238:** Anti-oligodendrocyte immunohistochemistry; **MBP:** Anti-myelin-basic-protein immunohistochemistry; **PAS:** periodic acid-Schiff; **PIXE:** proton-induced X-ray emission; **PLP:** Anti-proteolipid-protein immunohistochemistry.

Reference	MRI modality	MRI metrics	Species	Tissue state	Myelin histology	Meta-analysis
Continued on next page						
Abe et al. (2019)	DWI	AD, FA, RD	Mouse	perfusion fixed	PLP	-
Aojula et al. (2016)	DWI	AD, FA, MD, RD	Rat	in vivo	MBP	+
Argyridis et al. (2014)	DWI	MD	Mouse	perfusion fixed	LFB	+
Chandran et al. (2012)	DWI	FA, RD	Mouse	in vivo	LFB, MBP	+
Chang et al. (2017b)	DWI	FA, AD, RD	Mouse	perfusion fixed	MBP	-
Chang et al. (2017a)	DWI	FA, AD, RD, MD	Mouse	perfusion fixed	MBP	-
Choi et al. (2015)	DWI	FA, RD	Dog	fixed	Gold chloride	-
Grussu et al. (2017)	DWI	AD, FA, MD, RD	Human	fixed	PLP	-
Janve et al. (2013)	DWI	AD, FA, RD	Rat	perfusion fixed	LFB	+
Jelescu et al. (2016)	DWI	RD, RK	Mouse	in vivo	EM	-
Jespersen et al. (2010)	DWI	FA	Rat	perfusion fixed	AMG	-
Jito et al. (2008)	DWI	FA	Rats	in vivo	Toluidine blue	+
Kelm et al. (2016)	DWI	FA, MD, RD, MK, AK, RK	Mouse	perfusion fixed	EM	-
Kozłowski et al. (2008)	DWI	FA, AD, MD, RD	Rat	perfusion fixed	LFB, MBP	-
Kozłowski et al. (2014)	DWI	FA	Rats	fixed	MBP, Eriochrome-cyanine	-
Lehto et al. (2017)	DWI	AD, FA, MD, RD	Rat	in vivo	Gold chloride	+
Martirosyan et al. (2016)	DWI	FA	Rat	perfusion fixed	LFB	-
Moll et al. (2011)	DWI	AD, FA, MD, RD	Human	in situ / fixed	MBP	-
Mollink et al. (2019)	DWI	FA, MD, RD, AD	Human	fixed	PLP	+
Mottershead et al. (2003)	DWI	ADC, SDI	Human	fresh	LFB	-
Oakden et al. (2015)	DWI	AD, RD	Rat	in vivo / fixed	LFB	-
Peters et al. (2019)	DWI	FA, MD	Human	in vivo	LFB	-
Pol et al. (2019)	DWI	FA, MD	Mouse	in vivo	Solochrome	+
Praet et al. (2018)	DWI	AD, AK, DT, DK, FA, MD, MK, RD, RK	Mouse	in vivo	MBP	+
Schmierer et al. (2007b)	DWI	FA, MD	Human	fresh	LFB	-
Schmierer et al. (2008)	DWI	AD, FA, MD, RD	Human	fresh / fixed	LFB	-
Schwartz et al. (2005)	DWI	AI, iADC, tADC	Rat	perfusion fixed	Toluidine blue	-
Seehaus et al. (2015)	DWI	FA	Human	fixed	Gallyas	-
Seewann et al. (2009)	DWI	FA, ADC	Human	fixed	LFB, PAS, PLP	-
Soustelle et al. (2019)	DWI	RD	Mouse	perfusion fixed	MBP	+
Sundberg et al. (2010)	DWI	AD, FA, RD	Mouse	in vivo	MAB328	-
Takagi et al. (2009)	DWI	FA	Rat	excised	EM	-
Thiessen et al. (2013)	DWI	AD, FA, RD	Mouse	perfusion fixed	EM	+
Tu et al. (2016)	DWI	AD, FA, MD, RD, MTR at different ppm	Rat	in vivo	MBP	-
van der Voorn et al. (2011)	DWI	ADC, FA	Human	fixed	LFB	-
van Tilborg et al. (2018)	DWI	FA, RD	Rat	perfusion fixed	MBP	+
Wang et al. (2009)	DWI	AD, FA, RD, "Trace"	Rat	in vivo	LFB	-
Wang et al. (2015)	DWI	DBSI-RD	Human	fixed	LFB	-
Wei et al. (2013)	DWI	FA, RD	Dog	perfusion fixed	Gold chloride	-
Yano et al. (2018)	DWI	FA, MD, RD	Mouse	perfusion fixed	PLP	+
Beckmann et al. (2018)	MT	MTR	Mouse	in vivo	LFB	+
Bot et al. (2004)	MT	MTR	Human	fixed	LFB	-
Duhamel et al. (2019)	MT	ih-MTR	Mouse	in vivo	PLP-GFP fluorescence	-
Fatemi et al. (2011)	MT	MTR	Mouse	in vivo	LFB, MBP	+
Fjær et al. (2013)	MT	MTR	Mouse	in vivo	PLP	-
Fjær et al. (2015)	MT	MTR	Mouse	in vivo	LFB, PLP	+
Gareau et al. (2000)	MT	MTR	Guinea pig	in vivo	Solochrome-R-cyanine	+
Hakkarainen et al. (2016)	MT	MTR	Rats	perfusion fixed	Gold chloride	-
Janve et al. (2013)	MT	MP ('PSR')	Rat	perfusion fixed	LFB	-
Jelescu et al. (2016)	MT	MTR	Mouse	in vivo	EM	+
Kelm et al. (2016)	MT	MP ('PSR')	Mouse	perfusion fixed	EM	-
Khodanovich et al. (2017)	MT	MP ('MPF')	Mouse	fixed	LFB	+
Khodanovich et al. (2019)	MT	MP ('MPF')	Mouse	fixed	MBP	-
Lehto et al. (2017)	MT	T1sat, MTR	Rat	in vivo	Gold chloride	-
Moll et al. (2011)	MT	MTR	Human	in situ / fixed	MBP	-
Mottershead et al. (2003)	MT	MTR	Human	fresh	LFB	-
Odrobina et al. (2005)	MT	MTR, MP('MOb')	Rat	fixed	Toluidine blue	-
Reeves et al. (2015)	MT	MTR	Human	fixed	MBP	-
Schmierer et al. (2010)	MT	MTR	Human	fixed	MBP	-
Schmierer et al. (2004)	MT	MTR	Human	fresh	LFB	+
Schmierer et al. (2007a)	MT	MTR, MP('fb')	Human	fresh	LFB	+

(continued on next page)

Table 1 (continued)

Reference	MRI modality	MRI metrics	Species	Tissue state	Myelin histology	Meta-analysis
Schmierer et al. (2008)	MT	MTR, MP('fb')	Human	fresh / fixed	LFB	-
Seewann et al. (2009)	MT	MTR	Human	fixed	LFB, PAS, PLP	-
Soustelle et al. (2019)	MT	MP('f')	Mouse	perfusion fixed	MBP	+
Thiessen et al. (2013)	MT	MP('f'), MTR	Mouse	perfusion fixed	EM	+
Turati et al. (2015)	MT	MP('f')	Mouse	in vivo	Black Gold II, MBP	+
Underhill et al. (2011)	MT	MP('BPF'), MTR	Rat	in vivo	LFB	+
van der Voorn et al. (2011)	MT	MTR	Human	fixed	LFB	-
West et al. (2018)	MT	MP ('BPF')	Mouse	perfusion fixed	EM	-
Righart et al. (2017)	Other	T1w/T2w	Human	in situ	PLP	-
Soustelle et al. (2019)	Other	rSPF	Mouse	perfusion fixed	MBP	+
Argyridis et al. (2014)	QSM	Susceptibility	Mouse	perfusion fixed	LFB	+
Hametner et al. (2018)	QSM	Susceptibility	Human	in situ	LFB	-
Lodygensky et al. (2012)	QSM	Susceptibility	Rat	in vivo	Black Gold II	+
Pol et al. (2019)	QSM	Susceptibility	Mouse	in vivo	Solochrome	+
Soni et al. (2020)	QSM	Susceptibility	Mouse	in vivo	MBP	-
Stüber et al. (2014)	QSM	Susceptibility	Human	fixed	PIXE	-
Bagnato et al. (2018)	Relaxometry	R2*	Human	fixed	LFB and PLP	-
Bot et al. (2004)	Relaxometry	T1, T2	Human	fixed	LFB	-
Chen et al. (2017)	Relaxometry	MWF	Rat	perfusion fixed	EM	-
Hakkarainen et al. (2016)	Relaxometry	RAFFn, T1, T2	Rats	perfusion fixed	Gold chloride	-
Hametner et al. (2018)	Relaxometry	T1, R2*	Human	in situ	LFB	-
Jelescu et al. (2016)	Relaxometry	T2	Mouse	in vivo	EM	+
Kelm et al. (2016)	Relaxometry	MWF	Mouse	perfusion fixed	EM	-
Kozlowski et al. (2008)	Relaxometry	MWF	Rat	perfusion fixed	LFB, MBP	-
Kozlowski et al. (2014)	Relaxometry	MWF	Rats	in vivo	MBP, Eriochrome-cyanine	-
Laule et al. (2008)	Relaxometry	MWF	Human	fixed	LFB	-
Laule et al. (2006)	Relaxometry	MWF	Human	fixed	LFB	-
Laule et al. (2011)	Relaxometry	MWF	Human	fixed	LFB, MBP	-
Lehto et al. (2017)	Relaxometry	RAFF4	Rat	in vivo	Gold chloride	-
Mottershead et al. (2003)	Relaxometry	PD, T1, T2	Human	fresh	LFB	-
Oakden et al. (2015)	Relaxometry	MWF	Rat	in vivo / fixed	LFB	-
Odrobina et al. (2005)	Relaxometry	MWF, T1, T2	Rat	fixed	Toluidine blue	-
Pun et al. (2005)	Relaxometry	T1, MWF	Rat	fresh	Toluidine blue	-
Reeves et al. (2015)	Relaxometry	T1, T2, T2*	Human	fixed	MBP	-
Schmierer et al. (2010)	Relaxometry	T1	Human	fixed	MBP	-
Schmierer et al. (2004)	Relaxometry	T1	Human	fresh	LFB	+
Schmierer et al. (2007a)	Relaxometry	T1	Human	fresh	LFB	+
Schmierer et al. (2008)	Relaxometry	T1, T2	Human	fresh / fixed	LFB	-
Seewann et al. (2009)	Relaxometry	T1	Human	fixed	LFB, PAS, PLP	-
Soustelle et al. (2019)	Relaxometry	MWF	Mouse	perfusion fixed	MBP	+
Stüber et al. (2014)	Relaxometry	R1, R2*	Human	fixed	PIXE	-
Thiessen et al. (2013)	Relaxometry	T1	Mouse	perfusion fixed	EM	+
Wartjes et al. (2017)	Relaxometry	MWF, R1	Human	in situ	LFB	-
Webb et al. (2003)	Relaxometry	MWF	Rat	fresh	Toluidine blue	-
West et al. (2018)	Relaxometry	MWF	Mouse	perfusion fixed	EM	-

ies, we ran fixed-effects meta-analyses to estimate a cross-study validation effect size for each marker (Fig. 4). Except for mean diffusivity, we find meta-analytic evidence for correlations between histological markers and all markers investigated, with no clear marker having a stronger effect size than others. Finally, we show that choice of tissue type being investigated in different studies is unlikely to have affected our results (Supplementary Fig. S4).

3.3. P-value distribution bias and publication bias

Within the studies included in our meta-analyses, we then test for sources of bias. We find no evidence of abnormal p-value distribution, with a left-skewed p-curve and 80% of positive results providing p-values of 0.01 or lower (Supplementary Fig. S1). This indicates that conscious and unconscious bias towards significant results has likely not affected the study outcomes in the literature.

We then test for evidence of broader publication bias through a funnel plot (Supplementary Fig. S2). We find a significant correlation between sample size and effect size, where higher effect sizes in the literature are reported from studies with the smaller sample sizes.

While asymmetric funnel plots can arise from true heterogeneity in effect sizes across studies (Sterne et al., 2011), which may be the case in our studies, we perform a sensitivity analysis to test whether correcting for such asymmetry would alter the results. We perform

trim-and-fill procedures (Supplementary Fig. S3) based on (Duval and Tweedie, 2000) and find that if present, publication bias would inflate the validation effect size of some markers.

4. Discussion

Recommendations for *in vivo* human studies aiming to measure myelin with MRI and for future validation studies:

1. Recommendations for *in vivo* human studies aiming to measure myelin with MRI

- 1.1 Use acquisition protocols with multiple myelin-sensitive modalities (multimodal imaging).
- 1.2 Balance evidence from theoretical MR modelling and from histological validation studies when selecting MR markers to test hypotheses on myelin.
- 1.3 When selecting an MR marker based on individual validation studies, consider the validation design: a marker validated only in within-subject studies may not be sensitive to between-subject variability in myelination, and vice versa.

2. Recommendations for experimental protocols in validation studies

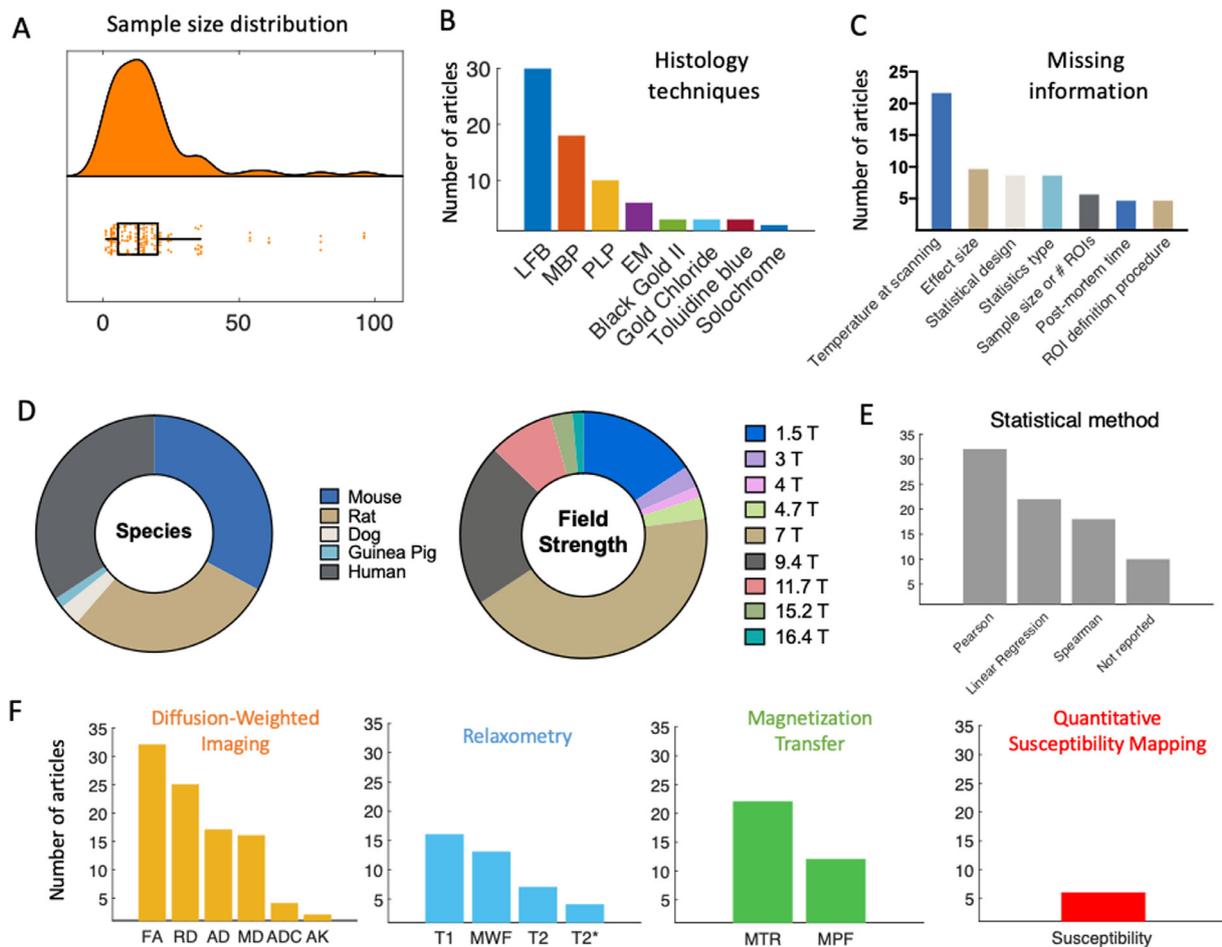


Fig. 3. Features of the MR-histology validation literature. **[A]** Sample size distribution of the included papers (median = 13 subjects). **[B]** The most frequently used histology techniques used in the articles are shown in decreasing order. **[C]** Due to lack of commonly agreed reporting standards, some relevant information may be missing from validation studies. The most commonly missing pieces of information on methods are reported here. **[D]** Pie chart of species and field strength used in each study. **[E]** Frequency of different statistical measures used for the validation analysis. **[F]** MR markers most commonly used in the literature, subdivided based on MR technique they are based on. The graphs show that DWI, MT and relaxometry are by far the most commonly explored techniques in validation studies. Within studies using DWI, tensor-based measures are predominant.

- 2.1 Report fixation protocols, and ideally use established ones.
- 2.2 Monitor and report scanning temperature.
- 2.3 In human studies, report and account for post-mortem time.
- 2.4 Use myelin histology specific for the experiment's objective, and where needed probe histology's specificity to the phenomenon of interest.
- 2.5 Use automated ROI definitions and co-registration between MR and histology.

3. Recommendations for statistical analyses in validation studies

- 3.1 Take the subject structure of the data into account - ROIs from the same subject are not independent, and multiple ROIs from multiple subjects cannot be pooled together without modelling the nested structure of the data.
- 3.2 Use correlation methods robust to outliers, and check distributional assumptions.
- 3.3 Covary for non-myelin factors such as axon density and iron.
- 3.4 Pre-register analyses and share data to reduce the impact of publication bias.

Our systematic review yielded 71 validation studies of microstructural MRI metrics to measure myelin, which included a wide range of MR markers and histological techniques. We then performed a detailed qualitative characterisation of this literature, as well as a quantitative meta-analysis of the reported effects sizes.

4.1. Many MRI markers correlate with myelin

Our meta-analyses of 7 commonly used MR markers (FA, RD, MD, susceptibility, R1, MTR and MP) show that there is evidence all these markers correlate with myelin, with the exception of MD. The effect sizes for each of these markers are pooled across a heterogeneous literature, which highlights the robustness of MR myelin markers to features of individual validation studies, such as field strength, animal models, and histological measure. This indicates that correlations between histological and MR-based markers of myelin are not restricted to any one setting, and is promising for translational uses of MR-based markers in contexts where obtaining histology is not possible.

4.2. Challenges in measuring myelin for in vivo studies in humans

One important question that the validation literature aims to tackle is whether one MR marker outperforms others in how well it captures myelin signals from a tissue. While some markers show stronger meta-

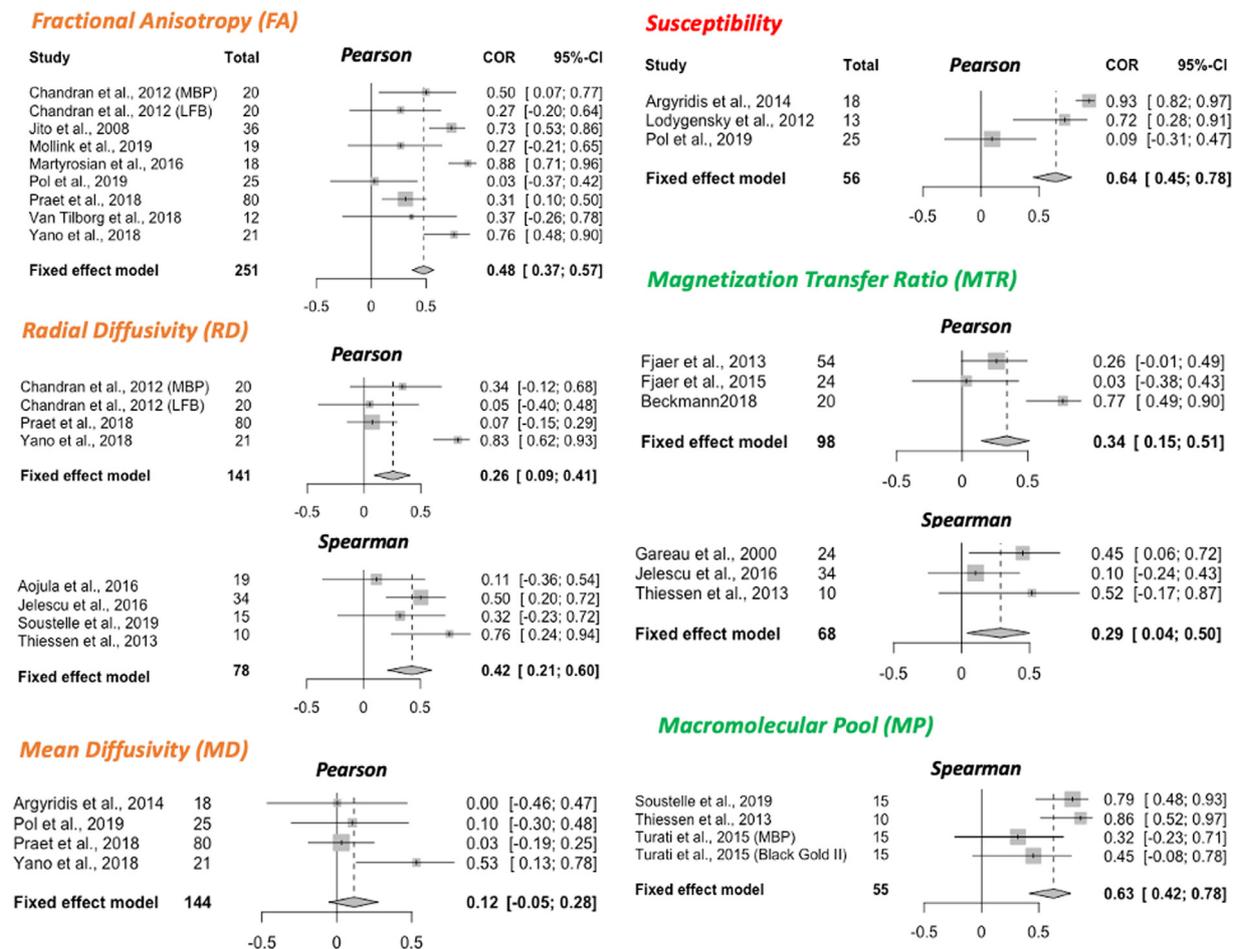


Fig. 4. Meta-analysis of correlations between histology and MR markers. Forest plots from meta-analyses of 6 MR markers (FA, RD, MD, susceptibility, MTR and MP). Spearman and Pearson correlations are pooled separately; the type of correlation coefficient is indicated at the top of each forest plot. All studies included use a between-subject statistical design; meta-analyses of studies with a within-subject design are reported in Fig. 5. For each study, sample size and effect size (R^2) with the respective confidence interval are provided. The fixed effect model results are reported at the bottom of each forest plot, together with pooled sample size across all included studies and with 95% Confidence Interval of the meta-analytic effect size.

analytic effect sizes than others, it is difficult to infer whether different markers relate to myelin to different extents. In particular, two key issues prevent making conclusive comparisons on how different MR markers compare with each other. One is the difference in methodologies between studies – effect sizes may be influenced by factors ranging from tissue processing to scanning temperature, and the small number of studies available makes it difficult to systematically account for these factors. The other is the inconsistent reporting: studies report results inconsistently (e.g. in terms of reported statistics) and sometimes fail to report crucial details. The heterogeneity in methodology also makes it difficult to systematically assess whether some factors such as tissue type affect the validity of different markers, e.g. it is impossible to disentangle whether some markers are more valid in white matter than gray matter or vice versa.

In this respect, our results underscore the importance of employing multimodal acquisitions, as observing similar effects across multiple markers is at present the best way to verify hypotheses related to myelin (**Recommendation 1.1**). This is particularly important for studies testing hypotheses on the amount of myelin, where the hypothesis is independent of the biophysical mechanism used to measure myelin. Techniques to explicitly combine multiple imaging modalities already exist (e.g. joint inference (Winkler et al., 2016), multimodal combination Mangeat et al. (2015)), and many studies have already found that multiple myelin markers can provide complementary information

(Callaghan et al., 2014; Eichert et al., 2020; Kolind et al., 2008; Lipp et al., 2019; Yeatman et al., 2014).

Another crucial aspect highlighted by our results is the need to balance evidence from theoretical models with evidence from validation studies, when selecting MR markers for a human *in vivo* study on myelination (**Recommendation 1.2**). This is particularly clear for diffusion MRI, where advanced models, such as the Spherical Mean Technique (Kaden et al., 2016), neurite orientation dispersion and density imaging (NODDI; Zhang et al. (2012)), and diffusion basis spectrum imaging (Wang et al., 2011), are becoming increasingly popular compared to tensor-based metrics such as FA, but are vastly under-represented in the validation literature (Alexander et al., 2019; Caspers and Axer, 2019). Therefore, it will be key to the continued success of these new methods to verify what they are sensitive to at the cellular level. In particular, well conducted validation studies will be critical to confirm the improved microstructural sensitivity of these markers compared to markers from simpler tensor models.

A notable absence from our meta-analyses is Myelin Water Fraction (MWF). While 14 studies in total examine MWF, a high number of these studies use a mixed within- and between- subject design ($n = 7$), do not report the type of static measure used ($n = 1$), or use sample sizes too small for inclusion ($n = 1$), which is why they did not fit our inclusion criteria for the meta-analysis. The remainder of the studies ($n = 5$) use different methodologies (within- and between-subject design) and

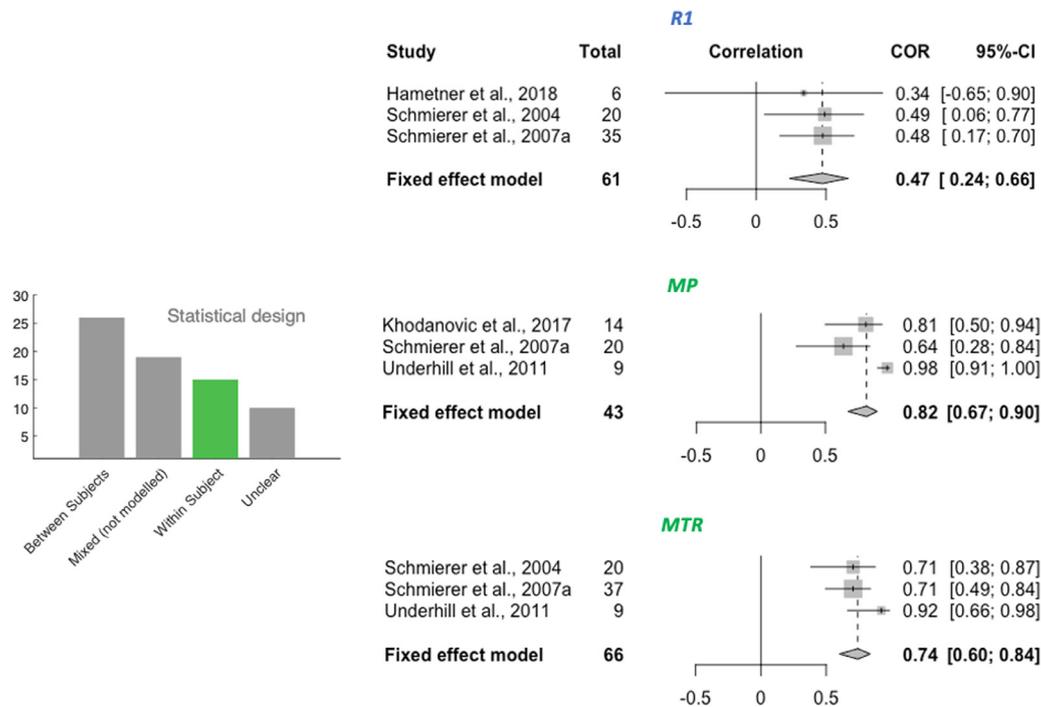


Fig. 5. Within-subject meta-analyses. For the three metrics R1, MP and MTR, a meta-analysis on within-subject correlations could be performed. The resulting effect size estimate for MTR is higher than its between-subject effect size (see Fig. 4).

outcome measures (Spearman and Pearson) from each other, making it difficult to pool them together.

A final recommendation for human *in vivo* studies is to take validation parameters into account when selecting which MR marker is most appropriate for a given study (**Recommendation 1.3**). Some markers are only validated with one type of statistical design, which limits their potential applications. In the studies we examined, relaxometry metrics such as R1 are more commonly validated with within-subject designs, which means it is unclear whether they would be able to pick up between-subject variability in myelination in the general population. Therefore, microstructural studies computing brain-behaviour correlations, where between-subject variability in myelination is key to the hypothesis tested (Johansen-Berg, 2010), might benefit from including markers that have been validated in between-subject designs in order to maximise their sensitivity to the phenomenon of interest.

4.3. Challenges in acquiring MRI data for validation studies

MR signals are affected by a variety of data acquisition choices. For instance, we find that the studies in our review validate MR metrics with either *in vivo* or *ex vivo* MR scanning, each of which has advantages and disadvantages. For studies using human brains, the option to combine *in vivo* scanning with *ex vivo* histology is rare (see Treit et al., 2019), whereas in animal studies, both *in vivo* and *ex vivo* scanning are possible. *In vivo* scanning has the advantage that it better resembles the parameters of *in vivo* human studies, while *ex vivo* scanning has the advantage that the tissue is scanned and processed for histology in a similar state. Also, higher resolution can be achieved with *ex vivo* imaging, which in studies of animals with smaller brains provides more anatomically comparable detail to human studies with larger voxel sizes (Lerch et al., 2012).

4.3.1. Ex vivo tissue fixation affects microstructural MRI

Tissue fixation is a necessary step for stopping tissue decomposition in *ex vivo* scans, but it also has a major influence on MR signals. Formalin fixation and the embedding medium for scanning can affect image

quality and MRI parameters (Dusek et al., 2019), even when excess fixative is being washed out of the tissue before scanning. This is because fixation works by linking amino acids in proteins, which means that the molecular tissue structure is inherently changed during the process (Thavarajah et al., 2012). Moreover, MR relaxation relies on water protons behaving differently depending on their molecular environment, and fixation does change this environment, for example by affecting membrane permeability (Shepherd et al., 2009). The same principle applies to diffusion-weighted signals, which are affected by the effects of fixation on water diffusion (Shepherd et al., 2009; Sun et al., 2005).

Fixation-dependent changes in MRI have two effects on validation studies. First, if validation evidence only comes from fixed tissue, the generalisability to fresh tissue is not guaranteed. Among the studies we analysed, only a few of them assessed tissue that was scanned fresh, whereas ideally, both states should be validated (e.g. Schmierer et al., 2008). Second, the variability in the type and duration of fixation can lead to variability in the MRI-based metrics. Longitudinal studies investigating tissue fixation report spatially and temporally varying effects on relaxation rates, MWF and also MP (Dawe et al., 2009; Seifert, 2019; Shatl et al., 2018). In the majority of studies we assessed, the fixation process ranged from a few days to a few weeks, which means the MR signal was likely very different between studies. Within an individual validation study, this influence may be minimized by using a standardised tissue processing protocol for all samples (**Recommendation 2.1**). However, systematic differences of fixation effects across different ROIs may still pose problems for the quantitative comparison between MRI and histology.

4.3.2. Scanning temperature affects microstructural MRI

Another factor to consider is temperature, which affects MRI parameters such as diffusion and relaxation (e.g. Birkel et al., 2016; Dhital et al., 2016). When scanning *in vivo*, the tissue is naturally kept at body temperature. However, the temperature conditions during post-mortem scanning are more flexible, allowing to either scan at room temperature or at a controlled temperature. Importantly, the tissue temperature could unintentionally change throughout the scanning session due to

gradient heating, making it important to monitor scanning temperature in *ex vivo* scans. If the temperature varied across samples or across ROIs within the same sample, this could induce unwanted variability in the MRI metrics. In the assessed *ex vivo* validation studies, the scanning temperature was frequently not reported, making it difficult to estimate its impact on the results and to compare microstructural MRI metrics across studies (**Recommendation 2.2**).

4.4. Challenges in obtaining ground-truth histological data for validation studies

4.4.1. No standard pipeline for histological quantification exists

Validation studies aim to obtain a ground-truth measure for myelin content, but we find that pipelines used to obtain myelin content histologically varied widely across the assessed papers. The vast majority of microscopic visualisation employed classical stainings (most frequently LFB and Gold-based stains) or immunohistochemistry approaches (most frequently anti-MBP and anti-PLP). These approaches have primarily been developed to visualise myelination patterns in the tissue, rather than quantify myelin content (e.g. see Carriel et al., 2017a; Kiernan, 2007; Thetiot et al., 2018; Woodhoo, 2018). Their advantage is that they allow imaging large fields of view, which is often needed in a validation study to compensate for the comparatively low spatial resolution of MRI relative to microscopy.

To analyse the resulting microscopy images, the majority of the assessed validation studies used either staining fraction (the relative amount of stained tissue) or average staining intensity in an area (the total amount of stain in the tissue) for quantification. However, both metrics come with limitations. Staining intensity is often assumed to scale linearly with myelin content. However, this is not always true, for example in cases where staining saturation can take place (e.g. in silver staining of white matter; Pistorio et al., 2006). Moreover, it is known that different samples (or different anatomical areas within the same sample) are not affected homogeneously by preparation steps such as fixation or antibody penetration (Dawe et al., 2009; Seehaus et al., 2015), which poses challenges in particular for between-subject validation designs. Using staining fraction to quantify myelin could circumvent this issue, if the image segmentation intensity threshold is adjusted flexibly (Mollink et al., 2019), depending on local intensity variations. However, it remains poorly understood to what extent staining fraction values are comparable across different staining methods.

One solution to address the limitations of these commonly used metrics would be to employ methods developed for quantification, rather than visualisation, of molecules. These methods include proton induced X-ray emission (PIXE, Stüber et al. (2014)) and lipid mass spectrometry (González de San Román et al., 2018), which could provide quantification of myelin for each imaging pixel, for example by estimating myelin concentration from sulfur and phosphorus (Stüber et al., 2014). With the exception of one study (Stüber et al., 2014), these methods have not been used in the validation papers we reviewed, but may be used in future studies to provide complementary information to traditional histological techniques.

Another disadvantage of traditional histology methods is that what they gain in field of view size they lose in resolution. Higher-resolution histological methods such as electron microscopy and coherent anti-Stokes Raman scattering microscopy can visualise even individual myelin sheaths within small tissue blocks (generally not more than a few mm of size). With sufficient image quality and appropriate image segmentation (Stikov et al., 2015b; Zaimi et al., 2016), these methods have the potential to estimate the myelin fraction of imaged tissue, as well as the packing density and total surface area of the myelin sheaths. While they are not as common in the validation literature, and often not feasible over large areas of the brain, these methods provide more detailed information compared to stainings or immunohistochemistry, and could provide further validation for microstructural MR markers (Sternberger et al., 1978; Vincze et al., 2008).

Finally, it is important to highlight that multiple different morphological aspects of the myelinated axon, including Nodes of Ranvier, sheath thickness, length and packing density - are all important to brain function (Almeida and Lyons, 2017; Xin and Chan, 2020). Most validation studies aim to quantify overall myelin content in a tissue, with the rationale that this is most closely related to the signals measured with MRI, and that for many applications in basic science and clinical translation, total myelin content measures can already be hugely beneficial. However, future studies may also want to assess what structural features of the myelinated axon are most closely related to different MR signals. This might allow us to go beyond testing hypotheses on the amount of myelin in a voxel.

4.4.2. Tissue conditions influence histological quantification

The success of any histological method does not only depend on its molecular principles underlying myelin visualisation, but also on experimental factors, such as tissue processing protocols (e.g. Werner et al., 2000). In our review, processing of the tissue for histology varied considerably across studies, with both paraffin embedding and cryosectioning being used with about equal frequency. While tissue fixation or freezing do not have a strong effect on lipids (Carriel et al., 2017a), extraction by solvents, such as used in paraffin embedding, extracts most lipids and only retains those covalently bound to protein (Carriel et al., 2017b; Kiernan, 2007). The myelin structure in the stained tissue is therefore inherently affected by how it was processed for histology and may not be comparable across studies that used different processing strategies (**Recommendation 2.1**).

In studies with human tissue, a further complication is that post-mortem time also influences histological staining. The length of the post-mortem time indicates the advance of the tissue autolysis and therefore the microstructural intactness of the tissue (Sele et al., 2019), but it can vary between a few hours and a few days between samples. While a considerable number of assessed papers did not report post-mortem times, in those who did, it varied between 4 hours and 3 days. Large variability in post-mortem times across tissue samples may induce unwanted variability in the histological quantification, and may need to be taken into account during the analysis as a covariate (**Recommendation 2.3**).

4.4.3. Histological specificity for myelin may depend on the experimental context

In our systematic review, most of the validation studies assessed only used one histological method. However, as different histological methods rely on wide-ranging and sometimes poorly understood molecular mechanisms, the agreement between MRI and histology will also depend on which histological marker is used. In the few papers that used multiple histological methods, validation results were not always the same for the different histological markers used (e.g. Kozłowski et al., 2008; Laule et al., 2011). This may be because the extent to which the histological quantification is an accurate representation of myelin content might differ based on the characteristics of the sample or pathology in question. Myelin is a highly complex substrate, with the dry mass consisting of about 70% lipids and 30% proteins (Gopalakrishnan et al., 2013; O'Brien and Sampson, 1965). Its composition can vary between tissue types (González de San Román et al., 2018), species (Gopalakrishnan et al., 2013) and in pathologies (Wheeler et al., 2008), and therefore different visualization strategies may be useful for different scenarios.

For example, immunohistochemistry can be used to target specific myelin constituents such as MBP and PLP, using antibodies that are attached to (most often through secondary antibodies) visualising elements, such as fluorescent molecules or diaminobenzidine tetrahydrochloride. If the target molecules are affected in the specific sample studied, this will affect the visualisation results. In studies using Shiverer mice, which have a genetic mutation for the MBP-gene (Molineaux et al., 1986), MBP-immunohistochemistry is well suited to demonstrate the genetic intervention, but could lead to a biased myelin quantification when used on Shiverer mice in a validation study.

Unlike immunohistochemistry, classical myelin stains are not specific to individual macromolecules of myelin, but make use of myelins biochemical characteristics (Kiernan, 2007). For example, LFB is likely attracted by the basic amino acids of myelins proteins and by phospholipids (de Almeida and Pearse, 1958; Kiernan, 2007; Klüver and Barrera, 1953; Salthouse, 1962a; 1962b), whereas Gold atoms present in Gold stains are chemically reduced by myelin lipids in formaldehyde fixed tissue, leading to the black appearance of gold staining in myelin fibres (Schmued et al., 2008; Schmued and Slikker, 1999). While these stains are less specific to individual myelin proteins, their success is also sample-dependent: Vincze et al. (2008) found that in comparison to anti-MBP immunohistochemistry, LFB only revealed myelination in later developmental stages.

In summary, the variability in histological methods within our review highlights the need to use myelin histology specific for each experiment's objective (**Recommendation 2.4**). In contexts where myelin histology has never been used before, this might mean probing the specificity of different histological methods to the phenomenon of interest.

4.4.4. Matching MRI and histology: ROI definition and image coregistration

Another factor that will affect MRI-histology correlations is the spatial coregistration between the two modalities. Even if both the microstructural MRI metric and the histology perfectly capture myelination, the correlation between the two will be low if the measures are taken from spatial locations that are not well matched between MR and histology. Here, the 2D nature of most histology techniques makes spatial coregistration challenging. Most of the assessed validation papers try to achieve spatial correspondence without coregistration (only 21 out of 70 studies use coregistration), but rather by manually outlining ROIs in the same anatomical location, identifying these ROIs based on landmarks.

The manual approach is commonplace, but has two pitfalls. First, placing ROIs may vary depending on the experimenter performing it, and automatic atlas-based ROI definitions are recommended practice in the neuroimaging field (Nichols et al., 2017). Second, coregistration may greatly improve the spatial correspondence between MR and histological images. While coregistration of MR images and histological images has unique challenges, such as the different spatial scales, changes in morphometry due to the tissue processing for histology, and potential cracks and folds in the tissue sections (Pichat et al., 2018), it has the advantage that ROI definition can be performed more reproducibly. Advances in automatic co-registration tools could aid a more widespread implementation in validation studies (Huszar et al., 2019) (**Recommendation 2.5**).

4.5. Challenges in statistical comparisons of MRI and histology

Our meta-analysis focussed on studies using a correlative approach to measure the accuracy of MR metrics for validation. However, there are two key issues that are not addressed by correlative accuracy studies. First, validity depends on accuracy and accuracy depends on precision. Many of the studies in our review focus on accuracy, i.e. the measured parameter being in agreement with the true underlying biological parameter of interest. However, precision, i.e. low variability in repeated measurements, and thus high reliability, is also crucial to validity of a metric. In the case of MRI metrics, repetition across time (e.g. Arshad et al., 2017; Lévy et al., 2018), across hardware (e.g. Bane et al., 2018; Leutritz et al., 2020) and across sequences (e.g. Stikov et al., 2015a) are all important, and, while outside the scope of this review, they are also a key prerequisite for accuracy.

Second, correlative approaches are key to assessing accuracy in histological validation studies, because for most MR metrics, there is no mathematical model capable of converting the units of the metric to readouts from myelin histology. However, the drawback of using shared variance as a measure for validity is that variance in the data is shaped

by a variety of factors, including measurement noise and underlying variance in 'ground-truth myelination'. Therefore, studies with high measurement noise or with low myelin variance (e.g. between subjects of the same group) may artificially deflate the true underlying correlation coefficient (Altman and Bland, 1983; Goodwin and Leech, 2006). Likewise, correlation does not equal causation and factors that affect both MRI and histology (such as post-mortem times) may artificially inflate or deflate correlation coefficients. A few studies have aimed to use non-correlative approaches for validation of MR metrics, for example by demonstrating effects from a given intervention on MR and histology (Lodygensky et al., 2012; Sampaio-Baptista et al., 2013), but fall outside the scope of this review.

4.5.1. Achieving robust and meaningful correlations

As highlighted above, MR markers within our review have been validated with either within-subject or between-subject designs, and each design provides different information about the marker. We also found that a sizable subset of studies ($n = 19$) pooled together multiple ROIs from multiple subjects, effectively measuring a mixture of both within and between-subject design, but without modelling each contribution independently. Additionally, in 8 studies, we could not deduce which variance was modeled. This limits the interpretability of the results, and fails to take the nested structure of the dataset into account. Therefore, future studies may want to perform analyses accounting for the subject structure in the data, thus maximising their interpretability (**Recommendation 3.1**).

In studies measuring within-subject variance (i.e. spatial covariance of MR and histological metrics), we found two key drawbacks. First, correlation metrics are often reported for individual subjects, making it difficult to pool results across studies. Second, the extent of variability may not be comparable to between-subject variability, as contrast between grey and white matter can be a strong driver of variance. This phenomenon is exemplified in Laule et al. (2006) and Peters et al. (2019), where correlation coefficients are lower in analyses including only ROIs from white matter, compared to analyses including a more heterogeneous set of ROIs. This is also reflected in our meta-analytical results for MTR (Figs. 4 and 5), where the correlation coefficients were lower in between-subject studies (95% CI for R^2 : 0.15 and 0.51) compared to within-subject studies (95% CI for R^2 : 0.60 and 0.84).

In studies measuring between-subject variance a notable source of variability was that the data belonged to multiple groups (e.g. different ages, different pathology severity) (Thiessen et al., 2013). While this provides the correlation with more power, spurious correlations are a concern when dealing with multiple groups. Therefore, it is important for future validation studies to check test distributional assumptions, check the raw data for spurious correlations, and use correlation methods that are robust to outliers (Salibian-Barrera and Zamar, 2002; Wilcox, 2016) (**Recommendation 3.2**), especially for studies using a between-subject design.

4.5.2. Including potential biological confounds as covariates

Myelin is not isolated from other microstructural features of brain tissue. In healthy tissue, myelin content is often related to iron and axon density. Iron occurs in high concentrations in oligodendrocytes and is important for the production and maintenance of myelin (Möller et al., 2019). It affects some MRI metrics similarly to myelin (Möller et al., 2019) and its distribution resembles myeloarchitecture (Fukunaga et al., 2010). Axonal density may also correlate with total amount of myelin, if axons are myelinated. This can be a problem for MR sequences that are sensitive to signals from axons, such as diffusion imaging, where axonal membranes affect diffusion signals even in the absence of myelin (Beaulieu, 2002). The concern about biological confounds is further corroborated by studies which perform MRI-histology correlations with various histological markers. For example, Jespersen et al. (2010) finds a correlation between FA and myelin stain of $r = 0.78$, and a correlation

between FA and cell density of $r = 0.72$, thus questioning the specificity of the FA-myelin correlation in the study.

For both iron and axons, colocalization can be even more pronounced in pathological samples. Conditions such as MS, or animal models such as Cuprizone-fed and Shiverer mice, are often used in validation studies, because of their known effect on myelin, but they also affect iron metabolism, due to iron's role in myelin maintenance (e.g. Hametner et al., 2013; Pandur et al., 2019; Sergeant et al., 2005) and axons, since myelin and axonal health are tightly coupled (Stassart et al., 2018).

Therefore, these biological confounds can often have an impact on the correlation coefficient between histology and MRI metrics. For example, an MRI marker sensitive to iron may yield a positive correlation with myelin histology, if myelin and iron colocalise in the tissue of interest. To test for myelin's unique contribution to correlations, one needs to quantify and correct for these biological confounds (**Recommendation 3.3**). Within the studies selected for our systematic review, only a few ($n = 7$) performed iron histology, while half of the studies considered axons ($n = 35$). Of these studies, only a small subset considered iron and axons as confounds in their analyses of myelin histology.

4.5.3. Biases in the validation literature

Publication bias and questionable research practices have long been established as a key factor hindering pooling of evidence across studies in biomedical research (Ahmed et al., 2012). In our analyses, we find no evidence of abnormal p -value distributions in the validation literature, but we do find an asymmetry in the correlation between effect size and sample size of validation studies, which is often considered an indication of publication bias.

Asymmetric funnel plots can be driven by many factors. Selective outcome reporting and selective analysis reporting are the most common. However, not all asymmetric funnel plots are due to issues that impact meta-analytic inference (Sterne et al., 2011). For instance, in some circumstances sampling variation and chance can lead to an asymmetric funnel plot without real publication bias. Moreover, in the case of validation studies, it is expected that true effect sizes would be heterogeneous, especially when pooling together studies that examined how the same MR marker relates to different histological markers.

Our results show that correcting for publication bias would impact meta-analytic validation evidence for some of the MR metrics we analysed. These results need to be interpreted with caution, as our meta-analyses include relatively few studies, but are strengthened by the fact that 10 out of 71 studies in our review do not report results for all the metrics they collect. Taken together, these observations suggest that measures to prevent publication bias may be useful for future validation studies. Pre-registered protocols with pre-specified power analyses can strengthen the robustness of effect size estimates, and even provide more accurate estimates than meta-analytic analyses themselves (Kvarven et al., 2020). Therefore, using pre-registration may help establish more accurate effect sizes for the correlations between MR metrics and myelin histology (**Recommendation 3.4**).

4.6. Conclusions and future directions

Our meta-analysis finds evidence for correlations between myelin histology and a range of MR markers: FA, RD, susceptibility, R1, MP and MTR, but not MD. These results verify that many MR markers are sensitive to myelin, but our analyses could not identify a single marker that is more sensitive to myelin than others. This suggests that for the time being, using multiple microstructural imaging markers in parallel may be the best way to test hypotheses related to myelin in humans in vivo.

We also find that the literature has a number of limitations. First, a wide variety of methodological approaches were used across the studies we assessed, making it challenging to estimate overall effect sizes and compare the effect sizes of different MR markers. Second, we find that heterogeneous and inconsistent reporting makes it difficult to assess

the quality of the studies, and the factors driving differences in results. Third, we find some evidence of inflated effect sizes due to publication bias. Tackling these issues will be crucial to improving our strategies to measure myelin in humans.

Credit authorship contribution statement

Alberto Lazari: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Ilona Lipp:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Project administration.

Acknowledgements

We thank Aman Badhwar for feedback on the meta-analysis. We thank Alexander S. Bates, Michiel Cottar, Nicole Eichert, Heidi Johansen-Berg, Maria Morozova, Ruairi Roberts, Zeena-Britt Sanders and Nikolaus Weiskopf for feedback on previous versions of the manuscript. We thank Edgar Liberis for his help with curating the online resource, Thomas Wassenaar for his advice on AMSTAR guidelines and Gunther Helms for his input on the macromolecular pool measure. IL is funded by the Max-Planck-Society. AL is supported by a PhD Studentship from the Wellcome Trust (109062/Z/15/Z). Fig. 1 was created with BioRender.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.neuroimage.2021.117744](https://doi.org/10.1016/j.neuroimage.2021.117744).

References

- Abe, Y., Komaki, Y., Seki, F., Shibata, S., Okano, H., Tanaka, K.F., 2019. Correlative study using structural MRI and super-resolution microscopy to detect structural alterations induced by long-term optogenetic stimulation of striatal medium spiny neurons. *Neurochem. Int.* 125 (2018), 163–174.
- Ahmed, I., Sutton, A.J., Riley, R.D., 2012. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ* 344 (7838), 1–10.
- Alexander, D.C., Dyrby, T.B., Nilsson, M., 2019. Imaging brain microstructure with diffusion MRI: practicality and applications. *NMR Biomed.* 32, e3841.
- de Almeida, D.F., Pearse, A.G., 1958. Comparative histochemistry of lipids in relation to myelination in rabbit brain. *J. Neurochem.* 3 (2), 132–138.
- Almeida, R.G., Lyons, D.A., 2017. On myelinated axon plasticity and neuronal circuit formation and function. *J. Neurosci.* 37 (42), 10023–10034.
- Alonso-Ortiz, E., Levesque, I.R., Pike, G.B., 2015. MRI-based myelin water imaging: a technical review. *Magn. Reson. Med.* 73, 70–81.
- Altman, D., Bland, J., 1983. Measurement in medicine: The analysis of method comparison studies. *J. R. Stat. Soc. Ser. D* 32 (3), 307–317.
- Aojula, A., Botfield, H., Patterson, J., Ii, M., Gonzalez, A.M., Abdullah, O., Logan, A., Sinclair, A., 2016. Diffusion tensor imaging with direct cytopathological validation: characterisation of decorin treatment in experimental juvenile communicating hydrocephalus. *Fluids Barriers CNS* 13 (9).
- Argyridis, I., Li, W., Johnson, G.A., Liu, C., 2014. Quantitative magnetic susceptibility of the developing mouse brain reveals microstructural changes in the white matter. *NeuroImage* 88, 134–142.
- Arshad, M., Stanley, J.A., Raz, N., 2017. Test-retest reliability and concurrent validity of in vivo myelin content indices: Myelin water fraction and calibrated T1w/T2w image ratio. *Hum. Brain Mapp.* 38, 1780–1790.
- Bagnato, F., Hametner, S., Boyd, E., Endmayr, V., Shi, Y., Ikonomidou, V., Chen, G., Pawate, S., Lassmann, H., Smith, S., Welch, E.B., 2018. Untangling the R2* contrast in multiple sclerosis: a combined MRI-histology study at 7.0 Tesla. *PLoS One* 13 (3), e0193839.
- Bane, O., Hectors, S.J., Wagner, M., Arlinghaus, L.L., Aryal, M.P., Cao, Y., Chenevert, T.L., Fennessy, F., Huang, W., Hylton, N.M., Kalpathy-Cramer, J., Keenan, K.E., Malwareenko, D.I., Mulkern, R.V., Newitt, D.C., Russek, S.E., Stupic, K.F., Tudorica, A., Wilmes, L.J., Yankeelov, T.E., Yen, Y.F., Boss, M.A., Taouli, B., 2018. Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: results from a multicenter phantom study. *Magn. Reson. Med.* 79 (5), 2564–2575.
- Barros, A.D., Arribarat, G., Combis, J., Chaynes, P., 2019. Matching ex vivo MRI with iron histology: pearls and pitfalls. *Front. Neuroanat.* 13, Article68.
- Barta, R., Kalantari, S., Laule, C., Vavasour, I.M., MacKay, A.L., Michal, C.A., 2015. Modeling T1 and T2 relaxation in bovine white matter. *J. Magn. Reson.* 259, 56–67.
- Basser, P.J., 2004. Scaling laws for myelinated axons derived from an electrotonic core-conductor model. *J. Integr. Neurosci.* 3 (2), 227–244.

- Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR Biomed.* 15 (7–8), 435–455.
- Beaulieu, C., 2014. The biological basis of diffusion anisotropy. In: Johansen-Berg, H., Behrens, T. (Eds.), *Diffusion MRI*, pp. 155–184.
- Beckmann, N., Giorgetti, E., Neuhaus, A., Zurbrugg, S., Accart, N., Smith, P., Perdoux, J., Perrot, L., Nash, M., Desrayaud, S., Wipfli, P., Frieauff, W., Shimshek, D.R., 2018. Brain region-specific enhancement of remyelination and prevention of demyelination by the CSF1R kinase inhibitor BLZ945. *Acta Neuropathol. Commun.* 6 (1), 9.
- Birkel, C., Langkammer, C., Golob-Schwarzl, N., Leoni, M., Haybaeck, J., Goessler, W., Fazekas, F., Ropele, S., 2016. Effects of formalin fixation and temperature on MR relaxation times in the human brain. *NMR Biomed.* 29 (4), 458–465.
- Bot, J.C.J., Blezer, E.L.A., Kamphorst, W., Lycklama, G.J., Ader, H.J., Castelijns, J.A., Ig, K.N., 2004. The spinal cord in multiple sclerosis: relationship of high-spatial-resolution quantitative MRI imaging findings to histopathologic results. *Neuroradiology* 233, 531–540.
- Button, K.S., Ioannidis, J.P.a., Mokrysz, C., Nosek, B.a., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14 (5), 365–376.
- Callaghan, M.F., Freund, P., Draganski, B., Anderson, E., Cappelletti, M., Chowdhury, R., Diedrichsen, J., Fitzgerald, T.H.B., Smittenaar, P., Helms, G., Lutti, A., Weiskopf, N., 2014. Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. *Neurobiol. Aging* 35 (8), 1862–1872. doi:10.1016/j.neurobiolaging.2014.02.008.
- Campbell, J.S.W., Leppert, I.R., Narayanan, S., Duval, T., Cohen-Adad, J., Pike, G.B., Stikov, N., 2017. Promise and pitfalls of g-ratio estimation with MRI. *NeuroImage* 182, 80–96.
- Carriell, V., Campos, A., Alaminos, M., Raimondo, S., Geuna, S., 2017. Staining methods for normal and regenerative myelin in the nervous system. In: *Histochemistry of Single Molecules*, 1560, pp. 219–229.
- Carriell, V., Campos, F., Aneiros-Fernandez, J., Kiernan, J., 2017. Tissue fixation and processing for the histological identification of lipids. In: *Histochemistry of Single Molecules*, 1560, pp. 219–229.
- Caspers, S., Axer, M., 2019. Decoding the microstructural correlate of diffusion MRI. *NMR Biomed.* 32, e3779.
- Chandran, P., Upadhyay, J., Markosyan, S., Lisowski, A., Buck, W., Chin, C.L., Fox, G., Luo, F., Day, M., 2012. Magnetic resonance imaging and histological evidence for the blockade of cuprizone-induced demyelination in C57BL/6 mice. *Neuroscience* 202, 446–453.
- Chang, E.H., Argyelan, M., Aggarwal, M., Chandon, T.-s.S., Karlsgodt, K.H., Mori, S., Malhotra, A.K., 2017. Diffusion tensor imaging measures of white matter compared to myelin basic protein immunofluorescence in tissue cleared intact brains. *Data Brief* 10, 438–443.
- Chang, E.H., Argyelan, M., Aggarwal, M., Chandon, T.S.S., Karlsgodt, K.H., Mori, S., Malhotra, A.K., 2017. The role of myelination in measures of white matter integrity: Combination of diffusion tensor imaging and two-photon microscopy of CLARITY intact brains. *NeuroImage* 147, 119–124.
- Chen, H.S.M., Holmes, N., Liu, J., Tetzlaff, W., Kozłowski, P., 2017. Validating myelin water imaging with transmission electron microscopy in a rat spinal cord injury model. *NeuroImage* 153, 122–130.
- Choi, J., Dickson, P., Calabrese, E., Chen, S., White, L., Ellingwood, M., Provenzale, J.M., 2015. Predicting degree of myelination based on diffusion tensor imaging of canines with mucopolysaccharidosis type I. *Neuroradiol.* J. 28 (6), 562–573.
- Cohen-Adad, J., 2014. What can we learn from T2* maps of the cortex? *NeuroImage* 93, 189–200.
- Cohen-Adad, J., 2018. Microstructural imaging in the spinal cord and validation strategies. *NeuroImage* 182, 169–183.
- Dawe, R.J., Bennett, D.A., Schneider, J.A., Vasireddi, S.K., Arfanakis, K., 2009. Post-mortem MRI of human brain hemispheres: T2 relaxation times during formaldehyde fixation. *Magn. Reson. Med.* 818, 810–818.
- De Santis, S., Drakesmith, M., Bells, S., Assaf, Y., Jones, D.K., 2014. Why diffusion tensor MRI does well only some of the time: variance and covariance of white matter tissue microstructure attributes in the living human brain. *NeuroImage* 89, 35–44.
- Dhital, B., Labadie, C., Stallmach, F., Möller, H.E., Turner, R., 2016. Temperature dependence of water diffusion pools in brain white matter. *NeuroImage* 127, 135–143.
- Does, M.D., 2018. Inferring brain tissue composition and microstructure via MR relaxometry. *NeuroImage* 182, 136–148.
- Duhamel, G., Prevost, V.H., Cayre, M., Hertenau, A., Mchinda, S., Carvalho, V.N., Varma, G., Durbec, P., Alsop, D.C., Girard, O.M., 2019. Validating the sensitivity of inhomogeneous magnetization transfer (ihMT) MRI to myelin with fluorescence microscopy. *NeuroImage* 199, 289–303.
- Dusek, P., Madai, V., Huelnhagen, T., Bahn, E., Matej, R., Sobesky, J., Niendorf, T., Acosta-Cabronero, J., Wuerfel, J., 2019. The choice of embedding media affects image quality, tissue R2*, and susceptibility behaviors in post-mortem brain MR. *Magn. Reson. Med.* 81, 2688–2701.
- Duval, S., Tweedie, R., 2000. Trim and fill: a simple funnel-plot-based method. *Biometrics* 56 (June), 455–463.
- Edwards, L.J., Kirilina, E., Mohammadi, S., Weiskopf, N., 2018. Microstructural imaging of human neocortex in vivo. *NeuroImage* 15, 184–206.
- Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.* 315 (7109), 629–634.
- Eichert, N., Papp, D., Mars, R. B., Watkins, K. E., 2020. Mapping human laryngeal motor cortex during vocalization. *Cerebral Cortex* 30 (12), 6254–6269.
- Fatemi, A., Wilson, M.A., Phillips, A.W., McMahon, M.T., Zhang, J., Smith, S.A., Arauz, E.J., Falahati, S., Gummadavelli, A., Bodagala, H., Mori, S., Johnston, M.V., 2011. In vivo magnetization transfer MRI shows dysmyelination in an ischemic mouse model of periventricular leukomalacia. *J. Cerebral Blood Flow Metab.* 31 (10), 2009–2018.
- Fjær, S., Bø, L., Lundervold, A., Myhr, K.-m., Pavlin, T., Torkildsen, Ø., Wergeland, S., 2013. Deep gray matter demyelination detected by magnetization transfer ratio in the cuprizone model. *PLoS One* 8 (12), e84162.
- Fjær, S., Bø, L., Myhr, K.M., Torkildsen, O., Wergeland, S., 2015. Magnetization transfer ratio does not correlate to myelin content in the brain in the MOG-EAE mouse model. *Neurochem. Int.* 83–84, 28–40.
- Fukunaga, M., Li, T.Q., Van Gelderen, P., De Zwart, J.A., Shmueli, K., Yao, B., Lee, J., Maric, D., Aronova, M.A., Zhang, G., Leapman, R.D., Schenck, J.F., Merkle, H., Duyn, J.H., 2010. Layer-specific variation of iron content in cerebral cortex as a source of MRI contrast. *Proc. Natl. Acad. Sci. USA* 107 (8), 3834–3839.
- Fünfschilling, U., Supplie, L.M., Mahad, D., Boretius, S., Saab, A.S., Edgar, J., Brinkmann, B.G., Kassmann, C.M., Tzvetanova, I.D., Möbius, W., Diaz, F., Meijer, D., Suter, U., Hamprecht, B., Sereda, M.W., Moraes, C.T., Frahm, J., Goebbels, S., Nave, K.A., 2012. Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature* 485 (7399), 517–521.
- Gareau, P.J., Rutt, B.K., Karlik, S.J., Mitchell, J.R., 2000. Magnetization transfer and multicomponent T2 relaxation measurements with histopathologic correlation in an experimental model of MS. *J. Magn. Reson. Imaging* 11 (6), 586–595.
- Gibson, K., Peterson, A., 1991. *Brain Maturation and Cognitive Development*. Aldine de Gruyter, New York.
- Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* 31 (32), 11597–11616.
- Goldman, L., Albus, J.S., 1968. Computation of impulse conduction in myelinated fibers; theoretical basis of the velocity-diameter relation. *Biophys. J.* 8 (5), 596–607.
- González de San Román, E., Bidmon, H.-J., Malisic, M., Susnea, I., Küppers, A., Hüblers, R., Wree, A., Nischwitz, V., Amunts, K., Huesgen, P.F., 2018. Molecular composition of the human primary visual cortex profiled by multimodal mass spectrometry imaging. *Brain Struct. Funct.* 223 (6), 2767–2783.
- Goodwin, L.D., Leech, N.L., 2006. Understanding correlation: factors that affect the size of r. *J. Exp. Educ.* 74 (3), 249–266.
- Gopalakrishnan, G., Awasthi, A., Belkaid, W., Jr, O.D.F., Liazoghli, D., Colman, D.R., Dhaunchak, A.S., 2013. Lipidome and proteome map of myelin membranes. *J. Neurosci. Res.* 334 (May 2012), 321–334.
- Grossman, R.I., Gomori, J.M., Ramer, K.N., Lexa, F.J., Schnall, M.D., 1994. Magnetization transfer: theory and clinical applications in neuroradiology. *Radiographics* 14 (2), 279–290.
- Grussu, F., Schneider, T., Tur, C., Yates, R.L., Tachrount, M., Deluca, G.C., Wheeler-Kingshott, C.A.M.G., 2017. Neurite dispersion: a new marker of multiple sclerosis spinal cord pathology? *Ann. Clin. Transl. Neurol.* 4 (9), 663–679.
- Hakkaraian, H., Sierra, A., Mangia, S., Garwood, M., Michaeli, S., Gröhn, O., Liimatainen, T., 2016. MRI relaxation in the presence of fictitious fields correlates with myelin content in normal rat brain. *Magn. Reson. Med.* 75 (1), 161–168.
- Hametner, S., Endmayr, V., Deistung, A., Palmrich, P., Prihoda, M., Haimburger, E., Menard, C., Feng, X., Haider, T., Leisser, M., Köck, U., Kaider, A., Höftberger, R., Robinson, S., Reichenbach, J.R., Lassmann, H., Traxler, H., Trattnig, S., Grabner, G., 2018. The influence of brain iron and myelin on magnetic susceptibility and effective transverse relaxation – a biochemical and histological validation study. *NeuroImage* 179 (June), 117–133.
- Hametner, S., Wimmer, L., Haider, L., Pfeifenbring, S., Brück, W., Lassmann, H., 2013. Iron and neurodegeneration in the multiple sclerosis brain. *Ann. Neurol.* 74 (6), 848–861.
- Harrer, M., Cuijpers, P., Furukawa, T., Ebert, D., 2019. Doing Meta-analysis in R: a Hands-on guide. PROTECT Lab Erlangen.
- Head, M.L., Holman, L., Lanfear, R., Kahn, A.T., Jennions, M.D., 2015. The extent and consequences of p-hacking in science. *PLoS Biol.* 13 (3), 1–15.
- Heath, F., Hurlley, S.A., Johansen-Berg, H., Sampaio-Baptista, C., 2018. Advances in non-invasive myelin imaging. *Dev. Neurobiol.* 78 (2), 136–151.
- Henkelman, R.M., Huang, X., Xiang, Q., Stanisz, G.J., Swanson, S.D., Bronskill, M.J., 1993. Quantitative interpretation of magnetization transfer. *Magn. Reson. Med.* 29 (6), 759–766.
- Hill, R.A., Li, A.M., Grutzendler, J., 2018. Lifelong cortical myelin plasticity and age-related degeneration in the live mammalian brain. *Nat. Neurosci.* 21 (May).
- Huszar, I., Pallegage-Gamarallage, M., Foxley, S., Tendler, B., Leonte, A., Hiemstra, M., Mollink, J., Smart, A., Bangerter-Christensen, S., Brooks, H., Turner, M., Ansgore, O., Miller, K., Jenkinson, M., 2019. Tensor image registration library: automated non-linear registration of sparsely sampled histological specimens to post-mortem MRI of the whole human brain. <https://www.biorxiv.org/content/10.1101/849570v1>.
- Janve, V.A., Zu, Z., Yao, S.Y., Li, K., Zhang, F.L., Wilson, K.J., Ou, X., Does, M.D., Subramaniam, S., Gochberg, D.F., 2013. The radial diffusivity and magnetization transfer pool size ratio are sensitive markers for demyelination in a rat model of type III multiple sclerosis (MS) lesions. *NeuroImage* 74, 298–305.
- Jelescu, I.O., Zurek, M., Winters, K.V., Veraart, J., Rajaratnam, A., Kim, N.S., Babb, J.S., Shepherd, T.M., Novikov, D.S., Kim, S.G., Fieremans, E., 2016. In vivo quantification of demyelination and recovery using compartment-specific diffusion MRI metrics validated by electron microscopy. *NeuroImage* 132, 104–114.
- Jensen, S.K., Yong, V.W., 2016. Activity-dependent and experience-driven myelination provide new directions for the management of multiple sclerosis. *Trends Neurosci.* 39(6), 356–365.
- Jespersen, S.N., Bjarkam, C.R., Nyengaard, J.R., Chakravarty, M.M., Hansen, B., Vosegaard, T., Østergaard, L., Yablonskiy, D., Nielsen, N.C., Vestergaard-Poulsen, P., 2010. Neurite density from magnetic resonance diffusion measurements at ultrahigh field: comparison with light microscopy and electron microscopy. *NeuroImage* 49 (1), 205–216.

- Jito, J., Nakasu, S., Ito, R., Fukami, T., Morikawa, S., Inubushi, T., 2008. Maturation changes in diffusion anisotropy in the rat corpus callosum: comparison with quantitative histological evaluation. *J. Magn. Reson. Imaging* 28 (4), 847–854.
- Johansen-Berg, H., 2010. Behavioural relevance of variation in white matter microstructure. *Curr. Opin. Neurol.* 23, 351–358.
- Kaden, E., Kelm, N.D., Carson, R.P., Does, M.D., Alexander, D.C., 2016. Multi-compartment microscopic diffusion imaging. *NeuroImage* 139, 346–359.
- Kelm, N.D., West, K.L., Carson, R.P., Gochberg, D.F., Ess, K.C., Does, M.D., 2016. Evaluation of diffusion kurtosis imaging in ex vivo hypomyelinated mouse brains. *NeuroImage* 124, 612–626.
- Khodanovich, M., Pishchelko, A., Glazacheva, V., Pan, E., Akulov, A., Svetlik, M., Tyumentseva, Y., Anan, T., 2019. Quantitative imaging of white and gray matter remyelination in the cuprizone demyelination model using the macromolecular proton fraction. *Cells* 8 (1204).
- Khodanovich, M.Y., Sorokina, I.V., Glazacheva, V.Y., Akulov, A.E., Nemirovich-Danchenko, N.M., Romashchenko, A.V., Tolstikova, T.G., Mustafina, L.R., Yarnykh, V.L., 2017. Histological validation of fast macromolecular proton fraction mapping as a quantitative myelin imaging method in the cuprizone demyelination model. *Sci. Rep.* 7 (March), 46686.
- Kiernan, J., 2007. Histochemistry of staining methods for normal and degenerating myelin in the central and peripheral nervous systems. *J. Histochem. J.* 39 (2), 87–106.
- Klüver, H., Barrera, E., 1953. A method for combined staining of cells and fibers in the nervous system. *J. Neuropathol. Exp. Neurol.* 12 (4), 400–403.
- Kolind, S.H., Laule, C., Vavasour, I.M., Li, D.K.B., Traboulsee, A.L., Mädler, B., Moore, G.R.W., Mackay, A.L., 2008. Complementary information from multi-exponential T2 relaxation and diffusion tensor imaging reveals differences between multiple sclerosis lesions. *NeuroImage* 40, 77–85.
- Kozlowski, P., Raj, D., Liu, J.I.E., Lam, C., Yung, A.C., Tetzlaff, W., 2008. Characterizing white matter damage in rat spinal cord with quantitative histology and histology. *J. Neurotrauma* 25, 653–676.
- Kozlowski, P., Rosicka, P., Liu, J., Yung, A.C., Tetzlaff, W., 2014. In vivo longitudinal myelin water imaging in rat spinal cord following dorsal column transection injury. *Magn. Reson. Imaging* 32 (3), 250–258.
- Kvarven, A., Ström, E., Johannesson, M., 2020. Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nat. Hum. Behav.* 4 (4), 423–434.
- Laule, C., Kozlowski, P., Leung, E., Li, D.K.B., Mackay, A.L., Moore, G.R.W., 2008. Myelin water imaging of multiple sclerosis at 7T: correlations with histopathology. *NeuroImage* 40, 1575–1580.
- Laule, C., Leung, E., Li, D.K.B., Traboulsee, A.L., Paty, D.W., Mackay, A.L., Moore, G.R.W., 2006. Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology. *Multiple Scler.* 12, 747–753.
- Laule, C., Vavasour, I.M., Leung, E., Li, D.K.B., Kozlowski, P., Traboulsee, A.L., Oger, J., Mackay, A.L., Moore, G.R.W., 2011. Pathological basis of diffusely abnormal white matter: insights from magnetic resonance imaging and histology. *Multiple Scler.* 17 (2), 144–150.
- Lee, Y., Morrison, B.M., Li, Y., Lengacher, S., Farah, M.H., Hoffman, P.N., Liu, Y., Tsingalia, A., Jin, L., Zhang, P.-w., Pellerin, L., Magistretti, P.J., Rothstein, J.D., 2012. Oligodendroglia metabolically support axons and contribute to neurodegeneration. *Nature* 487 (7408), 443–448.
- Lehto, L.J., Albers, A.A., Sierra, A., Tolppanen, L., Eberly, L.E., Mangia, S., Nurmi, A., Michaelis, S., Gröhn, O., 2017. Lysophosphatidyl choline induced demyelination in rat probed by relaxation along a fictitious field in high rank rotating frame. *Front. Neurosci.* 11, 1–14.
- Lerch, J.P., Gazdzinski, L., Germann, J., Sled, J.G., Henkelman, R.M., Nieman, B.J., 2012. Wanted dead or alive? The tradeoff between in-vivo versus ex-vivo MR brain imaging in the mouse. *Front. Neuroinform.* 6 (March), Article6.
- Leutritz, T., Samson, R.S., Curt, A., Helms, G., Freund, P., Weiskopf, N., 2020. Multiparameter mapping of relaxation (R1, R2*), proton density and magnetization transfer saturation at 3T: A multicenter dual-vendor reproducibility and repeatability study. *Hum. Brain Mapp.*
- Levesque, I.R., Pike, G.B., 2009. Characterizing healthy and diseased white matter using quantitative magnetization transfer and multicomponent T2 relaxometry: a unified view via a four-pool model. *Magn. Reson. Med.* 62 (6), 1487–1496.
- Lévy, S., Guertin, M.C., Khatibi, A., Mezer, A., Martinu, K., Chen, J.L., Stikov, N., Rainville, P., Cohen-Adad, J., 2018. Test-retest reliability of myelin imaging in the human spinal cord: Measurement errors versus region- and aging-induced variations. *PLoS One* 13 (1), 1–25.
- Lipp, I., Jones, D.K., Bells, S., Sgarlata, E., Foster, C., Stickland, R., Davidson, A.E., Tallantyre, E.C., Robertson, N.P., Wise, R.G., Tomassini, V., 2019. Comparing MRI metrics to quantify white matter microstructural damage in multiple sclerosis. *Hum. Brain Mapp.* 40 (10), 2917–2932.
- Liu, J., Dietz, K., Deloyht, J.M., Pedre, X., Kelkar, D., Kaur, J., Vialou, V., Lobo, M.K., Dietz, D.M., Nestler, E.J., Dupree, J., Casaccia, P., 2012. Impaired adult myelination in the prefrontal cortex of socially isolated mice. *Nat. Neurosci.* 15 (12), 1621–1623.
- Lodygensky, G.A., Marques, J.P., Maddage, R., Perroud, E., Sizonenko, S.V., Hüppi, P.S., Grueter, R., 2012. In vivo assessment of myelination by phase imaging at high magnetic field. *NeuroImage* 59 (3), 1979–1987.
- Lucchinetti, C.F., Popescu, B.F.G., Bunyan, R.F., Moll, N.M., Roemer, S.F., Lassmann, H., Brück, W., Parisi, J.E., Scheithauer, B.W., Giannini, C., Weigand, S.D., Mandrekar, J., Ransohoff, R.M., 2011. Inflammatory cortical demyelination in early multiple sclerosis. *N. Engl. J. Med.* 365 (23), 2188–2197.
- Lynn, J.D., Anand, C., Arshad, M., Homayouni, R., Rosenberg, D.R., Ofen, N., Raz, N., Stanley, J.A., 2020. Microstructure of human corpus callosum across the lifespan: regional variations in axon caliber, density, and myelin content. *Cereb. Cortex.*
- Mackay, A., Whittall, K., Adler, J., Li, D., Paty, D., Graeb, D., 1994. In vivo visualization of myelin water in brain by magnetic resonance. *Magn. Reson. Med.* 31 (6), 673–677.
- MacKay, A.L., Vavasour, I.M., Rauscher, A., 2009. MR relaxation in multiple sclerosis. *Neuroimaging Clin. N. Am.* 19, 1–26.
- Makinodan, M., Rosen, K.M., Ito, S., Corfas, G., 2012. A critical period for social experience-dependent oligodendrocyte maturation and myelination. *Science* 337 (6100), 1357–1360.
- Mancini, M., Karakuzu, A., Cohen-Adad, J., Cercignani, M., Nichols, T. E., Stikov, N., 2020. An interactive meta-analysis of MRI biomarkers of myelin. *ELife* 9, e61523.
- Mangeat, G., Govindarajan, S.T., Mainero, C., Cohen-Adad, J., 2015. Multivariate combination of magnetization transfer, T2* and B0 orientation to study the myelo-architecture of the in vivo human cortex. *NeuroImage* 119, 89–102.
- Martirosyan, N.L., Turner, G.H., Kaufman, J., Patel, A.A., Belykh, E., Kalani, M.Y.S., Theodore, N., Preul, M.C., 2016. Manganese-enhanced MRI offers correlation with severity of spinal cord injury in experimental models. *Open Neuroimaging J.* 10, 139–147.
- Mckenzie, I.A., Ohayon, D., Li, H., Al, E., 2014. Motor skill learning requires active central myelination. *Science* 346 (6207), 318–322.
- Mezer, A., Yeatman, J.D., Stikov, N., Kay, K.N., Cho, N.-J., Dougherty, R.F., Perry, M.L., Parvizi, J., Hua, L.H., Butts-Pauly, K., Wandell, B.A., 2013. Quantifying the local tissue volume and composition in individual brains with magnetic resonance imaging. *Nat. Med.* 19 (12), 1667–1672.
- Molineaux, S.M., Engh, H., de Ferra, F., Hudson, L., Lazzarini, R.A., 1986. Recombination within the myelin basic protein gene created the dysmyelinating shiverer mouse mutation. *Proc. Natl. Acad. Sci. USA* 83 (19), 7542–7546.
- Moll, N.M., Rietsch, A.M., Thomas, S., Ransohoff, A.J., Lee, J.-c., Fox, R., Chang, A., Ransohoff, R.M., Fisher, E., 2011. Multiple sclerosis normal-appearing white matter: pathology – imaging correlations. *Ann. Neurol.* 70, 764–773.
- Möller, H.E., Bossoni, L., Connor, J.R., Crichton, R.R., Does, M.D., Ward, R.J., Zecca, L., Zucca, F.A., Ronen, I., 2019. Iron, myelin, and the brain: neuroimaging meets neurobiology. *Trends Neurosci.* 42 (6), 384–401.
- Mollink, J., Hiemstra, M., Miller, K.L., Huszar, I.N., Jenkinson, M., Raaphorst, J., Wiesmann, M., Ansong, O., Pallegage-Gamarallage, M., van Cappellen van Walsum, A.M., 2019. White matter changes in the perforant path area in patients with amyotrophic lateral sclerosis. *Neuropathol. Appl. Neurobiol.* 45 (6), 570–585.
- 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, 1293–1301.
- Nichols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M.P., Poldrack, R.A., Poline, J.B., Proal, E., Thirion, B., Van Essen, D.C., White, T., Yeo, B.T., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* 20 (3), 299–303.
- Novikov, D., Fieremans, E., Jespersen, S., Kiselev, V., 2019. Quantifying brain microstructure with diffusion MRI: theory and parameter estimation. *NMR Biomed.* 32 (4), e3998.
- Oakden, W., Kwicien, J.M., Reilly, M.A.O., Dabrowski, W., Whyne, C., Finkelstein, J., Hynynen, K., Stanisz, G.J., 2015. Quantitative MRI in a non-surgical model of cervical spinal cord injury. *NMR Biomed.* 28, 925–936.
- O'Brien, J.S.O., Sampson, E.L., 1965. Lipid composition of the normal human brain: gray matter, white matter and myelin. *J. Lipid Res.* 6 (9), 537–544.
- Odrobina, E.E., Lam, T.Y., Pun, T., Midha, R., Stanisz, G.J., 2005. MR properties of excised neural tissue following experimentally induced demyelination. *NMR Biomed.* 18 (5), 277–284.
- Pan, S., Mayoral, S.R., Choi, H.S., Chan, J.R., Kheirbek, M.A., 2020. Preservation of a remote fear memory requires new myelin formation. *Nat. Neurosci.* 23 (4), 487–499.
- Pandur, E., Pap, R., Varga, E., Jánoska, G., Komoly, S., Fórizs, J., Sipos, K., 2019. Relationship of iron metabolism and short-term cuprizone treatment of c57bl/6 mice. *Int. J. Mol. Sci.* 20 (9).
- Peters, A., 2009. The effects of normal aging on myelinated nerve fibers in monkey central nervous system. *Front. Neuroanat.* 3 (JUL), 1–10.
- Peters, J.M., Struyven, R.R., Prohl, A.K., Vasung, L., Stajduhar, A., Taquet, M., Bushman, J.J., Lidov, H., Singh, J.M., Scherrer, B., Madsen, J.R., Prabhu, S.P., Sahin, M., Afacan, O., Warfield, S.K., 2019. White matter mean diffusivity correlates with myelination in tuberous sclerosis complex. *Ann. Clin. Transl. Neurol.* 6 (7), 1178–1190.
- Pichat, J., Eugenio, J., Yousry, T., Ourselin, S., Modat, M., 2018. A survey of methods for 3D histology reconstruction. *Med. Image Anal.* 46, 73–105.
- Pierpaoli, C., Chiro, D., Basser, J., Trace, D., 1996. Diffusion tensor MR imaging of the human brain. *Radiology* 201, 637–648.
- Piredda, G.F., Hilbert, T., Thiran, J.-P., Kober, T., 2020. Probing myelin content of the human brain with MRI: a review. *Magn. Reson. Med.*
- Pistorio, A.L., Hendry, S.H., Wang, X., 2006. A modified technique for high-resolution staining of myelin. *J. Neurosci. Methods* 153, 135–146.
- Pol, S., Sveinsson, M., Sudyn, M., Babek, N., Siebert, D., Bertolino, N., Modica, C.M., Preda, M., Schweser, F., Zivadinov, R., 2019. Teriflunomide's effect on glia in experimental demyelinating disease: a neuroimaging and histologic study. *J. Neuroimaging* 29 (1), 52–61.
- Praet, J., Manyakov, N.V., Muchene, L., Mai, Z., Terzopoulos, V., Backer, S.De, Verhoye, M., 2018. Diffusion kurtosis imaging allows the early detection and longitudinal follow-up of amyloid- β -induced pathology. *Alzheimer's Res. Ther.* 10 (1).
- Pun, T.W., Odrobina, E., Xu, Q.G., Lam, T.Y., Munro, C.A., Midha, R., Stanisz, G.J., 2005. Histological and magnetic resonance analysis of sciatic nerves in the tellurium model of neuropathy. *J. Peripheral Nervous Syst.* 10 (1), 38–46.
- Purger, D., Gibson, E.M., Monje, M., 2016. Myelin plasticity in the central nervous system. *Neuropharmacology* 110, 563–573.
- Reeves, C., Tachroum, M., Thomas, D., Michalak, Z., Liu, J., Ellis, M., Diehl, B., Miserocchi, A., Mcevoy, A.W., Eriksson, S., Yousry, T., Thom, M., 2015. Combined ex vivo

- 9.4T MRI and quantitative histopathological study in normal and pathological neocortical resections in focal epilepsy. *Brain Pathol.* 26, 319–333.
- Righart, R., Biberacher, V., Jonkman, L.E., Klaver, R., Schmidt, P., Buck, D., Berthele, A., Kirschke, J.S., Hemmer, B., Zimmer, C., 2017. Cortical pathology in multiple sclerosis detected by the T1/T2-weighted ratio from routine magnetic resonance imaging. *Ann. Neurol.* 82, 519–529.
- Rushton, W.A., 1951. A theory of the effects of fibre size in medullated nerve. *J. Physiol.* 115 (1), 101–122.
- Saab, A.S., Tzvetavona, I.D., Trevisiol, A., Baltan, S., Dibaj, P., Kusch, K., Möbius, W., Goetze, B., Jahn, H.M., Huang, W., Steffens, H., Schomburg, E.D., Pérez-Samartín, A., Pérez-Cerdá, F., Bakhtiari, D., Matute, C., Löwel, S., Griesinger, C., Hirrlinger, J., Kirchhoff, F., Nave, K.A., 2016. Oligodendroglial NMDA receptors regulate glucose import and axonal energy metabolism. *Neuron* 91 (1), 119–132.
- Salibian-Barrera, M., Zamar, R.H., 2002. Bootstrapping robust estimates of regression. *Ann. Stat.* 30 (2), 556–582.
- Salthouse, T.N., 1962. A quantitative histochemical method for estimating phospholipids. *Nature* 195 (4837), 187–188.
- Salthouse, T.N., 1962. Luxol fast blue: a new solvent azo dye with improved staining qualities for myelin and phospholipids. *Biotech. Histochem.* 37 (5), 313–316.
- Sampaio-Baptista, C., Khrapitchev, A.A., Foxley, S., Schlagheck, T., Scholz, J., Jbadi, S., Deluca, G.C., Miller, K.L., Taylor, A., Thomas, N., Kleim, J., Sibson, N.R., Bannerman, D., Johansen-berg, H., 2013. Motor skill learning induces changes in white matter microstructure and myelination. *The J. Neurosci.* 33 (50), 19499–19503.
- Schmierer, K., Parkes, H.G., So, P.-w., An, S.F., Brandner, S., Ordidge, R.J., Yousry, T.A., Miller, D.H., 2010. High field (9.4 Tesla) magnetic resonance imaging of cortical grey matter lesions in multiple sclerosis. *Brain* 133, 858–867.
- 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, 407–415.
- Schmierer, K., Tozer, D.J., Scaravilli, F., Altmann, D.R., Barker, G.J., Tofts, P.S., Miller, D.H., 2007. Quantitative magnetization transfer imaging in postmortem multiple sclerosis brain. *J. Magn. Reson. Imaging* 51, 41–51.
- Schmierer, K., Wheeler-Kingshott, C.A.M., Boulby, P.A., Scaravilli, F., Altmann, D.R., Barker, G.J., Tofts, P.S., Miller, D.H., 2007. Diffusion tensor imaging of post mortem multiple sclerosis brain. *NeuroImage* 35, 467–477.
- Schmierer, K., Wheeler-Kingshott, C.A.M., Tozer, D.J., Boulby, P.A., Parkes, H.G., Yousry, T.A., Scaravilli, F., Barker, G.J., Tofts, P.S., Miller, D.H., 2008. Quantitative magnetic resonance of postmortem multiple sclerosis brain before and after fixation. *Magn. Reson. Med.* 277, 268–277.
- Schmued, L., Bowyer, J., Cozart, M., Heard, D., Binienda, Z., Paule, M., 2008. Introducing Black-Gold II, a highly soluble gold phosphate complex with several unique advantages for the histochemical localization of myelin. *Brain Res.* 1229, 210–217.
- Schmued, L., Slikker, W., 1999. Black-gold: a simple, high-resolution histochemical label for normal and pathological myelin in brain tissue sections. *Brain Res.* 837 (1–2), 289–297.
- Schulze, R., 2005. Meta-analysis: a Comparison of Approaches. Hogrefe Publishing.
- Schwartz, E.D., Cooper, E.T., Fan, Y., Jawad, A.F., Chin, C.-I., Nissannov, J., Hackney, D.B., 2005. MRI diffusion coefficients in spinal cord correlate with axon morphometry. *Brain Imaging* 16 (1), 73–76.
- Schwarzer, G., et al., 2007. meta: an R package for meta-analysis.
- Seehaus, A., Roebroek, A., Bastiani, M., Fonseca, L., 2015. Histological validation of high-resolution DTI in human post mortem tissue. *Front. Neuroanat.* 9, 1–12.
- Seewann, A., Vrenken, H., Van der Valk, P., Blezer, E., Knol, D., Castelijns, J., Polman, C., Pouwels, P., Barkhof, F., Geurts, J., 2009. Diffusely abnormal white matter in chronic multiple sclerosis. *Arch. Neurol.* 66 (5), 601–609.
- Seifert, A.C., 2019. Formalin tissue fixation biases myelin – sensitive MRI. *MRM* 82, 1504–1517.
- Sele, M., Wernitznig, S., Lipovšek, S., Radulović, S., Haybaeck, J., Birk-Toelghofer, A.M., Wodlej, C., Kleinegger, F., Sygulla, S., Leoni, M., Ropele, S., Leitinger, G., 2019. Optimization of ultrastructural preservation of human brain for transmission electron microscopy after long post-mortem intervals. *Acta Neuropathol. Commun.* 7 (1), 144.
- Sergeant, C., Vesvres, M.H., Devès, G., Guillou, F., 2005. Calcium, potassium, iron, copper and zinc concentrations in the white and gray matter of the cerebellum and corpus callosum in brain of four genetic mouse strains. *Nucl. Instrum. Methods Phys. Res. Sect. B* 231 (1–4), 234–238.
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., Altman, D.G., Booth, A., Chan, A.W., Chang, S., Clifford, T., Dickersin, K., Egger, M., Götzsche, P.C., Grimshaw, J.M., Groves, T., Helfand, M., Higgins, J., Lasserson, T., Lau, J., Lohr, K., McGowan, J., Mulrow, C., Norton, M., Page, M., Sampson, M., Schünemann, H., Simer, I., Summerskill, W., Tetzlaff, J., Trikalinos, T.A., Tovey, D., Turner, L., Whitlock, E., 2015. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *BMJ* 349 (January), 1–25.
- Shatl, A., Uddin, N., Matsuda, K., Figley, C., 2018. Quantitative ex vivo MRI changes due to progressive formalin fixation in whole human brain specimens: longitudinal characterization of diffusion, relaxometry, and myelin water fraction measurements at 3T. *Front. Med.* 5, 1–15.
- Shea, B.J., Reeves, B.C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., Henry, D.A., 2017. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 358, 1–9.
- Shepherd, T.M., Thelwall, P.E., Stanisz, G.J., Blackband, S.J., 2009. Aldehyde fixative solutions alter the water relaxation and diffusion properties of nervous tissue. *Magn. Reson. Med.* 62 (1), 26–34.
- Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. p-Curve and effect size: correcting for publication bias using only significant results. *Perspect. Psychol. Sci.* 9 (6), 666–681.
- Sinclair, J.L., Fischl, M.J., Alexandrova, O., Heß, M., Grothe, B., Leibold, C., Kopp-Scheinflug, C., 2017. Sound-evoked activity influences myelination of brainstem axons in the trapezoid body. *J. Neurosci.* 37 (34), 8239–8255.
- Soni, N., Vegh, V., To, X.V., Mohamed, A.Z., Borges, K., 2020. Combined diffusion tensor imaging and quantitative susceptibility mapping discern discrete facets of white matter pathology post-injury in the rodent brain. *Front. Neurol.* 11, 1–16.
- Soustelle, L., Antal, M.C., Lamy, J., Rousseau, F., Armprach, J.-P., Loureiro de Sousa, P., 2019. Correlations of quantitative MRI metrics with myelin basic protein (MBP) staining in a murine model of demyelination. *NMR Biomed.* e4116.
- Stankoff, B., Joachim, J., Hartung, H.-P., Ku, P., 2016. Repair strategies for multiple sclerosis: challenges, achievements and perspectives. *Curr. Opin. Neurol.* 29 (3), 286–292.
- Stassart, R.M., Möbius, W., Nave, K.A., Edgar, J.M., 2018. The Axon-Myelin unit in development and degenerative disease. *Front. Neurosci.* 12 (JUL).
- Steadman, P.E., Xia, F., Ahmed, M., Mocle, A.J., Penning, A.R., Geraghty, A.C., Steenland, H.W., Monje, M., Josselyn, S.A., Frankland, P.W., 2020. Disruption of oligodendrogenesis impairs memory consolidation in adult mice. *Neuron* 105 (1), 150–164.e6.
- Stedehouder, J., Kushner, S.A., 2017. Myelination of parvalbumin interneurons: aparamonious locus of pathophysiological convergence in schizophrenia. *Mol. Psychiatry* 22 (1), 4–12.
- Sternberger, N.H., Itoyama, Y., Kies, M.W., Webster DeF., H., 1978. Myelin basic protein demonstrated immunocytochemically in oligodendroglia prior to myelin sheath formation. *PNAS* 75 (5), 2521–2524.
- Sterne, J.A., Sutton, A.J., Ioannidis, J.P., Terrin, N., Jones, D.R., Lau, J., Carpenter, J., Rücker, G., Harbord, R.M., Schmid, C.H., Tetzlaff, J., Deeks, J.J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D.G., Moher, D., Higgins, J.P., 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343 (7818), 1–8.
- Stikov, N., Boudreau, M., Levesque, I.R., Tardif, C.L., Barral, K., Pike, G.B., 2015. On the accuracy of T1 mapping: searching for common ground. *MRM* 522, 514–522.
- Stikov, N., Campbell, J.S.W., Stroh, T., Lavelee, M., Frey, S., Novak, J., Nuara, S., Ho, M.K., Bedell, B.J., Dougherty, R.F., Leppert, I.R., Boudreau, M., Narayanan, S., Duval, T., Cohen-Adad, J., Picard, P.A., Gasecka, A., Cote, D., Pike, G.B., 2015. Quantitative analysis of the myelin g-ratio from electron microscopy images of the macaque corpus callosum. *Data Brief* 4, 368–373.
- Stüber, C., Morawski, M., Schäfer, A., Labadie, C., Wähnert, M., Leuze, C., Streicher, M., Barapatre, N., Reimann, K., Geyer, S., Spemann, D., Turner, R., 2014. Myelin and iron concentration in the human brain: a quantitative study of MRI contrast. *NeuroImage* 93, 95–106.
- Sun, S.W., Neil, J.J., Liang, H.F., He, Y.Y., Schmidt, R.E., Hsu, C.Y., Song, S.K., 2005. Formalin fixation alters water diffusion coefficient magnitude but not anisotropy in infarcted brain. *Magn. Reson. Med.* 53 (6), 1447–1451.
- Sundberg, L.M., Herrera, J.J., Narayana, P.A., 2010. In vivo longitudinal MRI and behavioral lesions in experimental spinal cord injury. *J. Neurotrauma* 27, 1753–1767.
- Szucs, D., Ioannidis, J.P., 2020. Sample size evolution in neuroimaging research: an evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage* 221 (July), 117164.
- Takagi, T., Nakamura, M., Yamada, M., Hikishima, K., Momoshima, S., Fujiyoshi, K., Shibata, S., Okano, H.J., Toyama, Y., Okano, H., 2009. Visualization of peripheral nerve degeneration and regeneration: Monitoring with diffusion tensor tractography. *NeuroImage* 44 (3), 884–892.
- Thavarajah, R., Mudimbaimannar, V.K., Elizabeth, J., Rao, U.K., 2012. Chemical and physical basics of routine formaldehyde fixation. *JOMFP* 16 (3), 400–405.
- Thetiot, S., Freeman, Desmazieres, 2018. Immunohistochemical analysis of myelin structures. *Myelin Methods and Protocols.*
- Thiessen, J.D., Zhang, Y., Zhang, H., Wang, L., Buist, R., Del, M.R., Kong, J., Li, X.-m., Martin, M., 2013. Quantitative MRI and ultrastructural examination of the cuprizone mouse model of demyelination. *NMR Biomed.* 26 (11).
- van Tilborg, E., Achterberg, E.J., van Kammen, C.M., van der Toorn, A., Groenendaal, F., Dijkhuizen, R.M., Heijnen, C.J., Vanderschuren, L.J., Benders, M.N., Nijboer, C.H., 2018. Combined fetal inflammation and postnatal hypoxia causes myelin deficits and autism-like behavior in a rat model of diffuse white matter injury. *Glia* 66 (1), 78–93.
- Treit, S., Wheatley, B.M., Steve, T., 2019. Regional hippocampal diffusion abnormalities associated with subfield – specific pathology in temporal lobe epilepsy. *Epilepsia* 4, 544–554.
- Tu, T.W., Williams, R.A., Lescher, J.D., Jikaria, N., Turtzo, L.C., Frank, J.A., 2016. Radiological-pathological correlation of diffusion tensor and magnetization transfer imaging in a closed head traumatic brain injury model. *Ann. Neurol.* 79 (6), 907–920.
- Turati, L., Moscatelli, M., Mastropietro, A., Dowell, N.G., Zucca, I., Erbetta, A., Cordiglieri, C., Brenna, G., Bianchi, B., Mantegazza, R., Cercignani, M., Baggi, F., Minati, L., 2015. In vivo quantitative magnetization transfer imaging correlates with histology during de- and remyelination in cuprizone-treated mice. *NMR Biomed.* 28, 327–337.
- Underhill, H.R., Rostomily, R.C., Mikheev, A.M., Yuan, C., Yarnykh, V.L., 2011. Fast bound pool fraction imaging of the in vivo rat brain: association with myelin content and validation in the C6 glioma model. *NeuroImage* 54 (3), 2052–2065.
- Veraart, J., Fieremans, E., Novikov, D.S., 2019. On the scaling behavior of water diffusion in human brain white matter. *NeuroImage* 185 (2018), 379–387.
- Vincez, A., Mázl, M., Seress, L., Komoly, S., Ábrahám, H., 2008. A correlative light and electron microscopic study of postnatal myelination in the murine corpus callosum. *Int. J. Dev. Neurosci.* 26 (6), 575–584.
- van der Voorn, J., Pouwels, P., Powers, J., Kamphorst, W., Martin, J., Troost, D., Spreewegenberg, M.D., Barkhof, F., van den Knapp, M., 2011. Correlating quantitative MR imaging with histopathology in X-Linked adrenoleukodystrophy. *AJNR* 32, 481–489.
- Wang, S., Wu, E.X., Cai, K., Lau, H., 2009. Mild hypoxic-ischemic injury in the neonatal rat brain: longitudinal evaluation of white matter. *AJNR* 30, 1907–1913.

- Wang, Y., Sun, P., Wang, Q., Trinkaus, K., Schmidt, R.E., Naismith, R.T., Cross, A.H., Song, S.-k., 2015. Differentiation and quantification of inflammation, demyelination and axon injury or loss in multiple sclerosis. *Brain* 138, 1223–1238.
- Wang, Y., Wang, Q., Haldar, J.P., Yeh, F.C., Xie, M., Sun, P., Tu, T.W., Trinkaus, K., Klein, R.S., Cross, A.H., Song, S.K., 2011. Quantification of increased cellularity during inflammatory demyelination. *Brain* 134 (12), 3587–3598.
- Wartjes, J., Persson, A., Berge, J., Zeche, W., 2017. Myelin detection using rapid quantitative MR imaging correlated to macroscopically registered Luxol Fast Blue – stained brain specimens. *AJNR* 1096–1102.
- Waxman, S.G., 1980. Determinants of conduction velocity in myelinated nerve fibers. *Muscle Nerve* 3 (2), 141–150.
- Webb, S., Munro, C.A., Midha, R., Stanisz, G.J., 2003. Is multicomponent T2 a good measure of myelin content in peripheral nerve? *Magn. Reson. Med.* 49 (4), 638–645.
- Wei, P.T., Leong, D., Calabrese, E., White, L., Pierce, T., Platt, S., Provenzale, J., 2013. Diffusion tensor imaging of neural tissue organization: correlations between radiologic and histologic parameters. *Neuroradiol. J.* 26, 501–510.
- Weiskopf, N., Mohammadi, S., Lutti, A., 2015. Advances in MRI-based computational neuroanatomy: from morphometry to in-vivo histology. *Curr. Opin. Neurol.* 28, 313–322.
- Werner, M., Chott, A., Fabiano, A., Battifora, H., 2000. Effect of formalin tissue fixation and processing on immunohistochemistry. *Am. J. Surg. Pathol.* 24 (7), 1016–1019.
- West, K.L., Kelm, N.D., Carson, R.P., Gochberg, D.F., Ess, K.C., Does, M.D., 2018. Myelin volume fraction imaging with MRI. *NeuroImage* 182 (2016), 511–521.
- Wheeler, D., Venkata, V., Bandaru, R., Calabresi, P.A., Nath, A., Haughey, N.J., 2008. A defect of sphingolipid metabolism modifies the properties of normal appearing white matter in multiple sclerosis. *Brain* 131, 3092–3102.
- Wilcox, R.R., 2016. Comparing dependent robust correlations. *Br. J. Math. Stat. Psychol.* 69 (3), 215–224.
- Winkler, A.M., Webster, M.A., Brooks, J.C., Tracey, I., Smith, S.M., Nichols, T.E., 2016. Non-parametric combination and related permutation tests for neuroimaging. *Hum. Brain Mapp.* 37 (4), 1486–1511.
- de Winter, J.C., Gosling, S.D., Potter, J., 2016. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. *Psychol. Methods* 21 (3), 273–290.
- Wolff, S.D., Balaban, R.S., 1989. Magnetization transfer contrast (MTC) and tissue water proton relaxation in vivo. *Magn. Reson. Med.* 10 (1), 135–144.
- Woodhoo, A., 2018. *Myelin Methods and Protocols*. Humana Press.
- Xin, W., Chan, J.R., 2020. Myelin plasticity: sculpting circuits in learning and memory. *Nat. Rev. Neurosci.* 21 (12), 682–694.
- Yano, R., Hata, J., Abe, Y., Seki, F., Yoshida, K., Komaki, Y., Okano, H., Tanaka, K.F., 2018. Quantitative temporal changes in DTI values coupled with histological properties in cuprizone-induced demyelination and remyelination. *Neurochem. Int.* 119, 151–158.
- Yeatman, J.D., Wandell, B.A., Mezer, A.A., 2014. Lifespan maturation and degeneration of human brain white matter. *Nat. Commun.* 5, 1–12.
- Zaimi, A., Duval, T., Gasecka, A., Côté, D., Stikov, N., Cohen-Adad, J., 2016. AxonSeg: open source software for axon and myelin segmentation and morphometric analysis. *Front. Neuroinform.* 10, 1–13.
- Zatorre, R.J., Fields, R.D., Johansen-Berg, H., 2012. Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nat. Neurosci.* 15 (4), 528–536.
- Zhang, H., Schneider, T., Wheeler-kingshott, C.A., Alexander, D.C., 2012. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage* 61 (4), 1000–1016.
- Ziegler, G., Hauser, T.U., Moutoussis, M., Bullmore, E.T., Goodyer, I.M., Fonagy, P., Jones, P.B., Lindenberger, U., Dolan, R.J., 2019. Compulsivity and impulsivity traits linked to attenuated developmental frontostriatal myelination trajectories. *Nat. Neurosci.* 22 (6), 992–999.
- Zikopoulos, B., Barbas, H., 2010. Changes in prefrontal axons may disrupt the network in autism. *J. Neurosci.* 30 (44), 14595–14609.