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*

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 myelin histology 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.

Keywords: Myelin, MRI, validation, microstructural imaging, relaxometry, magnetization transfer, diffusion, histology

Highlights

- Systematic review and meta-analysis of studies validating microstructural imaging with myelin histology
- We find many MR markers are sensitive to myelin, including FA, RD, MP, MTR, Susceptibility, R1, but not MD
- Formal comparisons between MRI-based myelin markers are limited by methodological variability, inconsistent reporting and potential for publication bias
- Results emphasize the advantage of using multimodal imaging when testing hypotheses related to myelin 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; Waxman, 1980; Rushton, 1951). 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; Lee et al., 2012; Jensen and Yong, 2016). Changes in myelination occur as part of normal brain development (Gibson and Peterson, 1991; Ziegler et al., 2019) and aging (Peters, 2009; Hill et al., 2018). Moreover, recent studies have highlighted the possibility that myelination may change dynamically also in adulthood (Sampaio-Baptista et al., 2013), 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 neurological 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. The resolution of images acquired using a conventional MRI scanner lies in the millimetre range, whilst individual axons have diameters in the range of micrometres. Microstructural parameters can therefore only be estimated as summary measures 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 (Does, 2018; Edwards et al., 2018; Möller et al., 2019; Novikov et al., 2019). Macromolecules in the myelin sheath influence relaxation rates (as measured by the longitudinal relaxation rate R1, the transverse relaxation rate R2, the effective transverse relaxation rate R2*, and myelin water fraction MWF) as well as magnetization transfer (MT; as quantified by the magnetization transfer ratio MTR, and the macromolecular pool size MP). The structure of the myelin sheath hinders local water diffusion (as measured by diffusion-weighted imaging (DWI)) and diamagnetic myelin influences the local magnetic field strength (as measured by quantitative susceptibility mapping (QSM)) (Möller et al., 2019; Weiskopf et al., 2015). 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 (MacKay et al., 2009; Mezer et al., 2013; Alonso-Ortiz et al., 2015; Campbell et al., 2017; Heath et al., 2018; 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 (Levesque and Pike, 2009; Barta et al., 2015; Does, 2018) 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 (Cohen-Adad, 2018; Barros et al., 2019).

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 post-mortem[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. 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://lazaral.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 / relaxation rates or myelin water fraction based on its short relaxation rate; 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 (Pierpaoli et al., 1996; Beaulieu, 2014; De Santis et al., 2014) c) **MTI**: acquisitions applying off-resonance pulses to induce magnetization transfer effects and aiming to quantify them; 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 FOV 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 for what species was used, which part of the central nervous system was imaged (e.g. brain vs spinal cord, whole brain vs sections or smaller samples) and whether it came from healthy individuals or natural or induced pathologies.

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 if words like outlined or drawn or defined 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 ± sign, then this indicates mean and standard deviation across samples.

As temperature has an effect on MR relaxation times (Birkl 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). 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 Supplementary 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) statistical 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. 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.

Within included studies, separate meta-analyses were carried out for any MRI marker that was validated in more than two independent studies with more than three subjects. 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 corre-

lations 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 (Simonsohn et al., 2014; Harrer et al., 2019).

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 (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 *dmetar* 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/.

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); **lADC**: 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; **M0b**: 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 an 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
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, QSM	MD, Susceptibility	Mouse	perfusion fixed	LFB
Bagnato et al. (2018)	Relaxometry	R2*	Human	fixed	LFB and PLP
Beckmann et al. (2018)	MT	MTR	Mouse	in vivo	LFB
Bot et al. (2004)	Relaxometry, MT	T1, T2, MTR	Human	fixed	LFB
Chandran et al. (2012)	DWI	FA, RD	Mouse	in vivo	LFB, MBP
Chang et al. (2017b)	DWI	FA, AD, RD	Mouse	CLARITY fixation	MBP
Chang et al. (2017a)	DWI	FA, AD, RD, MD	Mouse	CLARITY fixation	MBP
Chen et al. (2017)	Relaxometry	MWF	Rat	perfusion fixed	EM
Choi et al. (2015)	DWI	FA, RD	Dog	fixed	Gold chloride
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
Grussu et al. (2017)	DWI	AD, FA, MD, RD	Human	fixed	PLP
Hakkarainen et al. (2016)	Relaxometry, MT	RAFFn, T1, T2, MTR	Rats	perfusion fixed	Gold chloride
Hametner et al. (2018)	Relaxometry, QSM	T1, R2*, Susceptibility	Human	in situ	LFB
Janve et al. (2013)	DWI, MT	AD, FA, RD, MP ('PSR')	Rat	perfusion fixed	LFB
Jelescu et al. (2016)	DWI, Relaxometry, MT	RD, RK, T2, MTR	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, Relaxometry, MT	FA, MD, RD, MK, AK, RK, MWF, 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
Kozlowski et al. (2008)	DWI, Relaxometry	FA, AD, MD, RD, MWF	Rat	perfusion fixed	LFB, MBP
Kozlowski et al. (2014)	DWI, Relaxometry	FA, MWF	Rats	in vivo (relaxometry) and fixed (DWI)	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

Continued on next page

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); **lADC**: 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; **M0b**: 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 an 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
Lehto et al. (2017)	DWI,Relaxometry, MT	AD, FA, MD, RD, RAFF4, T1sat, MTR	Rat	in vivo	Gold chloride
Lodygensky et al. (2012)	QSM	Susceptibility	Rat	in vivo	Black Gold II
Martirosyan et al. (2016)	DTI	FA	Rat	perfusion fixed	LFB
Moll et al. (2011)	DWI, MT	AD, FA, MD, RD, MTR	Human	in situ / fixed	MBP
Mollink et al. (2019)	DWI	FA, MD, RD, AD	Human	fixed	PLP
Mottershead et al. (2003)	DWI, Relaxometry, MT	ADC, SDI, PD, T1, T2, MTR	Human	fresh	LFB
Oakden et al. (2015)	DWI, Relaxometry	AD, RD, MWF	Rat	in vivo / fixed	LFB
Odrobina et al. (2005)	Relaxometry, MT	MWF, T1, T2, MTR, MP('M0b')	Rat	fixed	Toluidine blue
Peters et al. (2019)	DWI	FA, MD	Human	in vivo	LFB
Pol et al. (2019)	DWI, QSM	FA, MD, Susceptibility	Mouse	in vivo	Solochrome
Praet et al. (2018)	DWI	AD, AK, DT, DK, FA, MD, MK, RD, RK	Mouse	in vivo	MBP
Pun et al. (2005)	Relaxometry	T1, MWF	Rat	fresh	Toluidine blue
Reeves et al. (2015)	Relaxometry, MT	T1, T2, T2*, MTR	Human	fixed	MBP
Righart et al. (2017)	Other	T1w/T2w	Human	in situ	PLP
Schmierer et al. (2010)	Relaxometry, MT	T1, MTR	Human	fixed	MBP
Schmierer et al. (2004)	Relaxometry, MT	T1, MTR	Human	fresh	LFB
Schmierer et al. (2007a)	Relaxometry, MT	T1, MTR, MP('fb')	Human	fresh	LFB
Schmierer et al. (2007b)	DWI	FA, MD	Human	fresh	LFB
Schmierer et al. (2008)	DWI, Relaxometry, MT	AD, FA, MD, RD, T1, T2, MTR, MP('fb')	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, Relaxometry, MT	FA, ADC, T1, MTR	Human	fixed	LFB, PAS, PLP
Soni et al. (2020)	QSM	Susceptibility	Mouse	in vivo	MBP
Soustelle et al. (2019)	DWI, Relaxometry, MT, Other	RD, MWF, MP('f'), rSPF	Mouse	perfusion fixed	MBP
Stüber et al. (2014)	Relaxometry, QSM	R1, R2*, Susceptibility	Human	fixed	PIXE
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, Relaxometry, MT	AD, FA, RD, T1 MP('f'), MTR	Mouse	perfusion fixed	EM
Tu et al. (2016)	DWI	AD, FA, MD, RD, MTR at different ppm	Rat	in vivo	MBP
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)	DWI, MT	ADC, FA, MTR	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

Continued on next page

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); **lADC**: 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; **M0b**: 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 an 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
Wang et al. (2015)	DWI	DBSI-RD	Human	fixed	LFB
Warntjes et al. (2017)	Relaxometry	MWF, R1	Human	in situ	LFB
Wei et al. (2013)	DWI	FA, RD	Dog	perfusion fixed	Gold chloride
West et al. (2018)	Relaxometry, MT	MWF, MP ('BPF')	Mouse	perfusion fixed	EM
Yano et al. (2018)	DWI	FA, MD, RD	Mouse	perfusion fixed	PLP

3. Results

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

Our search yielded 385 unique articles (Figure 1). Of these articles, 294 were from Pubmed, 46 from Scopus, and 45 from our expert library. Abstract screening led to exclusion of 2532 studies, while full-text screening further excluded 62 studies, with a total of 70 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, 2, 3, 4, 5. To have a better overview of the literature, we provide quantitative summaries of our findings on the literature content (Figure 2). We find that validation studies have a median sample size of 13 (Figure 2A), 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, Figure 2F). It also employs a wide variety of species and field strengths used for experiments (Figure 2D). 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 (Figure 2F). We also note that many markers appear only once in the literature (e.g. RAFF4, ihMT), suggesting there are many more markers in the literature that lack thorough validation. We do not include these markers in Figure 2F 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 (Figure 2C).

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). To ensure comparability between studies included, separate meta-analyses were run for studies using between-subject and within-subject variance (Figure 3 and Supplementary 5 respectively). Moreover, studies using Pearson and Spearman coefficients were also pooled separately (meta-analyses are labelled Pearson or Spearman within Fig.3). On the included studies, we ran fixed-effects meta-analyses to estimate a cross-study validation effect size for each marker (Figure 3). 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.

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 Figure 2). 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 Figure 3). 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 Figure 4) based on (Duval and Tweedie, 2000) and find that if present, publication bias would inflate the validation effect size of some markers.

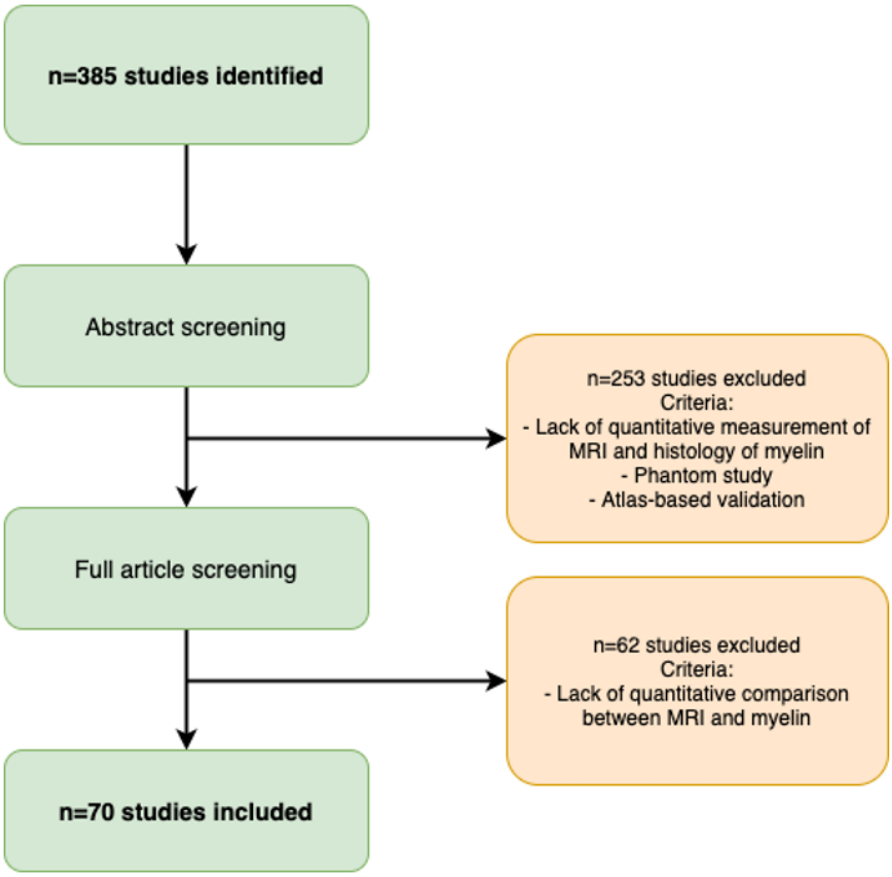

Figure 1: **Flow diagram of study selection.**

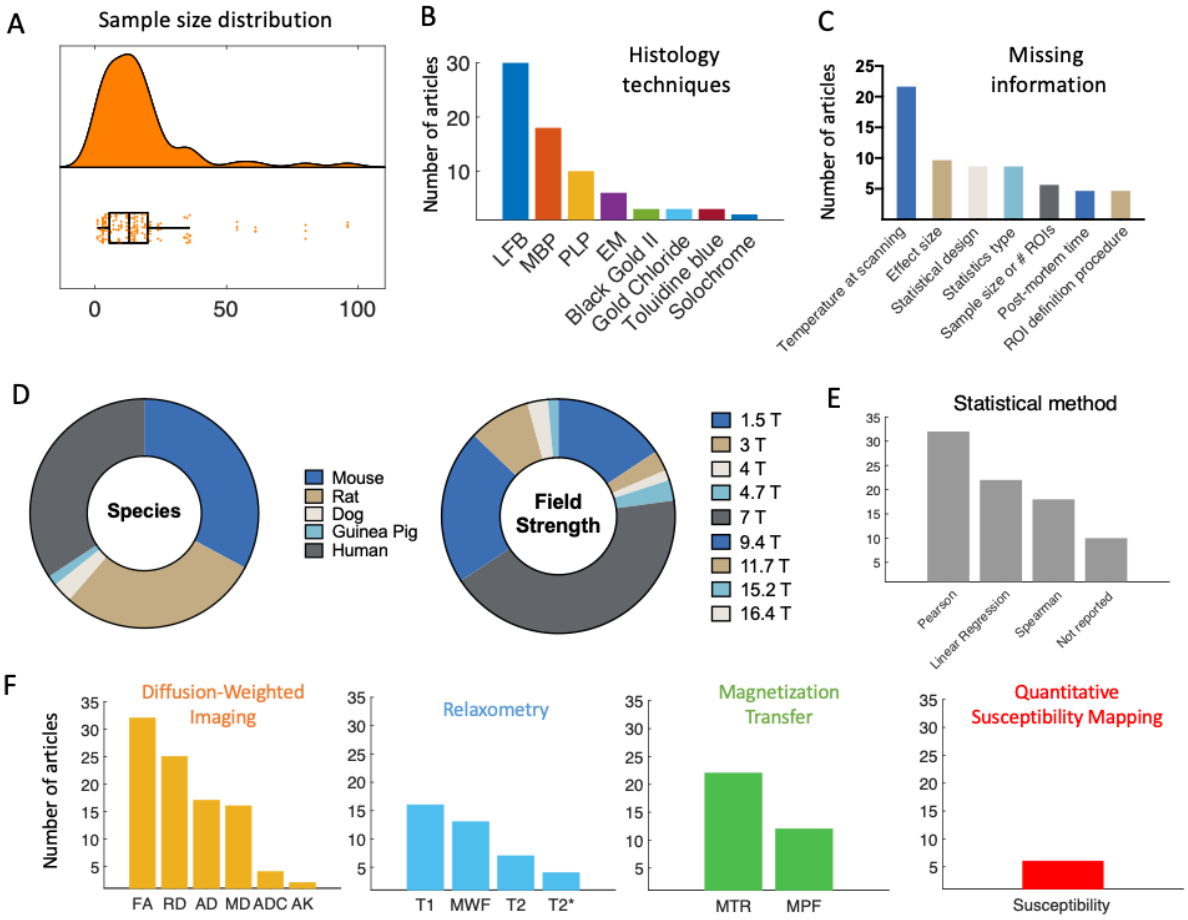

Figure 2: **Features of the MR-histology validation literature. [A]** Sample size distribution of the included papers (median = 13 subjects). **[B]** The most frequently used histology techniques used in the articles are shown in decreasing order. **[C]** Due to lack of commonly agreed reporting standards, 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.

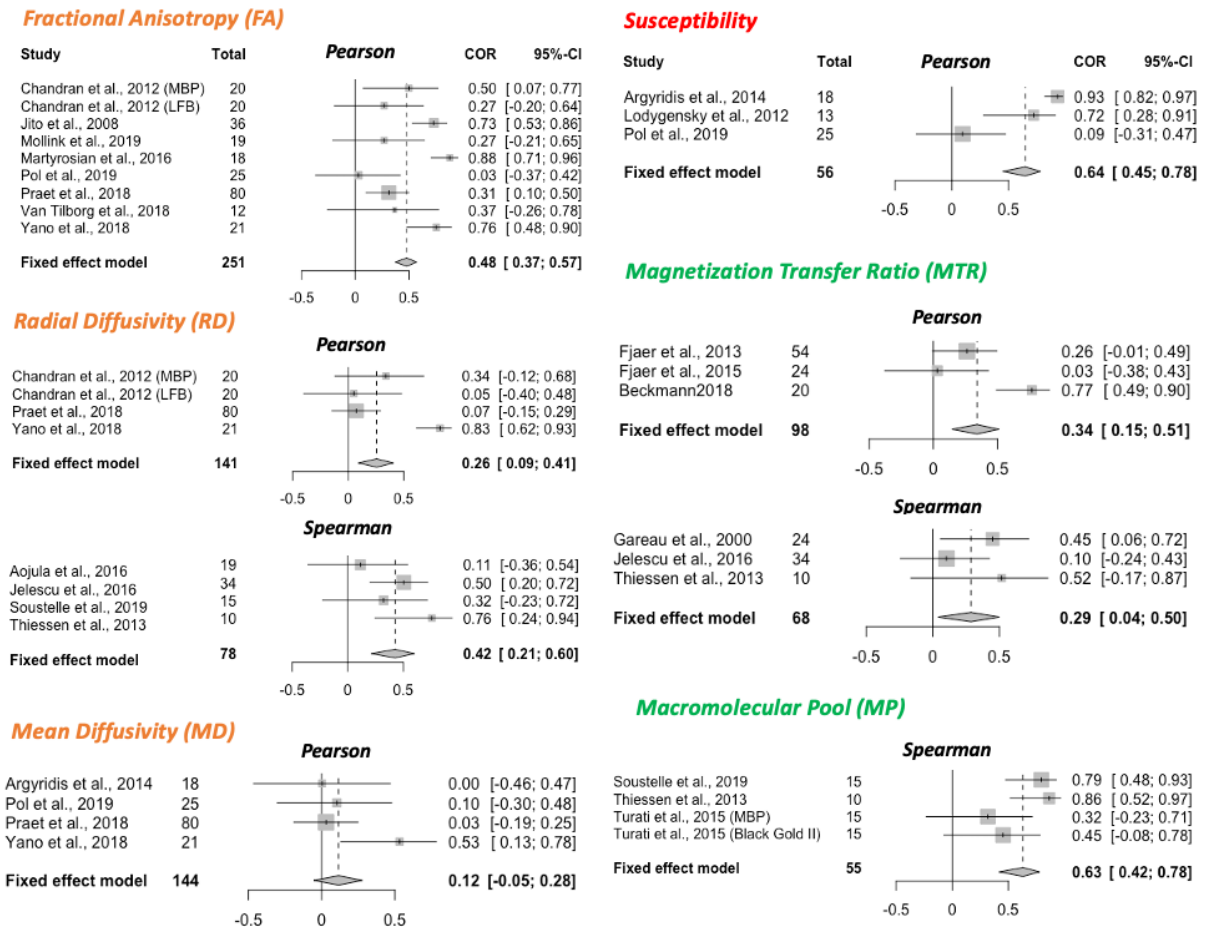

Figure 3: **Meta-analysis of correlations between histology and MR markers.** Forest plots from meta-analyses of 6 MR markers (FA, RD, MD, susceptibility, MTR and MP). Spearman and Pearson correlations are pooled separately; the type of correlation coefficient is indicated at the top of each forest plot. All studies included use a between-subject statistical design; meta-analyses of studies with a within-subject design are reported in Supplementary Figure 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.

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 Take validation parameters into account when selecting an MR marker for your study: 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

- 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 histologys 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 pool 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 70 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-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.

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)**. 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 (Lipp et al., 2019; Eichert et al., 2020).

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) and neurite orientation dispersion and density imaging (NODDI; Zhang et al. (2012), 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 13 studies in total examine MWF, a high number of these studies use a mixed within- and between- subject design, or do not report the type of static measure used, which makes them difficult to aggregate in a meta-analysis. Therefore, further validation studies for MWF, and especially studies examining between-subject variance, may be needed.

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. Relaxometry metrics such as R1 are generally 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).

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 (e.g. Birkl 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 WM (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). 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).

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 (Kiernan, 2007; Carriel et al., 2017b). 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. Kozlowski 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 to a lesser extent also by phospholipids (Kiernan, 2007; Klüver and Barrera, 1953; de Almeida and Pearse, 1958), 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 and Slikker, 1999; Schmued et al., 2008). 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, 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 re-

ported 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 GM and WM 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 (Figure 2, Supplementary 5), where the correlation coefficients were lower in between-subject studies (95% CI for R^2: .15 and .51) compared to within-subject studies (95% CI for R^2: .60 and .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 (Wilcox, 2016; Salibian-Barrera and Zamar, 2002) **(Recommendation 3.2)**.

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 =.78, and a correlation between FA and cell density of r = -.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 70 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.

Acknowledgements

We thank Aman Badhwar for feedback on the meta-analysis. We thank Alex Bates, Michiel Cottar, Nicole Eichert, Heidi Johansen-Berg, Maria Morozova, Ruairi Roberts, Zeena 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).

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. Neurochemistry International 125, 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 (Online) 344, 1–10.

Alexander, D.C., Dyrby, T.B., Nilsson, M., 2019. Imaging brain microstructure with diffusion MRI: practicality and applications. NMR in Biomedicine 32, e3841.

de Almeida, D.F., Pearse, A.G., 1958. Comparative histochemistry of lipids in relation to myelination in rabbit brain. Journal of Neurochemistry 3, 132–138.

Alonso-Ortiz, E., Levesque, I.R., Pike, G.B., 2015. MRI-based myelin water imaging: a technical review. Magnetic Resonance in Medicine 73, 70–81.

Altman, D., Bland, J., 1983. Measurement in medicine: The analysis of method comparison studies. Journal of the Royal Statistical Society. Series D (The Statistician) 32, 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 and Barriers of the CNS 13.

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. Human Brain Mapping 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, 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., Malyarenko, 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. Magnetic Resonance in Medicine 79, 2564–2575.

Barros, A.D., Arribarat, G., Combis, J., Chaynes, P., 2019. Matching ex vivo MRI with iron histology: Pearls and pitfalls. Frontiers in Neuroanatomy 13, Article 68.

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. Journal of Magnetic Resonance 259, 56–67.

Basser, P.J., 2004. Scaling laws for myelinated axons derived from an electrotonic core-conductor model. Journal of Integrative Neuroscience 3, 227–244.

Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system - A technical review. NMR in Biomedicine 15, 435–455. arXiv:15334406.

Beaulieu, C., 2014. The biological basis of diffusion anisotropy, in: Johansen-Berg, H., Behrens, T. (Eds.), Diffusion MRI. 2 ed., pp. 155–184.

Beckmann, N., Giorgetti, E., Neuhaus, A., Zurbruegg, 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 neuropathologica communications 6, 9.

Birkl, 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 in Biomedicine 29, 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 MR 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. Nature reviews. Neuroscience 14, 365–76.

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.

Carriel, V., Campos, A., Alaminos, M., Raimondo, S., Geuna, S., 2017a. Staining methods for normal and regenerative myelin in the nervous system, in: Histochemistry of Single Molecules. volume 1560, pp. 219–229.

Carriel, V., Campos, F., Aneiros-Fernandez, J., Kiernan, J., 2017b. Tissue fixation and processing for the histological identification of lipids, in: Histochemistry of Single Molecules. volume 1560, pp. 219–229.

Caspers, S., Axer, M., 2019. Decoding the microstructural correlate of diffusion MRI. NMR in Biomedicine 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., 2017a. Diffusion tensor imaging measures of white matter compared to myelin basic protein immunofluorescence in tissue cleared intact brains. Data in Brief 10, 438–443.

Chang, E.H., Argyelan, M., Aggarwal, M., Chandon, T.S.S., Karlsgodt, K.H., Mori, S., Malhotra, A.K., 2017b. 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., Kozlowski, 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 imagining of canines with mucopolysaccharidosis type I. The Neuroradiology Journal 28, 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. Postmortem MRI of human brain hemispheres: T2 relaxation times during formaldehyde fixation. Magnetic Resonance in Medicine 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., Hertanu, 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 postmortem brain MR. Magnetic Resonance in Medicine 81, 2688–2701.

Duval, S., Tweedie, R., 2000. Trim and Fill: A Simple Funnel-Plot-Based Method. Biometrics 56, 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. British Medical Journal 315, 629–634.

Eichert, N., Papp, D., Mars, R.B., Watkins, K.E., 2020. Mapping human laryngeal motor cortex during vocalization. biorxive doi:10.1093/cercor/bhaa182.

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. Journal of Cerebral Blood Flow and Metabolism 31, 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, 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. Neurochemistry International 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. Proceedings of the National Academy of Sciences of the United States of America 107, 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, 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. Journal of Magnetic Resonance Imaging 11, 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. Journal of Neuroscience 31, 11597–11616.

Goldman, L., Albus, J.S., 1968. Computation of impulse conduction in myelinated fibers; theoretical basis of the velocity-diameter relation. Biophysical Journal 8, 596–607.

González de San Román, E., Bidmon, H.J., Malisic, M., Susnea, I., Küppers, A., Hübbers, 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 Structure and Function 223, 2767–2783.

Goodwin, L.D., Leech, N.L., 2006. Understanding correlation: Factors that affect the size of r. Journal of Experimental Education 74, 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. Journal of Neuroscience Research 334, 321–334.

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? Annals of Clinical and Translational Neurology 4, 663–679. doi:10.1002/acn3.445.

Hakkarainen, 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. Magnetic Resonance in Medicine 75, 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, 117–133.

Hametner, S., Wimmer, I., Haider, L., Pfeifenbring, S., Brück, W., Lassmann, H., 2013. Iron and neurodegeneration in the multiple sclerosis brain. Annals of Neurology 74, 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 Biology 13, 1–15.

Heath, F., Hurley, S.A., Johansen-Berg, H., Sampaio-Baptista, C., 2018. Advances in noninvasive myelin imaging. Developmental Neurobiology 78, 136–151.

Hill, R.A., Li, A.M., Grutzendler, J., 2018. Lifelong cortical myelin plasticity and age-related degeneration in the live mammalian brain. Nature Neuroscience 21.

Huszar, I., Pallebage-Gamarallage, M., Foxley, S., Tendler, B., Leonte, A., Hiemstra, M., Mollink, J., Smart, A., Bangerter-Christensen, S., Brooks, H., Turner, M., Ansorge, 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. bioRxiv .

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 in Neurosciences 39(6), 356–65.

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, 205–216.

Jito, J., Nakasu, S., Ito, R., Fukami, T., Morikawa, S., Inubushi, T., 2008. Maturational changes in diffusion anisotropy in the rat corpus callosum: Comparison with quantitative histological evaluation. Journal of Magnetic Resonance Imaging 28, 847–854.

Johansen-Berg, H., 2010. Behavioural relevance of variation in white matter microstructure. Current Opinion in Neurology 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.

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. Scientific Reports 7, 46686.

Kiernan, J., 2007. Histochemistry of staining methods for normal and degenerating myelin in the central and peripheral nervous systems. Journal of Histotechnology 30, 87–106.

Klüver, H., Barrera, E., 1953. A method for combined staining of cells and fibers in the nervous system. J Neuropath Exp Neurol 12, 400–403.

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. Journal of 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. Magnetic Resonance Imaging 32, 250–258.

Kvarven, A., Strømland, E., Johannesson, M., 2020. Comparing meta-analyses and preregistered multiple-laboratory replication projects. Nature Human Behaviour 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 Sclerosis 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 Sclerosis Journal 17, 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, 443–448.

Lehto, L.J., Albors, A.A., Sierra, A., Tolppanen, L., Eberly, L.E., Mangia, S., Nurmi, A., Michaeli, S., Gröhn, O., 2017. Lysophosphatidyl choline induced demyelination in rat probed by relaxation along a fictitious field in high rank rotating frame. Frontiers in Neuroscience 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. Frontiers in Neuroinformatics 6, Article 6.

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. Human Brain Mapping .

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. Magnetic Resonance in Medicine 62, 1487–1496.

Lévy, S., Guertin, M.C., Khatibi, A., Mezer, A., Martinu, K., Chen, J.I., 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, 1–25.

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. Human Brain Mapping 40, 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. Nature Neuroscience 15, 1621–1623.

Lodygensky, G.A., Marques, J.P., Maddage, R., Perroud, E., Sizonenko, S.V., Hüppi, P.S., Gruetter, R., 2012. In vivo assessment of myelination by phase imaging at high magnetic field. NeuroImage 59, 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. The New England Journal of Medicine 365, 2188–97.

MacKay, A.L., Vavasour, I.M., Rauscher, A., 2009. MR Relaxation in Multiple Sclerosis. Neuroimag 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, 1357–1360.

Mancini, M., Karakuzu, A., Nichols, T., Cohen-Adad, J., Cercignani, M., Stikov, N., 2020. The quest for measuring myelin with MRI An interactive meta-analysis of quantitative comparisons with histology. bioRxiv .

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. The Open NeuroImaging Journal 10, 139–147.

Mckenzie, I.A., Ohayon, D., Li, H., Al., E., 2014. Motor skill learning requires active central myelination. Science 346, 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. Nature Medicine 19, 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. Proceedings of the National Academy of Sciences of the United States of America 83, 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 in Neurosciences 42, 384–401.

Mollink, J., Hiemstra, M., Miller, K.L., Huszar, I.N., Jenkinson, M., Raaphorst, J., Wiesmann, M., Ansorge, O., Pallebage-Gamarallage, M., van Cappellen van Walsum, A.M., 2019. White matter changes in the perforant path area in patients with amyotrophic lateral sclerosis. Neuropathology and Applied Neurobiology 45, 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. Nature Neuroscience 20, 299–303.

Novikov, D., Fieremans, E., Jespersen, S., Kiselev, V., 2019. Quantifying brain microstructure with diffusion MRI: Theory and parameter estimation. NMR in Biomedicine 32, e3998.

Oakden, W., Kwiecien, 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 in Biomedicine 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. Journal of Lipid Research 6, 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 in Biomedicine 18, 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. Nature Neuroscience 23, 487–499.

Pandur, E., Pap, R., Varga, E., Jánosa, G., Komoly, S., Fórizs, J., Sipos, K., 2019. Relationship of iron metabolism and short-term cuprizone treatment of c57bl/6 mice. International Journal of Molecular Sciences 20.

Peters, A., 2009. The effects of normal aging on myelinated nerve fibers in monkey central nervous system. Frontiers in Neuroanatomy 3, 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. Annals of Clinical and Translational Neurology 6, 1178–1190.

Pichat, J., Eugenio, J., Yousry, T., Ourselin, S., Modat, M., 2018. A survey of methods for 3D histology reconstruction. Medical Image Analysis 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. Magnetic Resonance in Medicine doi:10.1002/mrm.28509.

Pistorio, A.L., Hendry, S.H., Wang, X., 2006. A modified technique for high-resolution staining of myelin. Journal of Neuroscience 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. Journal of Neuroimaging 29, 52–61.

Praet, J., Manyakov, N.V., Muchene, L., Mai, Z., Terzopoulos, V., Backer, S.D., Torremans, A., Guns, P.j., Casteele, T.V.D., Bottelbergs, A., Broeck, B.V., Sijbers, J., Smeets, D., Shkedy, Z., Bijnens, L., Pemberton, D.J., Schmidt, M.E., Linden, A.V.D., Verhoye, M., 2018. Diffusion kurtosis imaging allows the early detection and longitudinal follow-up of amyloid- β -induced pathology. Alzheimer's Research & Therapy 10.

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. Journal of the Peripheral Nervous System 10, 38–46.

Purger, D., Gibson, E.M., Monje, M., 2016. Myelin plasticity in the central nervous system. Neuropharmacology 110, 563–573.

Reeves, C., Tachrount, 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 Pathology 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. The Journal of Physiology 115, 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, 119–132.

Salibian-Barrera, M., Zamar, R.H., 2002. Bootstrapping robust estimates of regression. Annals of Statistics 30, 556–582.

Sampaio-Baptista, C., Khrapitchev, A.A., Foxley, S., Schlagheck, T., Scholz, J., Jbabdi, 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 Journal of Neuroscience 33, 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., 2007a. Quantitative magnetization transfer imaging in postmortem multiple sclerosis brain. Journal of Magnetic Resonance 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., 2007b. 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. Magnetic Resonance in Medicine 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 Research 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 Research 837, 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.l., Nissanov, J., Hackney, D.B., 2005. MRI diffusion coefficients in spinal cord correlate with axon morphometry. Brain Imaging 16, 73–76.

Schwarzer, G., et al., 2007. meta: An r package for meta-analysis .

Seehaus, A., Roebroeck, A., Bastiani, M., Fonseca, L., 2015. Histological validation of high-resolution DTI in human post mortem tissue. Frontiers in Neuroanatomy 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, 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., Birkl-Toeglhofer, 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 neuropathologica communications 7, 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. Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms 231, 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., Gøtzsche, 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., Simera, 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 (Online) 349, 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. Frontiers in Medicine 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 (Online) 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. Magnetic Resonance in Medicine 62, 26–34.

Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. p-Curve and effect size: correcting for publication bias using only significant results. Perspectives on Psychological Science 9, 666–681.

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. Frontiers in Neurology 11, 1–16.

Soustelle, L., Antal, M.C., Lamy, J., Rousseau, F., Armspach, 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 in Biomedicine , e4116.

Stankoff, B., Joachim, J., Hartung, H.P., Ku, P., 2016. Repair strategies for multiple sclerosis: challenges, achievements and perspectives. Curr Opin Neurol 29, 286–292.

Stassart, R.M., Möbius, W., Nave, K.A., Edgar, J.M., 2018. The Axon-Myelin unit in development and degenerative disease. Frontiers in Neuroscience 12.

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, 150–164.e6.

Stedehouder, J., Kushner, S.A., 2017. Myelination of parvalbumin interneurons: A parsimonious locus of pathophysiological convergence in schizophrenia. Molecular Psychiatry 22, 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, 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, 1–8.

Stikov, N., Boudreau, M., Levesque, I.R., Tardif, C.L., Barral, K., Pike, G.B., 2015a. 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., Novek, 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., 2015b. Quantitative analysis of the myelin g-ratio from electron microscopy images of the macaque corpus callosum. Data in 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.

Sundberg, L.M., Herrera, J.J., Narayana, P.A., 2010. In vivo longitudinal MRI and behavioral studies in experimental spinal cord injury. Journal of Neurotrauma 27, 1753–1767.

Szucs, D., Ioannidis, J.P., 2020. Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (19902012) and of latest practices (20172018) in high-impact journals. NeuroImage 221, 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, 884–892.

Thavarajah, R., Mudimbaimannar, V.K., Elizabeth, J., Rao, U.K., 2012. Chemical and physical basics of routine formaldehyde fixation. JOMFP 16, 400–405.

Thetiot, S., Freeman, DesmazieresA, 2018. Immunohistochemical Analysis of Myelin Structures, in: 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 in Biomedicine 26.

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, 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. Annals of Neurology 79, 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 in Biomedicine 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, 2052–2065.

Veraart, J., Fieremans, E., Novikov, D.S., 2019. On the scaling behavior of water diffusion in human brain white matter. NeuroImage 185, 379–387.

Vincze, A., Mázló, M., Seress, L., Komoly, S., Ábrahám, H., 2008. A correlative light and electron microscopic

study of postnatal myelination in the murine corpus callosum. International Journal of Developmental Neuroscience 26, 575–584.

van der Voorn, J., Pouwels, P., Powers, J., Kamphorst, W., Martin, J., Troost, D., Spreeuwenberg, 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.

Warntjes, 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, 141–150.

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. The Neuroradiology Journal 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. American Journal of Surgical Pathology 24, 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, 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. British Journal of Mathematical and Statistical Psychology 69, 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. Human Brain Mapping 37, 1486–1511.

Woodhoo, A., 2018. Myelin Methods and Protocols. Humana Press.

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. Neurochemistry International 119, 151–158.

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. Frontiers in Neuroinformatics 10, 1–13.

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, 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. Nature Neuroscience 22, 992–999.

Zikopoulos, B., Barbas, H., 2010. Changes in prefrontal axons may disrupt the network in autism. Journal of Neuroscience 30, 14595–14609.

Supplementary Figures and Tables

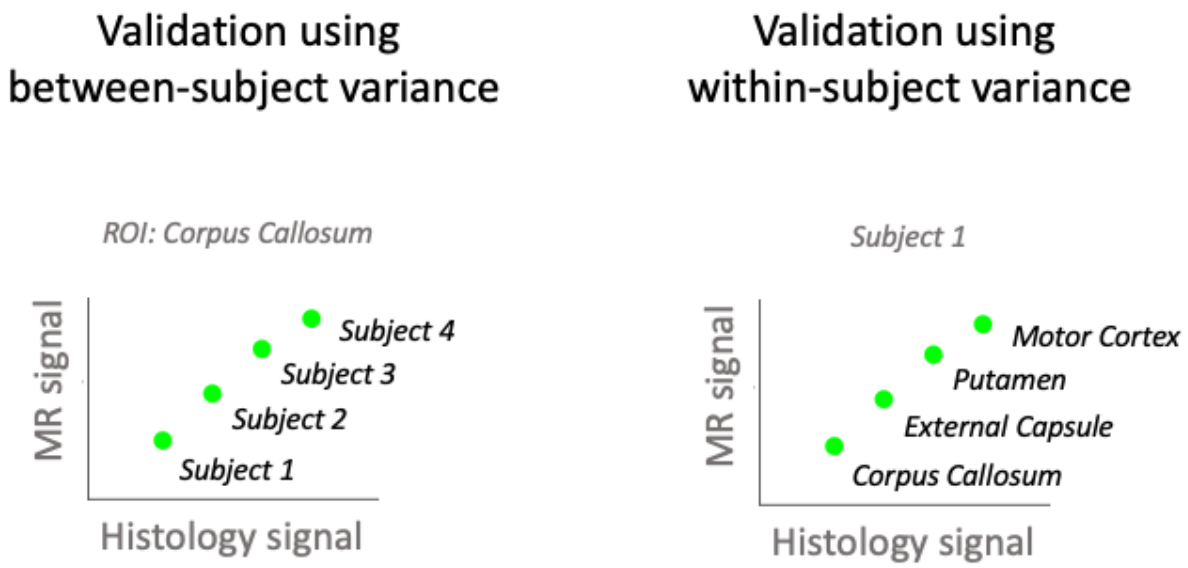

Figure 1: **Mock examples of between-subject vs within-subject design in validation studies.** In between-subject validation studies (left), each data point comes from a different subject and correlations are computed separately for each ROI. In within-subject validation studies (right), each data point 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.

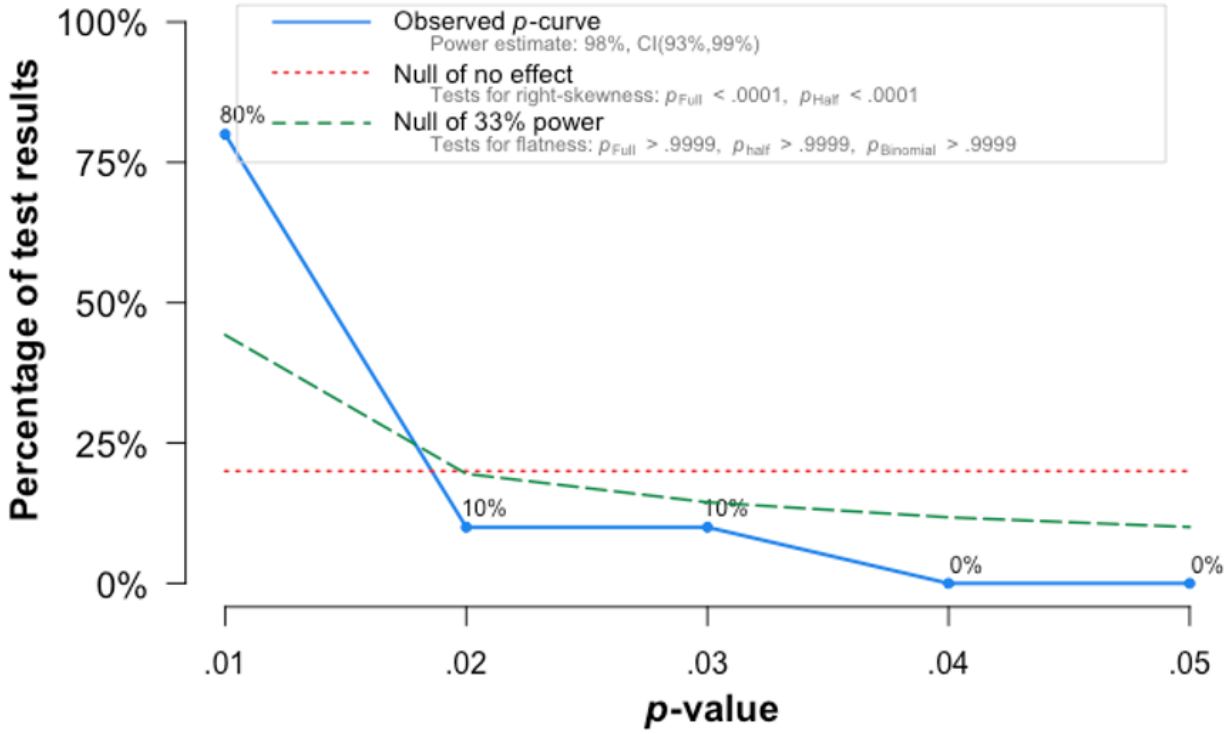

Figure 2: **p-hacking in the validation literature.** In fields where pressure to find significance biases published results p-value distributions tend to be right-skewed, i.e. to be characterized by p-values that are just below the 0.05 significance-threshold. We find evidence that the p-value distribution in the validation literature is, by contrast, left-skewed, with a majority of studies reporting p-values below 0.01.

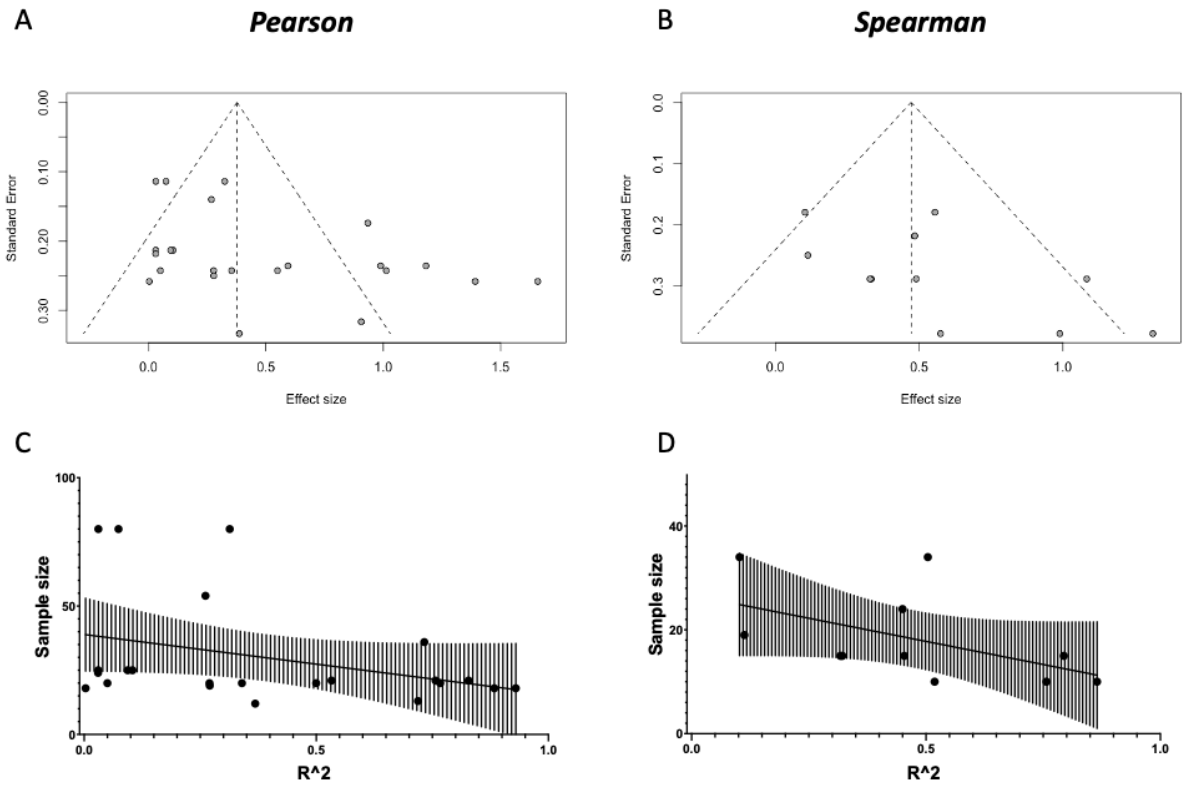

Figure 3: **Publication bias. [A and B]** Funnel plots of the studies in our quantitative meta-analyses (for Pearson and Spearman meta-analyses, respectively). We find some evidence for publication bias against low-sample size, low-effect articles (Eggers test: $t = 2.419$, $p = 0.02471$ for Pearson; $t = 2.053$, $p = 0.07025$ for Spearman). **[C and D]** To further characterize the relationship between Sample Size and Effect Size, we perform a correlation between the two, and find a relationship between higher sample sizes and lower effect sizes (Pearson: $r = -0.3759$, $p = 0.0771$; Spearman: $r = -0.6451$, $p = 0.0365$).

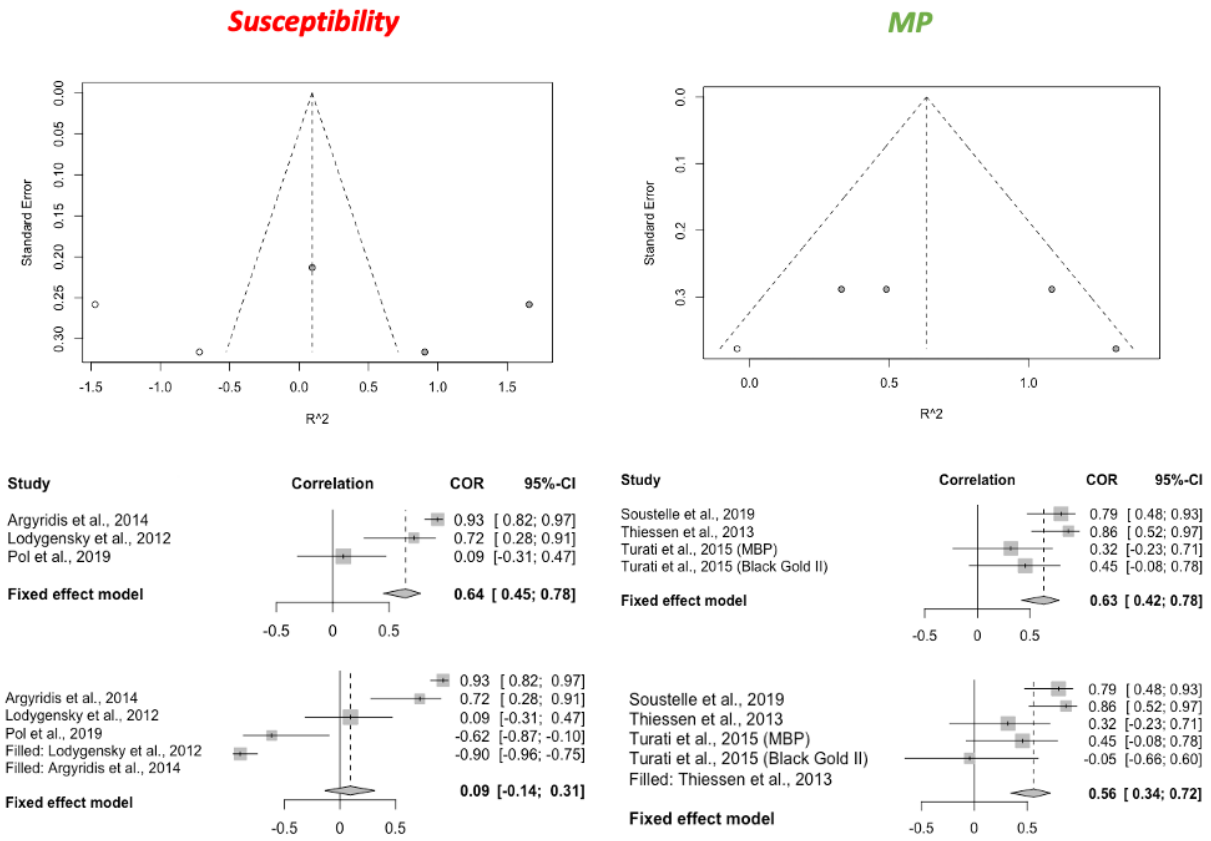

Figure 4: **Trim-and-fill correction for publication bias.** For both susceptibility and MP, we find evidence of publication bias (asymmetric funnel plots, top figures). We then use the trim-and-fill to correct for publication bias by imputing small-effect-small-sample studies that might have been performed, but not published. We then perform a meta-analysis on the imputed set of studies, and compare it to the complete case analysis in the main results.

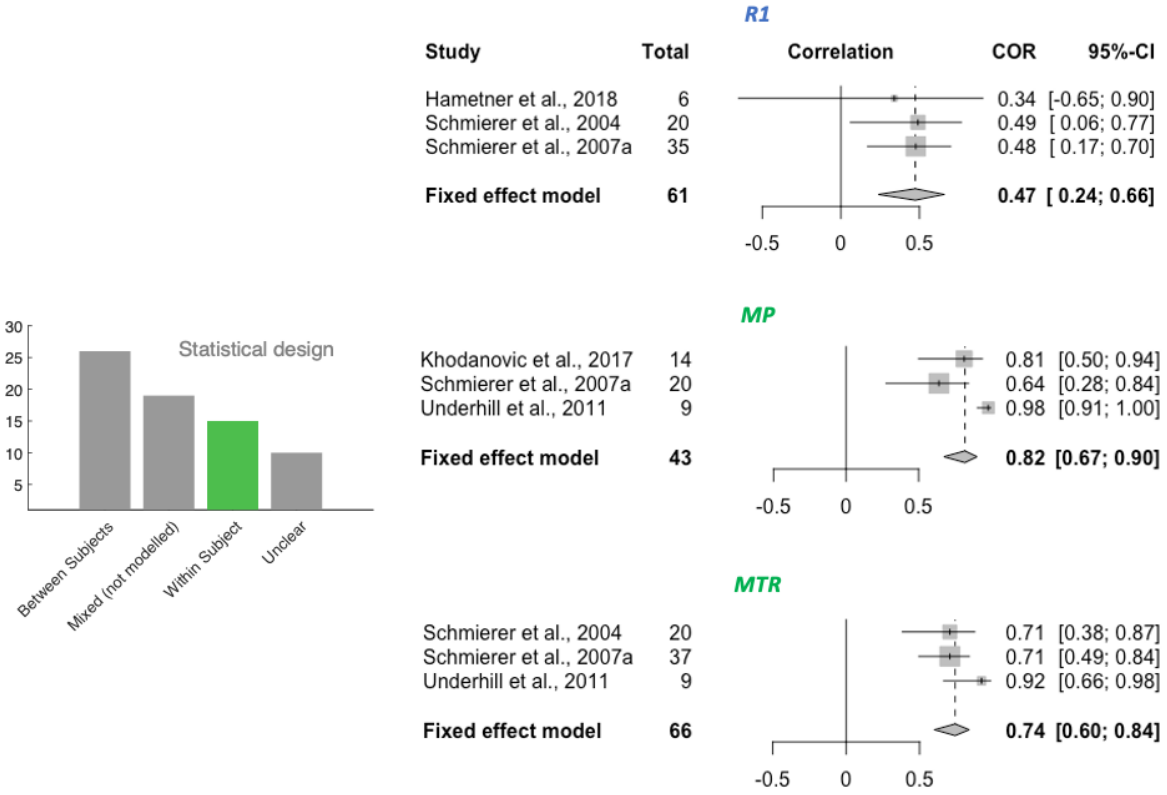

Figure 5: **Within-subject meta-analyses.** For the three metrics R1, MP and MTR, a meta-analysis on within-subject correlations could be performed. The resulting effect size estimate for MTR is higher than its between-subject effect size (see Figure 2).

Table 1: **Basic information of the assessed validation studies.** Provided are the species, condition/pathology studied, the tissue structure that underwent MR scanning, post-mortem time (if applicable), regions or tissue types of interest for the analysis and how they were defined. **Acronyms: ALS**: amytrophic lateral sclerosis; **APP/PS1**: Alzheimer mouse model; **Cuprizone**: Cuprizone-fed mice (D+R: demyelination and remyelination); **EAE**: allergic encephalomyelitis; **Kaolin**: Kaolin was used to induce communicating hydrocephalus; **LPS**: lipopolysaccharide-mediated animal model of MS, **MHII**: mild Hypoxic-Ischemic injury; **MPS**: mucopolysaccharidosis; **MS**: multiple sclerosis; **MSN-ChR2**: optical upregulation of Striatal medium spiny neurons **PLP-GFP**: Proteolipid protein - green fluorescent protein labelled mice; **TBI**: traumatic brain injury; **Thy1-eYFP-H**: mice that endogenously produce fluorescence signal; **TSC**: Tuberous sclerosis complex; **Shiverer**: Shiverer mice; **SCI**: spinal cord injury; **X-ALD**: X-Linked Adrenoleukodystrophy. **Anatomical structure: AC**: anterior commissure; **BG**: basal ganglia; **CC**: corpus callosum; **CP**: cerebellar peduncle; **CT**: cerebral cortex; **CST**: cortico-spinal tract; **DAWM**: diffusily abnormal white matter; **DTT**: dorsal tegmental tract; **DWMI**: diffuse white matter injury; **dGM**: deep gray matter; **GM**: gray matter; **NAWM**: normal appearing white matter; **OT**: optic tract; **PVWM**: Periventricular White Matter; **ST**: striatum; **Th**: Thalamus; **WM**: white matter; **WM-Ls**: white matter lesions.

Reference	Species	Pathology	Tissue structure	Post-mortem time	Regions / tissue types of interest	ROI definition
Abe et al. (2019)	Mouse	MSN-ChR2 mice	Whole brain	NA	dGM, CT	manual
Aojula et al. (2016)	Rat	Kaolin-induced hydocephalus and Controls	Whole brain	NA	CC and PVWM	manual for histology, coordinate-based for MRI
Argyridis et al. (2014)	Mouse	Different neonatal stages	Whole brain	NA	EC	manual
Bagnato et al. (2018)	Human	MS	1 cm coronal slices	4-83 hours	NAWM, DWMI, WM-Ls, thalamus, dGM, normal cortex	manual
Beckmann et al. (2018)	Mouse	Cuprizone	Whole brain	NA	CC, EC (within same ROI)	not reported
Bot et al. (2004)	Human	MS and Controls	Cervical spinal cord	8.2±1.7 hours	WM-Ls, NAWM, WM	manual
Chandran et al. (2012)	Mouse	Cuprizone	Whole brain	NA	WM	atlas-based
Chang et al. (2017b)	Mouse	Controls	Whole brain	NA	Various WM tracts	manual
Chang et al. (2017a)	Mouse	Thy1-eYFP-H mice and Controls	Whole brain	NA	Various WM tracts	manual
Chen et al. (2017)	Rat	SCI	3 mm spinal cord sections	NA	Fasciculus gracilis	manual
Choi et al. (2015)	Dog	MPS type 1 and Control	Whole brain	NA	ROIs across the brain, including WM and GM	manual
Duhamel et al. (2019)	Mouse	plp-GFP mice and Controls	Whole brain	NA	ROIs across the brain, including WM and GM	manual
Fatemi et al. (2011)	Mouse	Ischemic injury and control	Whole brain	NA	CC, IC, CP	manual
Fjær et al. (2013)	Mouse	Cuprizone and Controls	Whole brain	NA	CC, dGM, olfactory bulb, cerebellum and CT	manual
Fjær et al. (2015)	Mouse	EAE	Whole brain	NA	CC, dGM, olfactory bulb, cerebellum, CT	semi-automated
Gareau et al. (2000)	Guinea pig	EAE and Controls	Brain slice	NA	NAWM in CC	manual
Grussu et al. (2017)	Human	MS and Controls	Upper thoraric and upper lumbar spinal cord	not reported	GM, WM, lesional vs nonlesional	manual
Hakkarainen et al. (2016)	Rats	Controls	Whole brain	NA	ROIs across the brain	not reported
Hametner et al. (2018)	Human	Controls	Whole brain	range: 72-192 hours	WM, CT, Th, BG	manual
Janve et al. (2013)	Rat	LPS and control	Whole brain	NA	CC	not reported
Jelescu et al. (2016)	Mouse	Cuprizone and controls	Whole brain	NA	CC splenium	semi-automated
Jespersen et al. (2010)	Rat	Controls	Whole brain	NA	ROIs across the brain, including WM and GM	manual
Jito et al. (2008)	Rats	6 postnatal stages	Whole brain	NA	CC	manual

Continued on next page

Table 1: **Basic information of the assessed validation studies.** Provided are the species, condition/pathology studied, the tissue structure that underwent MR scanning, post-mortem time (if applicable), regions or tissue types of interest for the analysis and how they were defined. **Acronyms: ALS**: amytrophic lateral sclerosis; **APP/PS1**: Alzheimer mouse model; **Cuprizone**: Cuprizone-fed mice (D+R: demyelination and remyelination); **EAE**: allergic encephalomyelitis; **Kaolin**: Kaolin was used to induce communicating hydrocephalus; **LPS**: lipopolysaccharide-mediated animal model of MS, **MHII**: mild Hypoxic-Ischemic injury; **MPS**: mucopolysaccharidosis; **MS**: multiple sclerosis; **MSN-ChR2**: optical upregulation of Striatal medium spiny neurons **PLP-GFP**: Proteolipid protein - green fluorescent protein labelled mice; **TBI**: traumatic brain injury; **Thy1-eYFP-H**: mice that endogenously produce fluorescence signal; **TSC**: Tuberous sclerosis complex; **Shiverer**: Shiverer mice; **SCI**: spinal cord injury; **X-ALD**: X-Linked Adrenoleukodystrophy. **Anatomical structure: AC**: anterior commissure; **BG**: basal ganglia; **CC**: corpus callosum; **CP**: cerebellar peduncle; **CT**: cerebral cortex; **CST**: cortico-spinal tract; **DAWM**: diffusily abnormal white matter; **DTT**: dorsal tegmental tract; **DWMI**: diffuse white matter injury; **dGM**: deep gray matter; **GM**: gray matter; **NAWM**: normal appearing white matter; **OT**: optic tract; **PVWM**: Periventricular White Matter; **ST**: striatum; **Th**: Thalamus; **WM**: white matter; **WM-Ls**: white matter lesions.

Reference	Species	Pathology	Tissue structure	Post-mortem time	Regions / tissue types of interest	ROI definition
Kelm et al. (2016)	Mouse	TSC / Rictor	Whole brain	NA	CC midbody, CC genu, CC splenium (SCC), AC, EC, IC	manual
Khodanovich et al. (2017)	Mouse	Cuprizone (D+R) and Controls	Whole brain	NA	GM and WM	manual
Khodanovich et al. (2019)	Mouse	Cuprizone (D+R) and Controls	Whole brain	NA	ROIs across the brain, including WM and GM	manual
Kozlowski et al. (2008)	Rat	SCI and Controls	Spinal cord	NA	Fasciculus gracilis, fasciculus cuneatus, and CST	manual
Kozlowski et al. (2014)	Rats	SCI	Spinal cord	NA	Dorsal column	manual
Laule et al. (2008)	Human	MS	1 cm brain sections	not reported	WM, GM and lesions	manual
Laule et al. (2006)	Human	MS	1 cm brain sections	not reported	WM, GM and lesions	manual
Laule et al. (2011)	Human	MS	1 cm brain sections	not reported	DAWM and NAWM	manual
Lehto et al. (2017)	Rat	Controls	Whole brain	NA	Contralateral and ipsilateral CC and DTT	manual
Lodygensky et al. (2012)	Rat	Controls	Whole brain	NA	AC, CC	manual
Martirosyan et al. (2016)	Rat	SCI	Spinal cord (T9)	NA	Epicenter of injury	manual
Moll et al. (2011)	Human	MS	Whole brain	5.8±1 hours	WM, considering distance from lesion	manual
Mollink et al. (2019)	Human	ALS and controls	Brain tissue blocks including hippocampus	range: 0.5-5 days	Perforant path	tractography / manual
Mottershead et al. (2003)	Human	MS and Controls	Spinal cord, 2 cm pieces	72±39.2	MS lesions, diffuse damage, NAMW	manual
Oakden et al. (2015)	Rat	SCI	Caudal end of the C2 vertebra	NA	WM	manual
Odrobina et al. (2005)	Rat	Tellerium diet and Controls	Sciatic nerves	NA	Distal and proximal portion of nerve	not reported
Peters et al. (2019)	Human	Tuberculosis	Whole brain	NA	Surgically removed tissue blocks	NA
Pol et al. (2019)	Mouse	Teliflunomide and Controls	Whole brain	NA	CC (caudal medial, rostral lateral, rostral medial)	manual
Praet et al. (2018)	Mouse	APP/PS1 and controls	Whole brain	NA	ROIs across the brain, including WM and GM	manual on study based atlas
Pun et al. (2005)	Rat	Tellerium and Controls	2 cm pieces of sciatic nerves	NA	Nerves	unclear
Reeves et al. (2015)	Human	Epilepsy and Control	5 mm thick piece of cortex and underlying WM	0 (samples obtained during surgery)	Normal and pathological GM and WM	manual
Righart et al. (2017)	Human	MS	Whole brain	7 hours	Left inferior frontal gyrus, superior frontal gyrus, anterior cingulate gyrus, inferior parietal gyrus, and superior temporal gyrus	atlas (Freesurfer aparc)
Schmierer et al. (2010)	Human	MS	1 cm coronal brain slices	42±32 hours	Cortical GM lesions and nonlesional CT	manual
Schmierer et al. (2004)	Human	MS	1 cm coronal brain slices	14.2±8.5 hours	WM-Ls and NAWM	manual

Continued on next page

Table 1: **Basic information of the assessed validation studies.** Provided are the species, condition/pathology studied, the tissue structure that underwent MR scanning, post-mortem time (if applicable), regions or tissue types of interest for the analysis and how they were defined. **Acronyms: ALS**: amytrophic lateral sclerosis; **APP/PS1**: Alzheimer mouse model; **Cuprizone**: Cuprizone-fed mice (D+R: demyelination and remyelination); **EAE**: allergic encephalomyelitis; **Kaolin**: Kaolin was used to induce communicating hydrocephalus; **LPS**: lipopolysaccharide-mediated animal model of MS, **MHII**: mild Hypoxic-Ischemic injury; **MPS**: mucopolysaccharidosis; **MS**: multiple sclerosis; **MSN-ChR2**: optical upregulation of Striatal medium spiny neurons **PLP-GFP**: Proteolipid protein - green fluorescent protein labelled mice; **TBI**: traumatic brain injury; **Thy1-eYFP-H**: mice that endogenously produce fluorescence signal; **TSC**: Tuberous sclerosis complex; **Shiverer**: Shiverer mice; **SCI**: spinal cord injury; **X-ALD**: X-Linked Adrenoleukodystrophy. **Anatomical structure: AC**: anterior commissure; **BG**: basal ganglia; **CC**: corpus callosum; **CP**: cerebellar peduncle; **CT**: cerebral cortex; **CST**: cortico-spinal tract; **DAWM**: diffusily abnormal white matter; **DTT**: dorsal tegmental tract; **DWMI**: diffuse white matter injury; **dGM**: deep gray matter; **GM**: gray matter; **NAWM**: normal appearing white matter; **OT**: optic tract; **PVWM**: Periventricular White Matter; **ST**: striatum; **Th**: Thalamus; **WM**: white matter; **WM-Ls**: white matter lesions.

Reference	Species	Pathology	Tissue structure	Post-mortem time	Regions / tissue types of interest	ROI definition
Schmierer et al. (2007a)	Human	MS	Coronal brain slices	15±8 (range: 4.5-43) hours, scanning: 43±22 (range 10-107) hours	WM-Ls and NAWM	manual
Schmierer et al. (2007b)	Human	MS	1 cm coronal brain slices	16±6 (range: 7-28) hours, scanning: 46±25 (range 10-107) hours	WM-Ls and NAWM	manual
Schmierer et al. (2008)	Human	MS	1 cm coronal brain slices	17±6 (range: 7-28) hours, unfixed samples: 51±28 (range: 7-108) hours	WM-Ls and NAWM	manual
Schwartz et al. (2005)	Rat	Controls	Whole brain	NA	Various WM tracts	manual
Seehaus et al. (2015)	Human	Control	Whole brain	6 hours	CT and WM	NA
Seewann et al. (2009)	Human	MS	1 cm coronal brain slices	mean: 8.5 hours	WM-Ls, DAWM and NAWM	manual
Soni et al. (2020)	Mouse	TBI and Controls	Whole brain	NA	Middle CC, CC-EC ipsilateral and contralateral to injury	AMBMC mouse atlas / Allen mouse atlas
Soustelle et al. (2019)	Mouse	Cuprizone and Controls	Whole brain	NA	Medial and lateral CC, CT	manual
Stüber et al. (2014)	Human	Controls	Brain tissue blocks (pre/postcentral gyrus, posterior occipital lobe and subthalamic nucleus)	36 and 28 hours	GM and WM	NA
Sundberg et al. (2010)	Mouse	SCI and Controls	Spinal cord	NA	Area centred on injury	manual
Takagi et al. (2009)	Rat	Injury	Sciatic nerve	NA	Individual nerves	random selection of nerves
Thiessen et al. (2013)	Mouse	Cuprizone and Controls	Whole brain	NA	CC	manual
Tu et al. (2016)	Rat	TBI	Whole brain	NA	AC, CC, CT, CP, EC, PT, ST	manual
Turati et al. (2015)	Mouse	Cuprizone (D+R) and Controls	Whole brain	NA	CC	manual
Underhill et al. (2011)	Rat	C6 glioma model and Controls	Whole brain	NA	ROIs across the brain, including WM and GM	manual
van der Voorn et al. (2011)	Human	X-ALD and Controls	Coronal brain tissue sections	not reported	Areas with complete or active demyelination and NAWM	manual
van Tilborg et al. (2018)	Rat	Fetal inflammation and postnatal hypoxia	Whole brain	NA	Unclear	atlas-based

Continued on next page

Table 1: **Basic information of the assessed validation studies.** Provided are the species, condition/pathology studied, the tissue structure that underwent MR scanning, post-mortem time (if applicable), regions or tissue types of interest for the analysis and how they were defined. **Acronyms: ALS**: amytrophic lateral sclerosis; **APP/PS1**: Alzheimer mouse model; **Cuprizone**: Cuprizone-fed mice (D+R: demyelination and remyelination); **EAE**: allergic encephalomyelitis; **Kaolin**: Kaolin was used to induce communicating hydrocephalus; **LPS**: lipopolysaccharide-mediated animal model of MS, **MHII**: mild Hypoxic-Ischemic injury; **MPS**: mucopolysaccharidosis; **MS**: multiple sclerosis; **MSN-ChR2**: optical upregulation of Striatal medium spiny neurons **PLP-GFP**: Proteolipid protein - green fluorescent protein labelled mice; **TBI**: traumatic brain injury; **Thy1-eYFP-H**: mice that endogenously produce fluorescence signal; **TSC**: Tuberous sclerosis complex; **Shiverer**: Shiverer mice; **SCI**: spinal cord injury; **X-ALD**: X-Linked Adrenoleukodystrophy. **Anatomical structure: AC**: anterior commissure; **BG**: basal ganglia; **CC**: corpus callosum; **CP**: cerebellar peduncle; **CT**: cerebral cortex; **CST**: cortico-spinal tract; **DAWM**: diffusily abnormal white matter; **DTT**: dorsal tegmental tract; **DWMI**: diffuse white matter injury; **dGM**: deep gray matter; **GM**: gray matter; **NAWM**: normal appearing white matter; **OT**: optic tract; **PVWM**: Periventricular White Matter; **ST**: striatum; **Th**: Thalamus; **WM**: white matter; **WM-Ls**: white matter lesions.

Reference	Species	Pathology	Tissue structure	Post-mortem time	Regions / tissue types of interest	ROI definition
Wang et al. (2009)	Rat	MHII	Whole brain	NA	EC (ipsilesional and contralesional)	manual
Wang et al. (2015)	Human	MS	Section of cervical spinal cord	<10 hours	Voxels with positive histological staining	random selection of voxels
Warntjes et al. (2017)	Human	Controls	2 cm brain slices (one through caudate nucleus and one through thalamus)	20 hours to 3 days range (during which body was refrigerated)	GM and WM	NA
Wei et al. (2013)	Dog	Controls	Whole brain	NA	ROIs across the brain, including WM and GM	manual
West et al. (2018)	Mouse	Genetic model for hyper-/hypomyelination	Whole brain	NA	CC, AC, CT	manual
Yano et al. (2018)	Mouse	Cuprizone (D+R) and Controls	Whole brain	NA	CC	manual

Table 2: **Information on MRI imaging of the assessed validation studies.** Provided are imaging modality, field strength in Tesla (T), imaging resolution in terms of voxel size, and the tissue state and temperature during scanning. **Acronyms: Imaging: 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); **lADC**: 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; **M0b**: 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 an 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.

Reference	Modality	Metrics	T	Resolution (mm)	Tissue state	Temperature
Abe et al. (2019)	DWI	AD, FA, RD	7	0.125 x 0.125 x 0.125 mm	perfusion fixed	not reported
Aojula et al. (2016)	DWI	AD, FA, MD, RD	7	0.195 x 0.195 x 1mm	in vivo	body temperature
Argyridis et al. (2014)	QSM	Susceptibility	9.4	0.06 mm isotropic	perfusion fixed	not reported
	DWI	MD	9.4	unclear	perfusion fixed	not reported
Bagnato et al. (2018)	Relaxometry	R2*	7	0.7 mm isotropic	fixed	room temperature
Beckmann et al. (2018)	MT	MTR	7	0.094 0.094 x 0.5 mm	in vivo	body temperature
Bot et al. (2004)	Relaxometry	T1, T2	4.7	0.007 x 0.007 x 1 mm	fixed	not reported
	MT	MTR	4.7	0.007 x 0.007 x 1 mm	fixed	not reported
Chandran et al. (2012)	DWI	FA, RD	7	0.2 x 0.2 x 1 mm	in vivo	body temperature
Chang et al. (2017b)	DWI	FA, AD, RD	11.7	0.1 mm isotropic	CLARITY fixation	28°C

Continued on next page

Table 2: **Information on MRI imaging of the assessed validation studies.** Provided are imaging modality, field strength in Tesla (T), imaging resolution in terms of voxel size, and the tissue state and temperature during scanning. **Acronyms: Imaging: 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); **lADC**: 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; **M0b**: 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 an 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.

Reference	Modality	Metrics	T	Resolution (mm)	Tissue state	Temperature
Chang et al. (2017a)	DWI	FA, AD, RD, MD	11.7	0.1 mm isotropic	CLARITY fixation	28°C
Chen et al. (2017)	Relaxometry	MWF	7	0.07 x 0.07 x 05 mm	perfusion fixed	not reported
Choi et al. (2015)	DWI	FA, RD	7	0.1 mm isotropic	fixed	not reported
Duhamel et al. (2019)	MT	ih-MTR	11.75	0.3125 x 0.3125 x 1 mm	in vivo	37°C
Fatemi et al. (2011)	MT	MTR	9.4	0.083 x 0.81 x 0.8 mm	in vivo	body temperature
Fjær et al. (2013)	MT	MTR	7	0.2 x 0.2 x 0.23 mm	in vivo	body temperature
Fjær et al. (2015)	MT	MTR	7	0.2 x 0.2 x 0.23 mm	in vivo	body temperature
Gareau et al. (2000)	MT	MTR	4	0.160 x 0.3 x 0.5 mm	in vivo	body temperature
Grussu et al. (2017)	DWI	AD, FA, MD, RD	9.4	0.2 x 0.2 x 2 mm	fixed	35°C
Hakkarainen et al. (2016)	Relaxometry	RAFFn, T1, T2	9.4	0.156 x 0.156 x .35 mm	perfusion fixed	not reported
	MT	MTR	9.4	0.0293 x 0.0293 x 0.7 mm	perfusion fixed	not reported
Hametner et al. (2018)	Relaxometry	T1, R2*	7	0.43 x 0.43 x 0.65 mm	in situ	room temperature (20 deg)
	QSM	Susceptibility	7	0.43 x 0.43 x 0.65 mm	in situ	room temperature (20 deg)
Janve et al. (2013)	DWI	AD, FA, RD	9.4	0.167 x 0.167 x 0.167 mm	perfusion fixed	not reported
	MT	MP (PSR)	9.4	0.167 x 0.167 x 0.167 mm	perfusion fixed	not reported
Jelescu et al. (2016)	DWI	RD, RK	7	0.112 x 0.112 x 0.8 mm	in vivo	body temperature
	Relaxometry	T2	7	0.112 x 0.112 x 0.8 mm	in vivo	body temperature
	MT	MTR	7	0.112 x 0.112 x 0.8 mm	in vivo	body temperature
Jespersen et al. (2010)	DWI	FA*	16.4	0.1 x 0.1 x 0.5 mm	perfusion fixed	21°C
Jito et al. (2008)	DWI	FA	7	0.117 x 0.24 x 1 mm	in vivo	body temperature
Kelm et al. (2016)	DWI	FA, MD, RD, MK, AK, RK	15.2	0.15 x 0.15 x 0.15 mm	perfusion fixed	17 ± 0.5 °C
	Relaxometry	MWF	15.2	0.15 x 0.15 x 0.15 mm	perfusion fixed	17 ± 0.5 °C
	MT	MP (PSR)	15.2	0.15 x 0.15 x 0.15 mm	perfusion fixed	17 ± 0.5 °C
Khodanovich et al. (2017)	MT	MP ('MPF')	11.7	0.1 x 0.1 x 0.5 mm	fixed	body temperature
Khodanovich et al. (2019)	MT	MP ('MPF')	11.7	0.1 x 0.1 x 0.5 mm	fixed	body temperature
Kozlowski et al. (2008)	DWI	FA, AD, MD, RD	7	1 x 1 x 1 mm	perfusion fixed	not reported
	Relaxometry	MWF	7	0.078 x 0.078 x 1 mm	perfusion fixed	not reported
Kozlowski et al. (2014)	DWI	FA	7	1 x 1 x 1 mm (only ex vivo)	in vivo (relaxometry) and fixed (DWI)	body temperature / not reported
	Relaxometry	MWF	7	0.117 x 0.117 x 1 mm (ex vivo) 1 x 1 x 1.5 mm (in vivo)	in vivo (relaxometry) and fixed (DWI)	body temperature / not reported
Laule et al. (2008)	Relaxometry	MWF	7	0.234 x 0.234 x 1 mm	fixed	not reported

Continued on next page

Table 2: **Information on MRI imaging of the assessed validation studies.** Provided are imaging modality, field strength in Tesla (T), imaging resolution in terms of voxel size, and the tissue state and temperature during scanning. **Acronyms: Imaging: 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); **lADC**: 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; **M0b**: 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 an 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.

Reference	Modality	Metrics	T	Resolution (mm)	Tissue state	Temperature
Laule et al. (2006)	Relaxometry	MWF	1.5	in plane resolution not reported, 3mm thick slices	fixed	20°C
Laule et al. (2011)	Relaxometry	MWF	1.5 and 7	0.586 x 0.586 x 3 mm (1.5T), 0.234 x 0.234 x 1mm (7T)	fixed	not reported
Lehto et al. (2017)	Relaxometry	RAFF4	7	0.125 x 0.125 x 0.5 mm	in vivo	body temperature
	MT	T1sat, MTR	7	0.125 x 0.125 x 0.5 mm	in vivo	body temperature
	DWI	AD, FA, MD, RD	7	0.125 x 0.125 x 0.5 mm	in vivo	body temperature
Lodygensky et al. (2012)	QSM	Susceptibility	9.4	slice thickness of 0.5 mm, in-plane resolution between 0.0625 mm isotropic and 0.105 mm isotropic	in vivo	body temperature
Martirosyan et al. (2016)	DTI	FA	7	0.195 x 0.195 x 10 mm	perfusion fixed	not reported
Moll et al. (2011)	DWI	AD, FA, MD, RD	1.5	1.9 1.9 3 mm	in situ / fixed	body temperature
	MT	MTR	1.5	0.9 0.9 3 mm	in situ / fixed	body temperature
Mollink et al. (2019)	DWI	FA, MD, RD, AD	11.7	0.4 x 0.4 x 0.4 mm resolution.	fixed	not reported
Mottershead et al. (2003)	Relaxometry	PD, T1, T2	7	Single 1.5 mm slices, in plane resolution between 0.055 and 0.068 mm	fresh	29-38°C
	DWI	ADC, SDI	7	Single 1.5 mm slices, in plane resolution between 0.055 and 0.068 mm	fresh	29-38°C
	MT	MTR	7	Single 1.5 mm slices, in plane resolution between 0.055 and 0.068 mm	fresh	29-38°C
Oakden et al. (2015)	DWI	AD, RD	1.5	0.2 0.2 x 1 mm	in vivo / fixed	body temperature
	Relaxometry	MWF	1.5	0.2 0.2 x 1 mm	in vivo / fixed	body temperature
Odrobina et al. (2005)	Relaxometry	MWF, T1, T2	1.5	not reported	fixed	room temperature (20°C)
	MT	MTR, MP('M0b')	1.5	not reported	fixed	room temperature (20°C)
Peters et al. (2019)	DWI	FA, MD	3	1.72 x 1.72 x 2.2 mm	in vivo	body temperature
Pol et al. (2019)	DWI	FA, MD	9.4	0.078 0.078 0.250 mm	in vivo	body temperature
	QSM	Susceptibility	9.4	not reported	in vivo	body temperature
Praet et al. (2018)	DWI	AD, AK, DT, DK, FA, MD, MK, RD, RK	7	0.000214 x 0.000214 x 0.2 mm	in vivo	body temperature (37-37.3°C)
Pun et al. (2005)	Relaxometry	T1, MWF	1.5	not reported	fresh	20 °C
Reeves et al. (2015)	Relaxometry	T1, T2, T2*	9.4	0.136 x 0.136 x 0.5 mm	fixed	not reported
	MT	MTR	9.4	0.136 x 0.136 x 0.500 mm	fixed	not reported
Righart et al. (2017)	Other	T1w/T2w	1.5	1 x 1 x 1.5 mm	in situ	not reported
Schmierer et al. (2010)	Relaxometry	T1	9.4	0.117 x 0.156 x 1 mm	fixed	not reported
	MT	MTR	9.4	not reported	fixed	not reported

Continued on next page

Table 2: **Information on MRI imaging of the assessed validation studies.** Provided are imaging modality, field strength in Tesla (T), imaging resolution in terms of voxel size, and the tissue state and temperature during scanning. **Acronyms: Imaging: 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); **lADC**: 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; **M0b**: 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 an 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.

Reference	Modality	Metrics	T	Resolution (mm)	Tissue state	Temperature
Schmierer et al. (2004)	Relaxometry	T1	1.5	0.938 x 0.938 x 5 mm	fresh	mean 25.0°C (SD: 2.9°C)
	MT	MTR	1.5	0.938 x 0.938 x 5 mm	fresh	mean 25.0°C (SD: 2.9°C)
Schmierer et al. (2007a)	Relaxometry	T1	1.5	0.938 x 0.938 x 5 mm	fresh	room temperature
	MT	MTR, MP('fb')	1.5	0.938 x 0.938 x 5 mm	fresh	room temperature
Schmierer et al. (2007b)	DWI	FA, MD	1.5	2.5 x 2.5 x 5 mm	fresh	17.5 - 25.1°C
Schmierer et al. (2008)	MT	MTR, MP('fb')	1.5	0.94 x 0.94 x 5 mm	fresh / fixed	22.5°C
	Relaxometry	T1, T2	1.5	0.94 x 0.94 x 5 mm	fresh / fixed	22.5°C
	DWI	AD, FA, MD, RD	1.5	0.25 x 0.25 x 5 mm	fresh / fixed	22.5°C
Schwartz et al. (2005)	DWI	AI, iADC, tADC	9.4	0.039 x 0.039 x 0.5 mm	perfusion fixed	20°C
Seehaus et al. (2015)	DWI	FA	9.4	0.34 x 0.34 x 0.34 mm	fixed	30 °C
Seewann et al. (2009)	Relaxometry	T1	1.5	1 x 1 x 3 mm	fixed	not reported
	DWI	FA, ADC	1.5	2 x 2 x 8 mm	fixed	not reported
	MT	MTR	1.5	1 x 1 x 5 mm	fixed	not reported
Soni et al. (2020)	QSM	Susceptibility	9.4	0.1 x 0.1 x 0.3 mm	in vivo	body temperature (36.537°C)
Soustelle et al. (2019)	MT	MP('f')	7	0.1 x 0.1 x 0.750 mm	perfusion fixed	room temperature (20°C)
	DWI	RD	7	0.1 x 0.1 x 0.750 mm	perfusion fixed	room temperature (20°C)
	Other	rSPF	7	0.152 x 0.152 x 0.750 mm	perfusion fixed	room temperature (20°C)
	Relaxometry	MWF	7	0.152 x 0.152 x 0.750 mm	perfusion fixed	room temperature (20°C)
Stüber et al. (2014)	Relaxometry	R1, R2*	7	R1: 0.1 or 0.2 mm isotropic, R2*: 0.2 mm isotropic	fixed	not reported
	QSM	Susceptibility	7	0.070 mm isotropic	fixed	not reported
Sundberg et al. (2010)	DWI	AD, FA, RD	7	0.2 x 0.2 x 1 mm	in vivo	37°C
Takagi et al. (2009)	DWI	FA	7	0.31 x 0.31 x 0.94 mm	excised	not reported
Thiessen et al. (2013)	DWI	AD, FA, RD	7	0.098 0.098 0.750 mm	perfusion fixed	room temperature
	MT	MP('f'), MTR	7	0.098 0.098 0.750 mm	perfusion fixed	body temperature
	Relaxometry	T1	7	0.098 0.098 0.750 mm	perfusion fixed	room temperature
Tu et al. (2016)	DWI	AD, FA, MD, RD, MTR at different ppm	7	DWI: 0.2 x 0.2 x 0.2 mm; MT: 0.2 x 0.2 x 0.5 mm;	in vivo	body temperature
Turati et al. (2015)	MT	MP('f')	7	0.1 x 0.1 x 0.6 mm	in vivo	body temperature
Underhill et al. (2011)	MT	MP('BPF'), MTR	3	0.3 0.3 0.3 mm	in vivo	body temperature
van der Voorn et al. (2011)	DWI	ADC, FA	1.5	2 x 2 x 8 mm	fixed	room temperature (20-22°C)
	MT	MTR	1.5	not reported (3 mm sections)	fixed	room temperature (20-22°C)

Continued on next page

Table 2: **Information on MRI imaging of the assessed validation studies.** Provided are imaging modality, field strength in Tesla (T), imaging resolution in terms of voxel size, and the tissue state and temperature during scanning. **Acronyms: Imaging: 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); **lADC**: 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; **M0b**: 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 an 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.

Reference	Modality	Metrics	T	Resolution (mm)	Tissue state	Temperature
van Tilborg et al. (2018)	DWI	FA, RD	9.4	0.15 x 0.15 x 0.148 mm	perfusion fixed	not reported
Wang et al. (2009)	DWI	AD, FA, RD, "Trace"	7	Varying resolution: 0.250 x 0.250 x 0.5 mm for some scans, 0.313 x 0.313 x 0.7 mm for others	in vivo	body temperature
Wang et al. (2015)	DWI	DBSI-RD	4.7	0.250 x 0.250 x 0.5 mm	fixed	17°C
Warntjes et al. (2017)	Relaxometry	MWF, R1	3	0.7 x 0.7 x 4 mm	in situ	mean: 7.8° (SD: 3.1°C)
Wei et al. (2013)	DWI	FA, RD	7	0.1 mm isotropic	perfusion fixed	not reported
West et al. (2018)	Relaxometry	MWF	15.2	.150 x .150 x .150 mm	perfusion fixed	not reported
	MT	MP (BPF')	15.2	.150 x .150 x .150 mm	perfusion fixed	not reported
Yano et al. (2018)	DWI	FA, MD, RD	7	0.125 0.125 0.125 mm	perfusion fixed	not reported

Table 3: **Information on histology of the assessed validation studies.** Provided are information in tissue preparation, histological slice thickness, the histological method for myelin (and whether iron and axons were also considered) and the obtained histological metric. **Acronyms: AMG**: Autometallographic myelin stain; **LFB**: Luxol fast blue stain; **MA**: fraction of myelinated axons; **MAB238**: Anti-oligodendrocyte immunohistochemistry; **MBP**: Anti-myelin-basic-protein immunohistochemistry; **MSF**: myelin sheath fraction; **MST**: myelin sheath thickness; **MVF**: myelin volume fraction; **PAS**: periodic acid-Schiff; **PIXE**: proton-induced X-ray emission; **PLP**: Anti-proteolipid-protein immunohistochemistry.

Reference	Tissue preparation	Slice thickness	Myelin histology	Iron	Axons	Histology Metric
Abe et al. (2019)	cryosectioning	25 μm	PLP	-	+	staining fraction
Aojula et al. (2016)	perfusion fixation, cryosectioning	15 μm	MBP	-	-	staining fraction
Argyridis et al. (2014)	not reported	2 μm	LFB	+	-	contrasted and normalised luminance
Bagnato et al. (2018)	not reported	10 μm	LFB and PLP	+	-	staining intensity
Beckmann et al. (2018)	paraffin embedding	3 μm	LFB	-	-	staining intensity
Bot et al. (2004)	not reported	5 μm	LFB	-	+	normalised staining intensity
Chandran et al. (2012)	paraffin embedding	5 μm	LFB, MBP	-	-	staining intensity
Chang et al. (2017b)	CLARITY	3D	MBP	-	-	staining intensity
Chang et al. (2017a)	CLARITY	3D	MBP	-	-	normalised staining intensity
Chen et al. (2017)	ultrasectioning	1 μm	EM	-	+	staining fraction
Choi et al. (2015)	cryosectioning	50 μm	Gold chloride	-	+	staining intensity
Duhamel et al. (2019)	perfusion fixation, cryosectioning	20 μm	PLP-GFP fluorescence	-	-	normalised (background) staining intensity
Fatemi et al. (2011)	perfusion fixation, paraffin embedding / cryosectioning	20 μm / 40 μm	LFB, MBP	-	+	staining intensity
Fjær et al. (2013)	fixation, paraffin embedding	7 μm	PLP	-	-	staining fraction
Fjær et al. (2015)	fixation, paraffin embedding	7 μm	LFB, PLP	+	-	manual scoring, staining fraction
Gareau et al. (2000)	not reported	5 μm	Solochrome-R-cyanine	-	-	score of estimated 'myelin pallor'
Grussu et al. (2017)	paraffin embedding	10 μm	PLP	-	+	staining fraction

Continued on next page

Table 3: **Information on histology of the assessed validation studies.** Provided are information in tissue preparation, histological slice thickness, the histological method for myelin (and whether iron and axons were also considered) and the obtained histological metric. **Acronyms: AMG**: Autometallographic myelin stain; **LFB**: Luxol fast blue stain; **MA**: fraction of myelinated axons; **MAB238**: Anti-oligodendrocyte immunohistochemistry; **MBP**: Anti-myelin-basic-protein immunohistochemistry; **MSF**: myelin sheath fraction; **MST**: myelin sheath thickness; **MVF**: myelin volume fraction; **PAS**: periodic acid-Schiff; **PIXE**: proton-induced X-ray emission; **PLP**: Anti-proteolipid-protein immunohistochemistry.

Reference	Tissue preparation	Slice thickness	Myelin histology	Iron	Axons	Histology Metric
Hakkarainen et al. (2016)	cryosectioning	30 μm	Gold chloride	+	-	normalised staining intensity
Hametner et al. (2018)	fixation, cutting, paraffin embedding	10 μm	LFB	+	-	staining intensity
Janve et al. (2013)	paraffin embedding	10 μm	LFB	-	-	normalised staining intensity
Jelescu et al. (2016)	perfusion fixation, nanosectioning	2.36 nm	EM	-	+	staining fraction
Jespersen et al. (2010)	cryosectioning	40 μm	AMG	-	+	staining intensity
Jito et al. (2008)	perfusion fixation, semithin sectioning	750 nm	Toluidine blue	-	+	staining fraction
Kelm et al. (2016)	ultrasectioning	70 nm	EM	-	+	staining fraction
Khodanovich et al. (2017)	cryosectioning	10 μm	LFB	-	-	staining intensity
Khodanovich et al. (2019)	cryosectioning	10 μm	MBP	-	-	staining fraction
Kozlowski et al. (2008)	cryosectioning	20 μm	LFB, MBP	-	+	staining intensity (MBP), inverse staining intensity (LFB)
Kozlowski et al. (2014)	cryosectioning	20 μm	MBP, Eriochromecyanine	-	+	normalised staining intensity
Laule et al. (2008)	paraffin embedding	10 μm	LFB	-	-	staining intensity
Laule et al. (2006)	paraffin embedding	10 μm	LFB	-	-	staining intensity
Laule et al. (2011)	paraffin embedding	10 μm	LFB, MBP	-	+	normalised staining intensity
Lehto et al. (2017)	cryosectioning	30 μm	Gold chloride	-	-	staining intensity
Lodygensky et al. (2012)	perfusion fixation, cryosectioning	50 μm	Black Gold II	+	-	staining intensity
Martirosyan et al. (2016)	paraffin embedding	15 μm	LFB	-	-	staining intensity, percentage difference between injury epicenter to control animals
Moll et al. (2011)	cryosectioning	30 μm	MBP	-	+	staining intensity
Mollink et al. (2019)	paraffin embedding	6 μm	PLP	-	+	staining fraction
Mottershead et al. (2003)	fixation, paraffin embedding	10 μm	LFB	-	+	staining fraction
Oakden et al. (2015)	paraffin embedding	10 μm	LFB	-	-	visual classification of severity of pathology
Odrobina et al. (2005)	epon-araldite embedding	1 μm	Toluidine blue	-	-	staining fraction
Peters et al. (2019)	fixation and sectioning	13 μm	LFB	-	-	staining intensity
Pol et al. (2019)	perfusion fixation and cryosectioning	16 μm	Solochrome	-	-	staining intensity
Praet et al. (2018)	fixation, paraffin embedding	5 μm	MBP	-	-	staining intensity
Pun et al. (2005)	fixation, ultrasectioning	1 μm	Toluidine blue	-	+	staining fraction
Reeves et al. (2015)	not reported	7 μm	MBP	-	+	staining intensity
Righart et al. (2017)	paraffin embedding	10 μm	PLP	-	+	staining intensity
Schmierer et al. (2010)	paraffin embedding	5 μm	MBP	-	+	inverse staining intensity
Schmierer et al. (2004)	paraffin embedding	not reported	LFB	-	+	inverse staining intensity
Schmierer et al. (2007a)	fixation, paraffin embedding	not reported	LFB	-	+	inverse staining intensity
Schmierer et al. (2007b)	fixation, paraffin embedding	9.4±3.6 μm	LFB	-	+	inverse staining intensity
Schmierer et al. (2008)	paraffin embedding	5000 μm	LFB	-	+	inverse staining intensity
Schwartz et al. (2005)	epon immersion	1 μm	Toluidine blue	-	+	staining fraction (MVF, MST)
Seehaus et al. (2015)	cryosectioning	60 μm	Gallyas	-	-	staining intensity
Seewann et al. (2009)	paraffin embedding	10 μm	LFB, PAS, PLP	-	+	inverse staining intensity
Soni et al. (2020)	perfusion fixation, paraffin embedding	10 μm	MBP	-	-	manual scoring of immunoreactivity
Soustelle et al. (2019)	cryosectioning	60 μm	MBP	-	-	normalized staining intensity
Stüber et al. (2014)	cryosectioning	30 μm	PIXE	+	-	model-based estimation of myelin volume fraction

Continued on next page

Table 3: **Information on histology of the assessed validation studies.** Provided are information in tissue preparation, histological slice thickness, the histological method for myelin (and whether iron and axons were also considered) and the obtained histological metric. **Acronyms: AMG**: Autometallographic myelin stain; **LFB**: Luxol fast blue stain; **MA**: fraction of myelinated axons; **MAB238**: Anti-oligodendrocyte immunohistochemistry; **MBP**: Anti-myelin-basic-protein immunohistochemistry; **MSF**: myelin sheath fraction; **MST**: myelin sheath thickness; **MVF**: myelin volume fraction; **PAS**: periodic acid-Schiff; **PIXE**: proton-induced X-ray emission; **PLP**: Anti-proteolipid-protein immunohistochemistry.

Reference	Tissue preparation	Slice thickness	Myelin histology	Iron	Axons	Histology Metric
Sundberg et al. (2010)	perfusion fixation, cryosectioning	35 μm	MAB328	-	+	staining fraction
Takagi et al. (2009)	fixation and sectioning	80 nm	EM	-	+	staining fraction
Thiessen et al. (2013)	epon embedding	1 μm	EM	-	+	staining fraction (MA, MSF)
Tu et al. (2016)	cryosections	10 μm	MBP	-	+	normalised staining intensity
Turati et al. (2015)	perfusion fixation, cryosectioning	10 μm	Black Gold II, MBP	-	-	staining intensity
Underhill et al. (2011)	perfusion fixation, paraffin embedding	5 μm	LFB	-	+	normalized staining intensity
van der Voorn et al. (2011)	paraffin embedding	7 μm	LFB	-	+	inverse staining intensity
van Tilborg et al. (2018)	paraffin embedding	8 μm	MBP	-	-	staining fraction
Wang et al. (2009)	perfusion fixation, cryosectioning	10 μm	LFB	-	+	staining intensity
Wang et al. (2015)	paraffin embedding	5 μm	LFB	-	+	staining fraction
Warntjes et al. (2017)	fixation	4 μm	LFB	-	-	staining intensity
Wei et al. (2013)	cryosectioning	50000 μm	Gold chloride	-	-	inverse staining intensity
West et al. (2018)	ultrasectioning	0.07 μm	EM	-	+	staining fraction
Yano et al. (2018)	cryosectioning	25 μm	PLP	-	-	staining fraction

Reference Tissue preparation Slice thickness Myelin histology Iron Axons Histology Metric

Table 4: **Information on the employed statistical approach of the assessed validation studies.** Provided are information on the nature of data points that entered the correlation analysis, the method for coregistration between MRI and histology, sample size (number of subjects per group, total number of subjects, and number of data points for the analysis), the type of variance that was modelled in the statistical analysis, the reported statistic, and whether the study was selected for our meta-analysis (MA).

Reference	Correlation	Coregistration	N (Subjects/group)	N (Total subjects)	N (ROIs)	Statistical Design	Statistics	MA
Abe et al. (2019)	ROI to ROI	No coregistration	8	8	96 (in total across subjects)	Mixed (not modelled)	Pearson	-
Aojula et al. (2016)	ROI to ROI	No coregistration	4 kaolin-only, 6 kaolin + PBS (control intervention), 5 kaolin + decorin, 4 controls,	19	2	Between-subject	Spearman	+
Argyridis et al. (2014)	ROI to ROI	Not reported	18	18	3 (pooled into 1 weighted measure)	Between-subject	Not reported (likely linear regression)	+
Bagnato et al. (2018)	ROI to ROI	No coregistration	7	7	429 (in total across subjects)	Within-subject (mixed, modelled)	Pearson; linear regression	+
Beckmann et al. (2018)	ROI to ROI	Not reported	5	20	1	Between-subject	Pearson	+

Continued on next page

Table 4: **Information on the employed statistical approach of the assessed validation studies.** Provided are information on the nature of data points that entered the correlation analysis, the method for coregistration between MRI and histology, sample size (number of subjects per group, total number of subjects, and number of data points for the analysis), the type of variance that was modelled in the statistical analysis, the reported statistic, and whether the study was selected for our meta-analysis (MA).

Reference	Correlation	Coregistration	N (Subjects/group)	N (Total subjects)	N (ROIs)	Statistical Design	Statistics	MA
Bot et al. (2004)	ROI to ROI	No coregistration	11 MS, 2 controls	13	222 (in total across subjects)	Mixed (unclear)	Spearman; linear regression	-
Chandran et al. (2012)	ROI to ROI	Not reported	5 to 6 per group	20	1	Between-subject	Linear regression	+
Chang et al. (2017b)	ROI to ROI	Linear registration (12 dof affine)	4	4	14 (in total across subjects)	Mixed (not modelled)	Spearman; linear regression	-
Chang et al. (2017a)	ROI to ROI	Unclear	4	4	14 (in total across subjects)	Mixed (not modelled)	Spearman	-
Chen et al. (2017)	ROI to ROI	NA	3 to 6 per group	14	1	Between-subject	not reported	-
Choi et al. (2015)	ROI to ROI	Linear registration (affine, in Matlab)	2 MPS brains, 1 control	3	16	Within-subject	Correlation coefficient type not reported; Multivariate linear regression	-
Duhamel et al. (2019)	ROI to ROI	No coregistration	3	3	8 (7 ROIs + 1 "ROI" of relative GM/WM contrast)	Mixed (not modelled)	Pearson; linear regression	-
Fatemi et al. (2011)	ROI to ROI	Not reported	29 ischemic injury, 33 controls	61	25 (in total across subjects)	Unclear (likely mixed, not modelled)	Linear regression	-
Fjær et al. (2013)	ROI to ROI	No coregistration	6 controls, 48 cuprizone at varying stages	54	4	Between-subject	Linear regression	+
Fjær et al. (2015)	ROI to ROI	Not reported	6	24	5	Between-subject	Linear regression	+
Gareau et al. (2000)	unclear (likely ROI to ROI)	No coregistration	6x4 EAE at different time points, 6 controls	24	1	Between-subject	Spearman	-
Grussu et al. (2017)	ROI to ROI	Landmark-based non-linear registration	2 MS, 2 controls	4	48 (in total across subjects)	Mixed (not modelled)	Pearson; linear regression	-
Hakkarainen et al. (2016)	ROI to ROI	Not reported	5 to 6 per group	5	60 (in total across subjects)	Not reported (likely mixed, not modelled)	Pearson	-
Hametner et al. (2018)	ROI to ROI	No coregistration	6	6	2870 (in total across subjects) for R2*, 2809 (in total across subjects) for T1	Within-subject (mixed, modelled)	Linear regression	+
Janve et al. (2013)	ROI to ROI	Manual coregistration	8 LPS, 1 control	9	6	Mixed (not modelled)	Pearson	-

Continued on next page

Table 4: **Information on the employed statistical approach of the assessed validation studies.** Provided are information on the nature of data points that entered the correlation analysis, the method for coregistration between MRI and histology, sample size (number of subjects per group, total number of subjects, and number of data points for the analysis), the type of variance that was modelled in the statistical analysis, the reported statistic, and whether the study was selected for our meta-analysis (MA).

Reference	Correlation	Coregistration	N (Subjects/group)	N (Total subjects)	N (ROIs)	Statistical Design	Statistics	MA
Jelescu et al. (2016)	ROI to ROI	NA	2 x 12 Cuprizone, 10 controls	34	1	Between-subject	Partial Spearman (weight as covariate)	+
Jespersen et al. (2010)	ROI to ROI	No coregistration	3	3	8	Mixed (not modelled)	Pearson	-
Jito et al. (2008)	ROI to ROI	NA	6	36	1	Between-subject	Linear regression	+
Kelm et al. (2016)	ROI to ROI	No coregistration	4 TSC, 3 Rictor, 5 Controls	12	6	Mixed (not modelled)	not reported	-
Khodanovich et al. (2017)	ROI to ROI	No coregistration	7 cuprizone, 7 controls	14	6	Within-subject (mixed, modelled)	Pearson; linear regression	+
Khodanovich et al. (2019)	ROI to ROI	No coregistration	4 demyelination, 5 remyelination, 4 controls	13	1	Between-subject	Pearson; linear regression	+
Kozlowski et al. (2008)	ROI to ROI	No coregistration	6 at 3 weeks post injury, 4 at 8 weeks post injury, 6 controls (controls not used in correlation)	6 and 4 depending on the analysis	3	Mixed (not modelled)	Pearson	-
Kozlowski et al. (2014)	ROI to ROI	No coregistration	8	16	1	Between-subject	Pearson	-
Laule et al. (2008)	ROI to ROI	Landmark-based linear registration	3 MS patients, 10 samples	3	22-30 per sample	Within-subject	Not reported (likely linear regression)	-
Laule et al. (2006)	ROI to ROI	Landmark-based linear registration	13	13	23-45 per slice	Within-subject	Not reported (likely linear regression)	+
Laule et al. (2011)	ROI to ROI	Landmark-based linear registration	9 MS patients, 23 samples	9	Not reported	Within-subject	Not reported (likely linear regression)	+
Lehto et al. (2017)	ROI to ROI	No coregistration	21	21	12 and 2 depending on the analysis	Mixed (not modelled)	Pearson	-
Lodygensky et al. (2012)	ROI to ROI	Not reported	13 (across different ages)	13	2	Between-subject	Linear regression; Spearman mentioned but not reported	+

Continued on next page

Table 4: **Information on the employed statistical approach of the assessed validation studies.** Provided are information on the nature of data points that entered the correlation analysis, the method for coregistration between MRI and histology, sample size (number of subjects per group, total number of subjects, and number of data points for the analysis), the type of variance that was modelled in the statistical analysis, the reported statistic, and whether the study was selected for our meta-analysis (MA).

Reference	Correlation	Coregistration	N (Subjects/group)	N (Total subjects)	N (ROIs)	Statistical Design	Statistics	MA
Martirosyan et al. (2016)	ROI to ROI	No coregistration	15 from different types of SCI, 3 controls	18	1	Between-subject	Pearson	+
Moll et al. (2011)	ROI to ROI	No coregistration	4	4	12 (in total across subjects)	Mixed (not modelled)	Spearman	-
Mollink et al. (2019)	ROI to ROI	No coregistration	14 ALS and 5 Controls	19	1	Between-subject	Pearson	+
Mottershead et al. (2003)	ROI to ROI	No coregistration	4 MS patients, 1 control	5	108 (in total across subjects)	Mixed (not modelled)	Spearman	-
Oakden et al. (2015)	ROI to ROI	Manual coregistration	8 for acute group, 8 for chronic group	16	Variable and only specified for significant correlations	Mixed (not modelled)	Pearson	-
Odrobina et al. (2005)	unclear (likely ROI to ROI)	No coregistration	Not reported	Not reported	Not reported (likely 2)	Unclear (likely mixed, not modelled)	Not reported	-
Peters et al. (2019)	voxel-wise	3D registration	3	3 (1 reported)	not reported	Within-subject	Spearman	-
Pol et al. (2019)	ROI to ROI	No coregistration	13 Teliflunomide, 12 controls	25	3	Between-subject	Pearson	+
Praet et al. (2018)	ROI to ROI and voxelwise	3D stacking of histology, non-linear registration	48 Alzheimer's model, 32 wild type	80	1	Between-subject	Pearson (Bonferroni corrected for multiple comparisons); Bayesian multivariate linear regression	+
Pun et al. (2005)	unclear	Unclear	62 Tellerium, 34 controls	96	1	Between-subject	not reported	-
Reeves et al. (2015)	ROI to ROI	Manual coregistration	12 epilepsy, 1 control	13	3-4 (43 in total across subjects)	Mixed (not modelled)	Spearman; linear regression	-
Righart et al. (2017)	ROI to ROI	No coregistration	9	9	5 per patient; 36 in total across subjects	Mixed (not modelled)	Generalized estimating equations	-
Schmierer et al. (2010)	ROI to ROI	Manual coregistration	21	2	2	Mixed (not modelled)	Linear regression	-
Schmierer et al. (2004)	ROI to ROI	Manual coregistration	19	20	68 (in total across subjects)	Within-subject (mixed, modelled)	Pearson	+
Schmierer et al. (2007a)	ROI to ROI	Manual coregistration	35	35	129 (in total across subjects)	Within-subject (mixed, modelled)	Pearson	+

Continued on next page

Table 4: **Information on the employed statistical approach of the assessed validation studies.** Provided are information on the nature of data points that entered the correlation analysis, the method for coregistration between MRI and histology, sample size (number of subjects per group, total number of subjects, and number of data points for the analysis), the type of variance that was modelled in the statistical analysis, the reported statistic, and whether the study was selected for our meta-analysis (MA).

Reference	Correlation	Coregistration	N (Subjects/group)	N (Total subjects)	N (ROIs)	Statistical Design	Statistics	MA
Schmierer et al. (2007b)	ROI to ROI	No coregistration	16	16	44-51 depending on the analysis	Within-subject (mixed, modelled)	Pearson	-
Schmierer et al. (2008)	ROI to ROI	No coregistration	15	15	40-44 (in total across subjects)	Mixed (unclear whether modelled)	Linear regression	-
Schwartz et al. (2005)	ROI to ROI	No coregistration	3	3	6	Mixed (unclear whether modelled)	Pearson; linear regression	-
Seehaus et al. (2015)	voxel-wise	3D affine	1	1	voxel-wise (62782)	Within-subject	Pearson	-
Seewann et al. (2009)	ROI to ROI	Manual coregistration	10	10	42 (in total across subjects)	Mixed (modelled, but unclear whether result refers to within or between subject variance)	Pearson; Spearman mentioned but not reported	-
Soni et al. (2020)	ROI to ROI	No coregistration	36 TBI, 6 controls; correlation only for TBI group	36	3	Between-subject	Spearman	-
Soustelle et al. (2019)	ROI to ROI	Linear interpolation to common space	7 cuprizone-fed, 8 controls	15	3 (1/animal in each correlation)	Between-subject	Spearman	+
Stüber et al. (2014)	pixel-wise	Landmark-based registration	3	3	pixel-wise	Within-subject	Pearson; linear regression (univariate and multiple)	-
Sundberg et al. (2010)	ROI to ROI	No coregistration	15 injury, 9 control (two groups analysed separately)	15, 9 depending on the analysis	3	Mixed (not modelled)	Pearson	-
Takagi et al. (2009)	sample to sample	NA	10	70	1	Unclear	Pearson	-
Thiessen et al. (2013)	ROI to ROI	No coregistration	5 cuprizone, 5 controls	10	1	Between-subject (both within groups and pooled across groups) and within-group results have n $\leq$ 5 rather than n $\leq$ 10	Spearman	+
Tu et al. (2016)	ROI to ROI	No coregistration	5	25	7	Unclear (likely mixed, not modelled)	Pearson	-
Turati et al. (2015)	ROI to ROI	No coregistration	15 per mouse strain	15 for each analysis	1	Between-subject (pooled across groups)	Spearman	+

Continued on next page

Table 4: **Information on the employed statistical approach of the assessed validation studies.** Provided are information on the nature of data points that entered the correlation analysis, the method for coregistration between MRI and histology, sample size (number of subjects per group, total number of subjects, and number of data points for the analysis), the type of variance that was modelled in the statistical analysis, the reported statistic, and whether the study was selected for our meta-analysis (MA).

Reference	Correlation	Coregistration	N (Subjects/group)	N (Total subjects)	N (ROIs)	Statistical Design	Statistics	MA
Underhill et al. (2011)	ROI to ROI	No coregistration	5 and 4	9	1	Within-subject (mixed, modelled)	Pearson	+
van der Voorn et al. (2011)	ROI to ROI	Manual coregistration	15 patients, 5 controls	20	55 (in total across subjects)	Mixed (not modelled)	Pearson, linear regression analysis	-
van Tilborg et al. (2018)	unclear	Unclear	6	12	3	Between-subject	Linear regression	+
Wang et al. (2009)	ROI to ROI	No coregistration	15 at different ages	15	Unclear (likely 2)	Unclear (likely mixed, not modelled)	Pearson	-
Wang et al. (2015)	voxel-wise	Landmark-based linear coregistration (rigid body, based on 13 landmarks)	3	3	80 voxels (subset of total voxels selected for analysis)	Within-subject	Spearman	-
Warntjes et al. (2017)	pixel-wise	Manually coregistration by rotation, translation, and scaling	12	12	voxel-wise	Within-subject (mixed, modelled)	Spearman; linear regression	-
Wei et al. (2013)	ROI to ROI	3D stacking of histological images and linear coregistration	1	1	20	Within-subject	Pearson; linear regression	-
West et al. (2018)	unclear	NA	3 per group, 6 controls	15	unclear (likely 4)	Mixed (not modelled)	Linear regression	-
Yano et al. (2018)	ROI to ROI	No coregistration	18 cuprizone treated, 3 controls	21	1	Between-subject	Pearson	+

Table 5: Information on the statistical results of the assessed validation studies.

Reference	Modality	Result	Linear equation
Abe et al. (2019)	DWI	FA: r = .32 (p = .012), RD: r = -.41 (p = .001), AD: r = -.11 (p = .39)	NA
Aojula et al. (2016)	DWI	AD: r = .159, p = .541 (CC); r = -.360, p = 0.155 (PVWM); FA: r = .091, p = .729 (CC); r = .346, p = .174 (PVWM); MD: r = -.031, p = .903 (CC); r = -.495, p = .043 (PVWM); RD: r = -.115, p = .66 (CC); r = -.458, p = .064 (PVWM)	NA
Argyridis et al. (2014)	QSM	susceptibility: R2 = .93, p = .004	Susceptibility = -0.14 * myelin + 0.14
	DWI	MD: R2 = .003, p = .46	MD = -0.09 * myelin + 0.69
Bagnato et al. (2018)	Relaxometry	R2* and LFB: r = -.09, p < .01 (NAWM); r = -.05, p < .01 (DWMI); r = .99, p<.01 (WM-Ls); r = .83, p < .01 (thalamus); r = .82, p < .01 (dGM); shadow plaques: not significant; R2 = 0.014 (all ROIs combined); R2* and PLP: only CT reported, r = 0.57 p < .01	R2* = 0.02 * myelin
Beckmann et al. (2018)	MT	MTR: R2 = .7669 (p < .0001)	NA
Bot et al. (2004)	Relaxometry	T1: r = .71 (p < .001); T2: r = .77 (p < .001)	See paper
	MT	MTR: r = -.76, p < .001	See paper
Chandran et al. (2012)	DWI	FA: R2 = .27, p = .016 (LFB), R2 = .50, p = .0003 (MBP); RD: R2 = .34, p = .005 (MBP), R2 = 0.05; p = .31 (LFB)	not reported

Continued on next page

Table 5: Information on the statistical results of the assessed validation studies.

Reference	Modality	Result	Linear equation
Chang et al. (2017b)	DWI	AD: $p = .61$; FA: $p < .01$; RD: $p = .83$	Not reported
Chang et al. (2017a)	DWI	AD: $r = .0385$; FA: $r = .446$, $p < .05$; MD: $r = -.114$; RD: $r = -.195$	NA
Chen et al. (2017)	Relaxometry	MWF: $r = .823$, $p < .001$	NA
Choi et al. (2015)	DWI	FA: $p = 1.9 \ 10^{-5}$ (control), $p = .989$ and $p = .776$ (MPS); RD: $p < .0001$ (control), $p = .917$ and $p = .459$ (MPS)	Not reported
Duhamel et al. (2019)	MT	ih-MTR: between $r = .93$ and $r = .98$, $p < .0001$, depending on the specific sequence	See paper
Fatemi et al. (2011)	MT	MTR: $R2 = .695$, $p < .0001$; correlations with LFB not reported	not reported
Fjær et al. (2013)	MT	MTR: $R2 = .338$, $p < .0001$ (medial CC); $R2 = .426$, $p < .0001$ (lateral CC); $R2 = .208$, $p < .001$ (dGM); $R2 = .052$, $p = .1048$ (CT)	intercept = -3.002, slope 0.004 (medial CC); intercept -2.910, slope = 0.002 (lateral CC); intercept = -1.414, slope = 0.018 (dGM); intercept = -1.273, slope = 0.347 (CT)
Fjær et al. (2015)	MT	MTR and LFB: not reported. MTR and PLP: $R2 = 0.01$, $p = .63$ (CC), $R2 = .08$, $p = .2$ (dGM), $R2 < .01$, $p = .97$ (CT);	not reported
Gareau et al. (2000)	MT	MTR: $R2 = .45$, $p < .001$	NA
Grussu et al. (2017)	DWI	AD: $r = .75$, $p < .001$; FA: $r = .42$, $p = .025$; MD: $r = .67$; $p < .001$; RD: $r = .61$; $p < .001$	See paper
Hakkarainen et al. (2016)	Relaxometry	T1: $r = .77$; T2: $r = .18$; TRAFF 1: $r = .33$; T RAFF2: $r = .47$; T RAFF3: $r = .56$; T RAFF4: $r = .83$; T RAFF5: $r = .84$ (all $p < .001$)	NA
	MT	MTR: $r = .34$ ($p < .001$)	NA
Hametner et al. (2018)	Relaxometry	R2*: $r = .027$; T1 (log): $r = -.583$; for results separate for tissue class see paper	R2* = 44.7355 0.002942 * myelin; T1 (log) = 7.2738 0.001928 * myelin
	QSM	Susceptibility: $r = -.352$; for results separate for tissue class see paper	Susceptibility = 0.008758 0.000196 * myelin
Janve et al. (2013)	DWI	AD: $r = .03$ (lesion), $r = -.08$ (combined); FA: $r = .16$ (lesion), $r = .27$ (combined); RD: $r = -.40$ (lesion), $r = -.49$ (combined)	NA
	MT	MP: $r = .87$ (lesion), $r = .85$ (all)	NA
Jelescu et al. (2016)	DWI	RD: $r = -.71$, $p = .0004$; RK: $r = .14$, $p = .5638$	NA
	Relaxometry	T2: $r = -.64$, $p = .0024$	NA
	MT	MTR: $r = .32$, $p = .1821$	NA
Jespersen et al. (2010)	DWI	FA: $r = .78$, $p = 7 \ 10^{-8}$	NA
Jito et al. (2008)	DWI	FA: $r = .856$, $p < .0001$	not reported
Kelm et al. (2016)	DWI	FA: not significant; MD: $R2 = .35$; RD: $R2 = .37$; MK: $R2 = .48$; AK: not significant; RK: $r = .49$	NA
	Relaxometry	not reported	NA
	MT	not reported	NA
Khodanovich et al. (2017)	MT	MP: $R2 = .897$, $p < .001$ (all); $r = .870$, $p = .007$ (control); $r = .927$, $p = 0.002$ (Cuprizone)	MP = 0.124 * myelin + 6.278
Khodanovich et al. (2019)	MT	MP: $r = .90$, $p < .001$ (CC); $r = .83$, $p < .001$ (Caudate/putamen); $r = .80$, $p < .001$ (Hippocampus); $r = .88$, $p < .001$ (CT)	MP = 0.4-0.6 * myelin + 6.2-7.84 (slope and intercept vary depending on anatomical area)

Continued on next page

Table 5: Information on the statistical results of the assessed validation studies.

Reference	Modality	Result	Linear equation
Kozlowski et al. (2008)	DWI	AD: r = -.27, p = .07) for LFB at 3 weeks, r = .24 (p = .16) for LFB at 8 weeks, r = -.35 (p = .02) for MBP at 3 weeks, r = .03 (p = .88) for MBP at 8 weeks; FA: r = .33 (p = .02) for LFB at 3 weeks, r = .72 (p < .001) for LFB at 8 weeks, r = -.66 (p < .0001) for MBP at 3 weeks, r = .22 (p = .19) for MBP at 8 weeks; MD: r = -.60 (p < .001) for LFB at 3 weeks, r = -.57 (p < .001) for LFB at 8 weeks, r = .26 (p = .08) for MBP at 3 weeks, r = -.19 (p = .26) for MBP at 8 weeks; RD: r = -.49 (p < .001) for LFB at 3 weeks, r = -.71 (p < .001) for LFB at 8 weeks, r = .51 (p < .001) for MBP at 3 weeks, r = -.21 (p = .22) for MBP at 8 weeks	NA
	Relaxometry	MWF: r = .64 (p < .001) for LFB at 3 weeks, r = .88 (p < .001) for LFB at 8 weeks, r = -.28 (p = .07) for MBP at 3 weeks, r = .15 (p = .40) for MBP at 8 weeks	NA
Kozlowski et al. (2014)	DWI	not reported	NA
	Relaxometry	MWF: r = .63, p = .0047 (EC, in vivo), r = .74, p = .0001 (EC, ex vivo), r = .64, p = .05 (MBP, in vivo), r = .63, p = .038 (MBP, ex vivo)	NA
Laule et al. (2008)	Relaxometry	MWF: R2 between .56 and .95 (mean: .78), all p < .0001; R2 between 0 and .79 (mean R2 = .43, p < .0001) just in WM ROIs	NA
Laule et al. (2006)	Relaxometry	MWF: R2 between .45 and .92 (mean .67), all p < .0001; mean R2 = .29 (p < .0001) just in WM ROIs	NA
Laule et al. (2011)	Relaxometry	MWF and LFB: R2 between .48 and .95 (mean: .74); MWF and MBP: R2 between .15 and .47 (mean: .31)	NA
Lehto et al. (2017)	Relaxometry	RAFF4: r = -.742, p < .0001 (CC), r = -.745, p = .001 (DTT)	NA
	MT	MTR: r = .741, p < .001 (CC), r = .719, p = .001 (DTT); T1sat: r = -.741, p < .001 (CC), r = -.705, p = .001 (DTT)	NA
	DWI	AD: r = .714, p < .001 (CC), r = .279, p = .223 (DTT); FA: r = .662, p < .001 (C), r = .438, p = .053 (DTT); MD: r = .708, p < .001 (CC), r = .112, p = .610 (DTT); RD: r = .257, p = .048 (CC), r = -.079, p = .739 (DTT)	NA
Lodygensky et al. (2012)	QSM	Susceptibility: R2 = .673, p < .0003 (AC); R2 = .76, p < .0001 (CC)	Susceptibility = - 0.0107 * myelin - 0.0017 (AC); Susceptbility = - 0.015 * myelin + 0.004 (CC)
Martirosyan et al. (2016)	DTI	FA: r = -.94 (p = .0001)	NA
Moll et al. (2011)	DWI	AD: r = -.41 (p = .004); FA: r = .38 (p = .008); MD: r = -.58 (p < .0001), RD: r = -.52 (p = .0002)	NA
	MT	MTR: r = 0.63 (p < .0001)	NA
Mollink et al. (2019)	DWI	FA: r = .52, p = .03	NA
Mottershead et al. (2003)	Relaxometry	PD: r = -.72 (p < .001); T1: r = -.78 (p < .001); T2: r = -.75 (p < .001)	NA
	DWI	ADC: r = -.45 (p = .001); SDI: r = .51 (p < .001)	NA
	MT	MTR: r = -.29 (p = .075)	NA
Oakden et al. (2015)	DWI	AD: r = .39; RD: r = .57	NA
	Relaxometry	MWF: r = -.06	NA
Odrobina et al. (2005)	Relaxometry	MWF: r = .71 ± .16 (p < .001); T1: r = -.88 ± 12 (p < .0001); T2: r = -.91 ± .08 (p < .0001)	NA
	MT	MTR: r = .56 ± .22 (p < .0005); MP: r = .77 ± .15 (p < .0001)	NA
Peters et al. (2019)	DWI	FA: r = .10 (CT), r = .09 (Tuber), r = .45 (Perituber), r = .19 (WM), r = .70 (All); MD: r = .04 (CT), r = .07 (Tuber), r = -.63 (Perituber), r = -.60 (WM), r = -.85 (All)	NA

Continued on next page

Table 5: Information on the statistical results of the assessed validation studies.

Reference	Modality	Result	Linear equation
Pol et al. (2019)	DWI	FA: r = -.240, p = .675 (caudal medial, controls), r = -.130, p = .802 (rostral lateral, controls), r = -.147, p = .785 (rostral medial, controls), r = -.122, p = .825 (caudal medial, Teliflunomide), r = -.265, p = .643 (rostral lateral, Teliflunomide), r = .017, p = .981 (rostral medial, Teliflunomide); MD: r = .592, p = .202 (caudal medial, controls), r = .377, p = . 438 (rostral lateral, controls), r = .221, p = .675 (rostral medial, controls), r = -.083, p = .896 (caudal medial, Teliflunomide), r = -.237, p = . 667 (rostral lateral, Teliflunomide), r = .069, p = .903 (rostral medial, Teliflunomide);	NA
	QSM	Susceptibility: r = -.190, p = .723 (caudal medial, controls), r = -.214, p = .675 (rostral lateral, controls), r = .005, p = .987 (rostral medial, controls), r = .601, p = .141 (caudal medial, Teliflunomide), r = .243, p = .658 (rostral lateral, Teliflunomide), r = .211, p = .695 (rostral medial, Teliflunomide)	NA
Praet et al. (2018)	DWI	AD: r = -.479, p = .0362; AK: r = .376 (p = .4851); FA: r = -.560 (p = .0025); MD: r = -.175, p = 1; MK: r = .504 (p = .0173); RD: r = .272, p = 1; RK: r = .518 (p = .0110); Voxelwise statistics: DT: r = .5545, DK: r = .4509, DT/DK: .6374	Not reported
Pun et al. (2005)	Relaxometry	MWF: r = .77, T1: r = .88	NA
Reeves et al. (2015)	Relaxometry	T1: R2 = .270 (p < .005); T2: R2 = .235 (p < .005); T2*: p < .05	Not reported
	MT	MTR: p < .005	Not reported
Righart et al. (2017)	Other	NA	Not reported
Schmierer et al. (2010)	Relaxometry	T1: not reported	Not reported
	MT	MTR: r = .52, p = .02	Not reported
Schmierer et al. (2004)	Relaxometry	T1: r = .70, p < .001	NA
	MT	MTR: r = -.84, p < .001	NA
Schmierer et al. (2007a)	Relaxometry	T1: r = .69, p < .001	NA
	MT	MP: r = -.80, p < .001; MTR: r = -.84, p < .001	NA
Schmierer et al. (2007b)	DWI	FA: r = -.79, p < .001; MD: r = .68, p < .001	NA
Schmierer et al. (2008)	MT	MP: r = -.72 (unfixed) and r = -.86 (fixed), all p < .01; MTR: r = -.83 (unfixed) and r = -.68 (fixed)	Not reported
	Relaxometry	T1: r = .77 (unfixed) and r = .89 (fixed), T2: r = .82 (unfixed) and r = .92 (fixed), all p < .01	Not reported
	DWI	AD: r = .53 (unfixed) and r = .8 (fixed) ; FA: r = -.78 (unfixed) and r = -.83 (fixed); MD: r = .73 (unfixed) and r = .78 (fixed); RD: r = .74 (unfixed) and .81 (fixed), all p < .01	Not reported
Schwartz et al. (2005)	DWI	AI: r = -.58, p = .0125 (MVF) , r = 0.19, p = 0.4536 (MST); lADC: r = -.48, p = .0437 (MVF), r = .43, p = .0717 (MST); tADC: r = -.74, p = .0004 (MVF), r = .34, p = .1701 (MST)	Not reported
Seehaus et al. (2015)	DWI	FA: r = .487, MD: r = -.542, RD: r = -.554	NA
Seewann et al. (2009)	Relaxometry	T1 and LFB: r = .68, p < .01; T1 and PLP: r = .41, p < .05	NA
	DWI	ADC and LFB: r = .28; ADC and PLP: r = .16 ; FA and LFB: r = -.63, p < .01; FA and PLP: r = -.35, p < .05	NA
	MT	MTR and LFB: r = -.63, p < .01; MTR and PLP: r = -.27	NA
Soni et al. (2020)	QSM	susceptibility: R2 = .1, p = .02 (only reported for 1 out of 3 ROIs)	NA
Soustelle et al. (2019)	MT	MP: All: r = .87, p < .05 (medial CC), r = .91, p < .05 (lateral CC); r = .89, p < .05 (CT); Controls only: MP: r = .36 (medial CC); r = .55 (lateral CC); r = .86 (CT); Cuprizone only: r = .64 (medial CC); r = .82 (lateral CC); r = .25 (CT);	NA
	DWI	RD: All: r = -.7, p < .05 (medial CC), r = -.63, p > .05 (lateral CC); r = -.21, p > .05 (CT); Controls only: r = .38 (medial CC), r = .40 (lateral CC), r = .45 (CT); Cuprizone only: r = -.07 (medial CC), r = .14 (lateral CC), r = .79 (CT);	NA
	Other	rSPF: r = .86, p < .05 (medial CC), r = .76, p < .05 (lateral CC), r = .52, p > .05 (CT)	NA
	Relaxometry	MWF: r = .68, p < .05 (medial CC), r = .69, p < .05 (lateral CC), r = .20, p > .05 (CT)	NA

Continued on next page

Table 5: Information on the statistical results of the assessed validation studies.

Reference	Modality	Result	Linear equation
Stüber et al. (2014)	Relaxometry	R1/R2*: $p < .001$; R2*: $r = 0.805$	R1: slope: 1.066 ± 0.016, intercept: 1.132 ± 0.009; R2*: slope: 38.59 ± 0.50, intercept: 45.86 ± 0.29,
	QSM	Susceptability: $p < .001$	Susceptibility: slope: - 0.069 ± 0.002, intercept: - 0.025 ± 0.001
Sundberg et al. (2010)	DWI	AD: $r = .4825$, $p = .0216$ (lateral), $r = .4799$, $p = .0282$ (dorsal); FA: $r = -.8235$, $p = .0382$ (dorsal column of caudal zone);	NA
Takagi et al. (2009)	DWI	FA: $r = .2387$, $p = .4796$ (myelin sheath density); $r = -.6941$, $p = .0178$ (Myelin sheath thickness)	NA
Thiessen et al. (2013)	DWI	AD: $r = -.75$, $p < .001$ (MSF); $r = -.89$, $p < .01$ (MA); FA: $r = .84$, $p < .01$ (MSF), $r = .7$, $p < .001$ (MA); MD: $r = .81$, $p < .01$ (MSF); $r = .81$, $p < .01$ (MA); RD: $r = .87$, $p < .01$ (MSF); $r = -.77$, $p < .001$ (MA); for correlation separately for each group see paper	NA
	MT	MP: $r = .93$; $p < .001$ (MSF); $r = .81$, $p < .01$ (MA); MTR: $r = .72$, $p < .001$ (MSF); $r = .73$, $p < .001$ (MA); for correlation separately for each group see paper; Pearson: $r = 0.98$	Myelin = 5.37 * MP - 0.25
	Relaxometry	T1: $r = -.66$, $p < .001$ (MSF); $r = -.73$ (MA); for correlation separately for each group see paper	NA
Tu et al. (2016)	DWI	AD: $r = -.01$; FA: $r = .13$, MD: $r = -.17$, RD: $r = -.46$ ($p < .01$); MTR-3.5: $r = .38$ ($p < .01$), MTR-20: $r = .28$	NA
Turati et al. (2015)	MT	MP: in C57BL/6 mice: $r = .743$, $p = .002$ (BGII); $r = .589$, $p = .023$ (MBP); in SJL/J mice: $r = .586$, $p = .024$ (BGII) and $r = .536$, $p = .042$ (MBP)	NA
Underhill et al. (2011)	MT	MP: $r = .99$ ($p < .001$), $r = .91$, $p = .03$ (just GM), $r = .95$, $p = .0047$ (just WM); MTR: $r = .96$ ($p < .001$), $r = .93$, $p = .021$ (just GM), $r = -.21$, $p = .79$ (just WM); for results on four-parameter estimation of MP, see paper	MP = 0.21 * myelin + 3.9
van der Voorn et al. (2011)	DWI	ADC: $r = -.637$ ($p < .01$); FA: $r = .766$ ($p < .01$)	Myelin = -13.7 + 143.9 * FA
	MT	MTR: $r = .727$ ($p < .01$)	Not reported
van Tilborg et al. (2018)	DWI	FA: $R2 = .341$, $p = .046$ (M1 motor), $R2 = .334$, $p = .049$ (M2 motor), $R2 = .428$, $p = .021$ (sensory)	not reported
Wang et al. (2009)	DWI	FA: $r = .681$, $p < .01$; RD: $r = 0.528$, $p < .01$; AD and Trace showed no significant correlation	NA
Wang et al. (2015)	DWI	DBSI-RD: $r = -.84$ ($p < .0001$), $r = -.42$ ($p = .039$), $r = -.82$ ($p < .0001$)	Available for each subject - see paper
Warntjes et al. (2017)	Relaxometry	R1: $r = .63 \pm .12$; MWF: $r = 0.74 \pm 0.11$	R1: slope: 0.065 ± 0.009; intercept: 1.19 ± 0.10
Wei et al. (2013)	DWI	FA: $r = .12$, $p = 0.68$ (WM), $r = .61$, $p = .32$ (GM). RD: $r = .55$, $p = .15$ (WM), $r = .594$, $p = .28$ (GM)	Not reported
West et al. (2018)	Relaxometry	MWF: $R2 = .66$	MWF = .056 * myelin 0.014
	MT	MWF: $R2 = .70$	MWF = .056 * myelin 0.014
Yano et al. (2018)	DWI	FA: $r = .87$, $p = 3.89 \times 10^{-7}$; MD: $r = -.73$, $p = 1.54 \times 10^{-4}$; RD: $r = -.91$, $p = 1.04 \times 10^{-8}$	NA