Harmonization of multi-scanner *in vivo* magnetic resonance spectroscopy: ENIGMA consortium task group considerations

Ashley D. Harris[1,2], Houshang Amiri[3], Mariana Bento[2,4], Ronald Cohen[5], Christopher R. K. Ching[6], Christina Cudalbu[7,8], Emily L. Dennis[9], Arne Doose[10], Stefan Ehrlich[10], Ivan I. Kirov[11], Ralf Mekle[12], Georg Oeltzschner[13], Eric Porges[5], Roberto Souza[2,14], Friederike I. Tam[10], Brian Taylor[15], Paul M. Thompson[6], Yann Quidé[16], Elisabeth A. Wilde[9], John Williamson[5], Alexander P. Lin[17] and Brenda Bartnik-Olson[18]*

[1]Department of Radiology, University of Calgary, Calgary, AB, Canada, [2]Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada, [3]Neuroscience Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran, [4]Department of Biomedical Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada, [5]Department of Clinical and Health Psychology, College of Public Health and Health Professions, University of Florida, Gainesville, FL, United States, [6]Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, Los Angeles, CA, United States, [7]CIBM Center for Biomedical Imaging, Lausanne, Switzerland, [8]Animal Imaging and Technology, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [9]TBI and Concussion Center, Department of Neurology, University of Utah, Salt Lake City, UT, United States, [10]Division of Psychological and Social Medicine and Developmental Neurosciences, Faculty of Medicine, Technische Universität Dresden, Dresden, Germany, [11]Department of Radiology, Center for Advanced Imaging Innovation and Research, New York University Grossman School of Medicine, New York, NY, United States, [12]Center for Stroke Research Berlin, Charité-Universitätsmedizin Berlin, Berlin, Germany, [13]Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, MD, United States, [14]Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada, [15]Division of Diagnostic Imaging, Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [16]School of Psychology, University of New South Wales (UNSW), Sydney, NSW, Australia, [17]Center for Clinical Spectroscopy, Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States, [18]Department of Radiology, Loma Linda University Medical Center, Loma Linda, CA, United States

Magnetic resonance spectroscopy is a powerful, non-invasive, quantitative imaging technique that allows for the measurement of brain metabolites that has demonstrated utility in diagnosing and characterizing a broad range of neurological diseases. Its impact, however, has been limited due to small sample sizes and methodological variability in addition to intrinsic limitations of the method itself such as its sensitivity to motion. The lack of standardization from a data acquisition and data processing perspective makes it difficult to pool multiple studies and/or conduct multisite studies that are necessary for supporting clinically relevant findings. Based on the experience of the ENIGMA MRS work group and a review of the literature, this manuscript

provides an overview of the current state of MRS data harmonization. Key factors that need to be taken into consideration when conducting both retrospective and prospective studies are described. These include (1) MRS acquisition issues such as pulse sequence, RF and B0 calibrations, echo time, and SNR; (2) data processing issues such as pre-processing steps, modeling, and quantitation; and (3) biological factors such as voxel location, age, sex, and pathology. Various approaches to MRS data harmonization are then described including meta-analysis, mega-analysis, linear modeling, ComBat and artificial intelligence approaches. The goal is to provide both novice and experienced readers with the necessary knowledge for conducting MRS data harmonization studies.

Introduction

Magnetic resonance spectroscopy (MRS) is a non-invasive technique permitting detection and quantification of metabolites *in vivo*. Despite its potential to provide unique and useful diagnostic, prognostic, and biological information on healthy brain function and neurological and psychiatric disorders, MRS studies have generally been constrained by small sample sizes and poor reproducibility, limiting their interpretation and generalization. We suggest that this is one reason that MRS has not had the large-scale uptake of other MR research modalities such as functional magnetic resonance imaging (fMRI) or diffusion imaging. As MR research enters the information era, there is a growing need for big data analytics, robust pooling and harmonization of data across multiple sites. This paradigm shift motivated the establishment of The Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium in 2009 with the initial goal of identifying replicable associations between common genetic variants and brain structure and function ([1], [2]). The successes of this initiative are widespread with over 50 working groups, including the MRS working group, with recent publications pooling imaging data from over 2,000 scientists from 45 countries to achieve the largest neuroimaging sample sizes for multiple disorders including schizophrenia [total $N = 9,572$; 4,474 cases; ([3])], bipolar disorder [total $N = 12,000$; 3,500 cases; ([4])], major depressive disorder [total $N = 10,105$; 2,148 cases; ([5])], obsessive-compulsive disorder [total $N = 3,665$; 1,905 cases; ([6])], and 22q11.2 deletion syndrome [22q11DS; total $N = 944$; 474 cases; ([7]); also see Thompson et al. ([2])]. However, MRS research has not yet embraced data aggregation to this level and thus has not achieved studies with sample sizes of this magnitude. We believe this is largely due to difficulties in technical standardization and harmonization.

Multi-site studies offer an efficient way to increase sample size, which can be difficult at single sites due to a limited population for recruitment. Alternatively, when a single site has multiple scanners, it is possible that a single study will acquire data across multiple scanners for logistical reasons. Having multiple scanners at a single center also provides a controlled approach to investigate scanner effects as data from the same subjects, acquired by the same research team, can be acquired on multiple scanners relatively easily. A significant challenge with MRS is that even if there are two identical scanners, running the same software, operated by the same research team, there may be differences in their performance and ultimately metabolite quantification. Thus, for clarity, in this manuscript *multi-scanner* will indicate multiple individual scanners, which may be run at the same center, operated by the same team or at different locations. *Multi-vendor* will denote different scanner vendors (e.g., Canon, General Electric, Philips, and Siemens etc.) whether at the same location or multiple locations. *Multi-site* will refer to scanners at different locations that are operated by different teams and staff (so one institution may have multiple sites).

In this review, two approaches of data aggregation and multi-scanner studies will be considered: prospective and retrospective. Prospective development of multi-site/multi-vendor/multi-scanner MRS protocols should consider the technical factors and processing pipelines that will impact metabolite quantification. There are various approaches to this challenge, and each may be appropriate in different scenarios depending on the research question. One approach is to use the same scanner vendor and model, with the same software and the exact same parameters. However, this does not guarantee homogeneity of results as scanner characteristics such as eddy currents and frequency drift are dependent on the individual scanner ([8]). Further, it is likely not feasible

to include only sites with identical scanners, as software upgrades are performed with inconsistent timelines. A lowest common denominator approach, where all sites use the same acquisition, catering to the least advanced scanner, regardless of the manufacturer, will minimize sequence variability. Even so, there are differences within the most conventional sequences between vendors limiting data homogeneity (9). This approach also results in underuse of some scanners' capabilities. An alternative is to take advantage of the most advanced capabilities on each scanner, acquiring data with the highest possible quality on each scanner and then combine results as effects rather than raw data. This approach carries the risk of not achieving the desired advantage of pooling multiple sites for increased power if the data quality is too diverse across scanners. Ultimately, a goal of harmonization is a balance between data consistency, data quality and data quantity.

There is a wealth of MRS data, published and unpublished, which may be aggregated using retrospective study designs to deepen our understanding of the brain and disease processes while leveraging the power of MRS data for future studies. With retrospective aggregation of data, there is little control over data acquisition parameters. While a scientist designing a study with retrospective data aggregation may choose inclusion and exclusion criteria in an effort to have highly homogeneous data, highly restrictive criteria may be self-defeating in that very few studies can be included in the analysis. As such, there is a need to consider how both technical and methodological differences will be handled. Furthermore, aggregating retrospective data requires biological factors to be considered as well.

The goal of this review is to present the current state of MRS data harmonization including the unique technical/methodological and population/biological challenges associated with aggregating MRS data in both prospective and retrospective study designs (Figure 1). Considering the ENIGMA MRS working group's goal of developing data harmonization strategies, we will discuss several solutions employed to reduce these effects. Finally, we will present some interpretations and considerations related to these strategies and their application to study design in both prospective and retrospective multi-scanner, multi-site studies.

Technical factors

Unlike MR imaging where quantitative imaging is more recent, MR spectroscopy studies have always been inherently quantitative. Initially, and still practiced, metabolites were quantified as a ratio to another metabolite (often creatine). This was followed by water referencing, and more recently, water referencing with appropriate corrections for water and metabolite parameters (T_1, T_2, and proton density) yielding

absolute units based on the assumptions used (10, 11). One challenge with MRS quantification is that measurement variability depends on many factors, including hardware (scanner manufacturer, field strength, transmitter/receiver coils, passive shim elements impacting magnetic field inhomogeneities, and frequency drifts, etc.), acquisition parameters (localization method, echo time, repetition time, and voxel size, etc.), data quality signal to noise ratio (SNR), linewidth, and data analysis (pre-processing, fitting, and quantification). The next sections will overview how some of these factors impact harmonization of data between scanners, vendors and sites. For more in-depth background on MRS as a method, readers are directed to Oz et al. (12), Tomiyasu and Harada (13), Lin et al. (14), and Cecil (15).
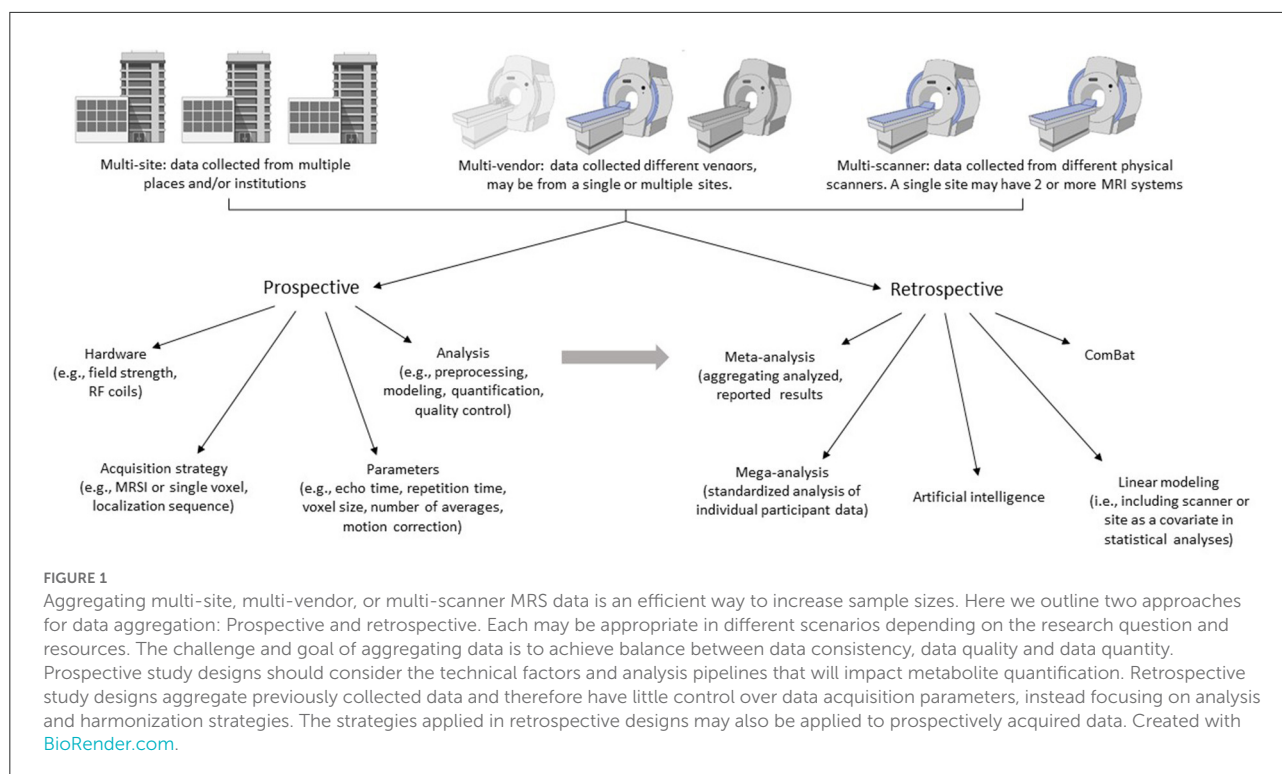
Hardware

Harmonization efforts may need to consider the available field strength and hardware. Different MRI scanners (even from the same vendor) have different RF coils used for transmit and receive. This is further complicated by many sites, particularly those at 7T, which use home-made RF coils. Additionally, the RF coil may impose additional limitations, such as maximum RF peak power. For prospective studies, scan protocols might have to be adjusted during study planning, so that they can be similarly executed on all types of participating hardware; however, this is challenging when aggregating data retrospectively.

Acquisition

Localization

One of the first decisions in performing an MRS study is the localization approach. For single voxel spectroscopy (SVS), traditionally the choice was between STimulated Echo Acquisition Mode [STEAM, (16)] or Point RESolved Spectroscopy [PRESS, (17)], and more recently spin echo full intensity acquired localized [SPECIAL (18)], localization by adiabatic selective refocusing [LASER (19)], and semi adiabatic localization by adiabatic selective refocusing [semiLASER/sLASER (20)]. The selection of localization method has multiple consequences including possible limitations on echo time, differences in chemical shift displacement and SNR. SemiLASER is recommended in a recent consensus and is available for General Electric, Siemens and Philips scanners (21). Even so, it remains a research tool which cannot always be implemented at clinical sites with clinical systems due to limited experience with MRS or institutional regulations preventing the installation of research-based (e.g., works in progress) acquisitions. In addition, the use of an advanced sequence, such as semi-LASER, requires sites to

FIGURE 1
Aggregating multi-site, multi-vendor, or multi-scanner MRS data is an efficient way to increase sample sizes. Here we outline two approaches for data aggregation: Prospective and retrospective. Each may be appropriate in different scenarios depending on the research question and resources. The challenge and goal of aggregating data is to achieve balance between data consistency, data quality and data quantity. Prospective study designs should consider the technical factors and analysis pipelines that will impact metabolite quantification. Retrospective study designs aggregate previously collected data and therefore have little control over data acquisition parameters, instead focusing on analysis and harmonization strategies. The strategies applied in retrospective designs may also be applied to prospectively acquired data. Created with BioRender.com.

obtain or generate basis sets for particular TE values, which can be difficult for novices. In contrast, STEAM and PRESS are available from all vendors as standard, product sequences. STEAM enables shorter echo times but has half the signal amplitude of PRESS (21). While PRESS and sLASER have comparable SNR at 3T, PRESS has on the order of ∼12%/ppm chemical shift displacement effects, compared to sLASER which shows 2%/ppm (21, 22).

Magnetic resonance spectroscopic imaging (MRSI) combines MRS with spatial encoding methods to map spectral information over multiple voxels. The choice of acquisition methods (TE, volume selection/excitation scheme, and phase encoding scheme) will depend primarily on the study-specific spatial and spectral information requirements, while taking into account MRSI's limitations (23). PRESS and sLASER sequences are most widely implemented, with sLASER favored at higher field strengths owing to large chemical shift displacement errors with PRESS at ≥3T (20). The long acquisition times associated with MRSI has limited its incorporation into many clinical studies; however, there is a growing body of work focused on accelerating scan times. More advanced MRSI acquisition techniques using spatial-spectral encoding techniques with echo-planar readouts [EPSI (24); PEPSI (25)] or other non-Cartesian trajectories (26–28) are available. In addition, accelerated MRSI techniques using shorter repetition times or k-space under sampling [sensitivity encoding (SENSE), generalized auto-calibrating partially parallel acquisition (GRAPPA), multi-band/simultaneous multi-slice] have been

developed [see Vidya Shankar et al. (29) and Bogner et al. (30) and references within]. These advancements remain largely in the research setting with limited implementation on clinical systems due to the need for specialized technical expertise to obtain good quality data. However, a recent consensus statement outlined vendor neutral recommendations for the acquisition, processing, and analysis of MRSI (23). As with SVS, even when the same parameters and acquisition approaches are used, substantial between-scanner differences can remain with MRSI data, this aggregating multi-site MRSI data will still need to account for differences between scanners (31).

Adjustments: RF calibration, shimming, water suppression calibration

There are multiple adjustments performed prior to the MRS acquisition that can have profound effects on data quality. These, in turn, may impact data harmonization. While many research centers use advanced methods to control these processes, many cannot be directly controlled by the operator. Thus, unless common vendors and systems are being used in a multi-scanner/multi-site study, the potential effects of differences in these adjustments may have to be accepted and their impact on data quality at a minimum recognized but likely controlled for in data analyses or aggregation steps. Each of these factors and their effect on data quality for harmonization will be briefly described here.

Radio frequency (RF) calibration is performed at the beginning of an exam and aims to optimize the RF power or RF voltage to yield the desired flip angle, typically with the aim of obtaining maximum signal (32). For SVS, this calibration is for the center of the voxel while the calibration is at the center of the volume of interest in MRSI. Performance of the RF calibration will ultimately affect sequence performance and data quality, particularly for PRESS localization. While most sites are not able to manipulate its efficacy, recognizing the impact on data quality may impact study design (i.e., choosing a single vendor) or data aggregation.

Linewidth directly impacts resolving overlapping resonances and complicated spectral patterns for reliable detection and quantification of metabolites. Narrow lines are a direct consequence of a homogeneous B**0** field, and in MRS, the required field homogeneity is generally higher than for most MR imaging applications resulting in an increased importance of shimming. Shimming generally refers to the process of (locally) removing macroscopic main magnetic field B0 inhomogeneities by applying small additional magnetic fields generated by shim coils. These are in addition to the passive shim elements that are also in place to develop a homogeneous main magnetic field and can impact field drift (described below). Often second-order shims are available on clinical systems, but this is not universal, with some systems having only first-order shims. Some research systems, particularly those at 7 T, have partial or full third-degree shims. While recommendations for SVS and MRSI shim approaches exist (33), customizing the shim methodology and hardware is generally not possible, particularly on clinical systems. Suggested criteria and corresponding linewidths have been defined for "excellent", "adequate" and "acceptable" categories for multiple field strengths and various anatomical regions (33) which can be used as a practical approach to define thresholds for acceptable field homogeneity for data harmonization across scanners and sites.

Finally, if water suppression (WS) is applied, both the WS method and its calibration can impact data quality and thus its harmonization. The goal of WS is to yield an MR spectrum in which the water signal is smaller than, or about the same in amplitude as the largest metabolite peak [NAA in the healthy brain (34)]. Alternatives to WS have been proposed [e.g., metabolite cycling, (35)], though they are largely still in development. On clinical systems, water suppression enhanced through T_1 effects (WET) with three or four CHemical Shift Selective (CHESS) WS pulses [WET, (36)], is the most commonly available WS method. A more elaborate method, VAriable Power and Optimized Relaxation Delays [VAPOR, (37)], uses a sequence of inversion pulses with varying inter-pulse delays and flip angles to minimize the water signal, although this approach is more available on research systems. Typically, the flip angle of a single WS pulse is varied during calibration to minimize the residual water

signal amplitude. As with all pre-scan adjustments, differences in WS methods should be recognized as a contributing factor to vendor effects during harmonization, since systematic differences in residual water amplitude can affect the spectral modeling process.

Acquisition parameters

Accurate quantification of metabolite concentrations or ratios ideally requires long repetition times (TR > 2,000 ms for $B_0 \leq 3T$) and short echo times (TE < 30 ms) to minimize signal loss due to T_1 and T_2 relaxation effects (38). Of these, TE will have the largest impact on the number of metabolites detected and spectral quality. Short-TE spectra have higher SNR and facilitate the detection of metabolites that have short T2 relaxation times and/or J-modulation due to J-couplings [e.g., glutamate, glutamine, and myo-inositol (mI)]. However, shorter TE spectra are complicated by broad high molecular weight macromolecules (MM) (39). Analysis methods are available to separate metabolites from lipid/macromolecule signals (see below), which can be employed in processing pipelines when aggregating both prospective and retrospective data sets, but the overlap between metabolites and MMs (as well as between coupled metabolites) remains a substantial source of measurement uncertainty. In contrast, longer TE (135, 144, and 288 ms) acquisitions can be used to reduce macromolecule signals, due to their short T_2 relaxation times. Optimizing the TE can also be used to optimize detection of a metabolite, for example, TE = 144 ms is used to improve the detection of lactate (40), which is inverted as a result of phase modulation by J-coupling. An alternative example is the recommendation of TE = 97 ms for the specific detection of 2-hydroxyglutarate using PRESS (41) or TE = 110 ms recommended for the detection of 2-hydroxyglutarate when using sLASER (42). While the choice of TE is typically determined by the study design (i.e., metabolite of interest) or the advantage of one technique over another (i.e., flatter baseline), the scanner gradient technology, and maximum peak B1 will constrain the minimum TE achievable on individual systems.

Two other important considerations are the voxel size and the number of averages (or number of transients) as these parameters both directly impact the SNR. When aggregating data from two studies that acquired different numbers of transients in the protocol, one approach may be to remove transients to match data acquisition parameters across scanners. However, deliberately reducing signal-to-noise and data quality is unsatisfactory, particularly if it impacts the net result (i.e., affects whether group differences are detected or not). Given the profound impact of voxel size on data quality, this is more challenging to harmonize directly. Alternatively, SNR and/or linewidth could be included as a covariate in analyses to mitigate the effects of data quality between sites or, in fact, between data

sets. An additional confound of having different voxel sizes is the inclusion of different anatomy, which is discussed below. Other acquisition parameter selections such as number of spectral points and spectral bandwidth have less obvious effects on data harmonization procedures.

Motion correction

Motion is an issue with all MRS studies; first it degrades data quality by affecting frequency and phase. Motion also will degrade the shim and water suppression which further impacts data quality. Finally, motion will change the voxel localization, which may not actually degrade spectral quality, but data will be from a different location than expected (mislocalization) which will impact the interpretation of results. While it is not possible to determine whether differences in data quality are a result of motion or other acquisition factors, the impact of motion on data harmonization needs to be considered. At a base level, differences in immobilization techniques (*via* participant instructions or head restraints) or periodic B0 field changes resulting from respiratory or cardiac motion in individual participants across sites can contribute to differences in the extent of motion artifact between sites, an issue that may exist in both prospective and retrospective data harmonization. To mitigate the effects of motion, both retrospective and prospective correction methods exist, and a recent consensus paper recommends the use of an advanced motion correction that provides at least 5% stability or better (43). For prospectively designed studies, novel prospective motion correction approaches using internal navigation methods may be used. These approaches use additional pulse sequence elements to discern the position and shim parameters. Alternatively, external tracking systems that employ external cameras to continuously track motion combined with real-time B0 shim correction can minimize the impact of motion. We direct readers to Andronesi et al. (43) and references within for a review of the most advanced and effective motion correction methods. However, challenges with implementing such methods and ensuring comparable methods exist at all sites may prevail and, in those instances, guidelines for acceptable data quality should be based on the question to be addressed. A number of recent consensus papers outline key quality metrics [SNR, metabolite and unsuppressed water resonance linewidth, residual water signal, line shape, Cramér-Rao lower bounds (CRLB) of the line fit of the metabolites of interest, fit quality (residuals of the fit) and the presence of artifacts] to be evaluated when determining spectral quality (21, 38). There has also been work on automated ^{1}H MRS quality control pipelines or combined approaches including both automated steps and visual spectra inspection as well as approaches with automated artifact pattern recognition for quality assessment (44–48).

Analysis

Regardless of the acquisition parameters, common analyses are desired for harmonization. Even this can be challenging as there are multiple vendor-proprietary MRS data formats that export data in varying levels of "rawness", e.g., all transients across all coils, coil combined, or in some cases, all transients are averaged prior to data export. The level of data rawness will impact both the degree of processing that can be applied as well as how common the analysis pipeline can be (for example, frequency-and-phase correction requires all transients), particularly in multi-vendor studies.

(Pre)Processing, modeling, and quantification

Modern MRS data analysis typically includes three key phases (11):

1) Data (pre-)processing, a series of operations preparing raw data into final spectra that are passed on to a modeling algorithm.
2) Modeling/fitting, the process of approximating/representing the measured signal.
3) Quantification, the process of converting model parameters into quantitative metabolite level estimates.

To accomplish these steps, a wide variety of data analysis software are available. While many widely used software packages such as LCModel (49) and jMRUI (50) have been distributed as compiled, closed-source binaries, there has been a notable shift toward modular open-source solutions (38, 51–53). LCModel, which has dominated MRS analysis, has also been made open access, but is not actively developed anymore. Many of these software packages have been compiled at https://www.mrshub.org.

Preprocessing

Recent MRS expert consensus has made several recommendations for preprocessing operations to maximize data quality, including receiver coil combination, alignment and removal of corrupted transients, eddy-current correction, etc., offering a template for harmonizing analysis workflows (11). However, there is no single comprehensive MRS data preprocessing pipeline that is universally agreed upon. As a result, MRS analysis workflows differ between research groups, and are frequently informed by personal preference and organic development over time, leading to incongruent results (54). While the effects of preprocessing pipeline diversity on metabolite estimates remain to be systematically investigated, it is difficult to imagine that this does not impact results and in turn affect the aggregation data across vendors at a minimum. Moreover, given the multitude of vendor-proprietary MRS file formats of varying levels of "rawness" developing a

comprehensive MRS data preprocessing pipeline across vendors has so far proved elusive. Most recently, a standardized MRS-specific data storage format ("NIfTI-MRS") and data/metadata accumulation logic ("BIDS-MRS") have been proposed, which should mitigate pipeline heterogeneity, aid in large-scale data aggregation, and improve integration with other neuroimaging modalities (55, 56).

Modeling

The diversity of spectral modeling methods is another substantial source of discrepancy and variability in MRS. Recent expert consensus recommends linear-combination modeling (LCM) for most scenarios, i.e., approximating the measured data as the weighted sum of (measured or simulated) basis functions for metabolite and macromolecular signals, which is reflected by the majority of fitting algorithms available and used. Most LCM algorithms (11, 44, 49, 51–53, 57–61) include parameters for signal amplitudes, zero- and first-order phase adjustments and small frequency shifts. They differ considerably, however, in the way they estimate smooth baseline fluctuations and line shape parameters and employ different strategies to solve the (typically ill-posed) optimization problem, e.g., different ways to search for optimal starting values. The (perhaps unsurprising) resulting sensitivity of metabolite level estimates to small differences in the modeling process has long been informally considered an "open secret" in the MRS community, but is increasingly studied (62–64). These types of discrepancies can impact retrospective analyses, particularly when performing meta-analyses as data is not analyzed using a common process.

In some cases, harmonization of the modeling process across-site and within-site is possible assuming transparent reporting of as many details as possible (65). For example, to account for small differences between MRS sequence implementations in multi-site, multi-vendor studies, basis sets should be simulated as accurately as possible for each scanner, specifically incorporating the actual RF pulse waveforms, gradient schemes, sequence timing, etc. (9, 53, 62, 66–69). However, even with common acquisition parameters (as common as possible) and using vendor-specific basis sets, it should be understood that substantial variability remains between vendors and scanners (9).

An additional challenge in modeling is handling the baseline and macromolecules (MM). Expert recommendation is to include an experimentally measured macromolecular spectrum in the basis set, as mathematical approximation (by splines or polynomials, as implemented in most fitting software packages) "does not completely reproduce the *in-vivo* spectral pattern" (39) particularly at higher field strength (70, 71). However, measuring the macromolecules *in vivo* is technically challenging and time consuming thus often not feasible. Alternatively, MM can be (and often are) simulated based on the available parameters including frequency in ppm (usually to 0.9 ppm),

linewidth, relative amplitude, number of components per resonance, etc. (39); however, this approach likely does not completely reproduce the *in vivo* spectral pattern. While this approach should also be useful for harmonization of multi-site studies (either prospective or retrospective), MM variance between scanners/vendors/sites is not well understood and still requires investigation.

Quantification

The final analysis step is the conversion of modeling parameters reflecting metabolite signal amplitudes into units representing concentrations. The most common approach is the use of an internal reference standard, i.e., a signal that is affected by the same proportionality factors including RF coil loading. For proton (^{1}H) brain MRS, this standard is typically the internal water signal or the total creatine signal (creatine + phosphocreatine), though other references such as total NAA (NAA + NAAG) have been used. Using the tissue water signal as the reference requires an additional short measurement without water suppression. While many simply report this water-referenced quantification [reported as institutional units (i.u.)] it does not represent actual tissue concentration (i.e., true molar or molal units). Recommended procedures are provided in a recent expert consensus paper (11) to correct for relaxation and proton density in different tissues to approximate true, absolute concentrations. This consensus paper also underlines that long TR and short TE reduce the importance of relaxation correction, while acknowledging that their choice is often determined by the study goals. Measurement of individual relaxation times and tissue water content is prohibitively long in most clinical and research scenarios, and literature values are therefore most commonly used (72–76). The use of literature values complicates matters, as relaxation times and water content are field-dependent, and may vary in healthy and pathologic tissues (77–81). Quantification methods using external signals also exist but are beyond the scope of this text.

Biological factors

Biological factors are crucial to consider in retrospective studies. In prospective study designs, effects of biology are generally considered *a priori* within the study designs, so are less of an issue for harmonization. For example, as metabolites and possibly macromolecules change with age, it is not possible to blindly combine data from different age ranges – this is also the case for many other imaging modalities. An additional challenge with MRS is that metabolite levels are linked to the tissue fraction in the voxel. Age-related atrophy therefore can impact metabolite levels and present as metabolite concentration decreases while in fact reflecting a decrease in tissue volume (82). Further, when considering the voxel and its anatomical

location, metabolite levels vary throughout the brain with metabolite differences seen between white and gray matter and different lobes (83). This is further complicated by differences in water signal and signal relaxation (both metabolites and water) between these regions. While these factors can be prospectively accounted for, the ability to correct them retrospectively is limited to what level of raw data is aggregated. As with the above-mentioned technical factors, many of these factors are not independent and may further interact or be addressed methodologically in data aggregation and will be outlined below.

Voxel placement and anatomy

The selection of anatomical targets (the voxel location) for a multisite study demands considerations beyond those of a single-scanner study. These considerations are compounded when using multiple platforms and vendors. While the scientific question at hand is paramount in selecting locations (i.e., prospective studies), voxel placement by those collecting data needs to be reliable and consistent. Typically, to ensure that data collection is consistent across sites, anatomical landmarks for voxel placement are recommended that are easy to identify on structural scans and require minimal anatomical sophistication among study staff. Common landmarks include midline, skull, genu of the corpus callosum and ventricles. As a guiding heuristic, manualized instructions that are pragmatic, with visual examples of each step involved in placement, should be provided to minimize judgments. Voxel rotations may present additional challenges, particularly across vendors that may differ in voxel manipulation flexibility, implementation and display of the voxel during placement and additional features such as the direction of the chemical shift displacement of the water signal (10, 38). This can result in the same procedure for voxel placement producing data collected in slightly different locations. While solutions to aid in reproducible placement of voxels across participants and sites such as SmartExam, AutoAlign (AA) scout, and AutoVOI (84) are available, they are not universally available and differ between vendors. Additionally, these tools need to be on all scanners to be useful for prospectively designed studies. Quantifying accuracy or voxel overlap between subjects can be difficult, as a voxel is generally of fixed volume, but intracranial volume (ICV) varies, so normalized voxels (e.g., using SPM, FSL, etc.) will produce voxels of varying sizes. This means that tools commonly used to quantify voxel overlap, e.g., the Sørensen-Dice Coefficient (85) or intensity images are limited in how high a value can be achieved, as they are evaluating overlap of areas that may no longer be the same size and are subject to the anatomical registration procedures used (i.e., selection of registration template and registration options such as linear, non-linear etc.). This does not mean that these tools should not be used, but rather to highlight their impact on interpretation.

Age

The development of the brain lasts from birth until adulthood and brain metabolite concentrations are affected by normal development and aging. MRS studies have demonstrated developmental, age-dependent regional changes in NAA, creatine (Cr), choline (Cho), mI, glutamate/glutamine (Glx), and GABA (86–91) with the most significant changes seen in the first year of life (92–97). In neonates and infants, early rapid increases in NAA and Cr correspond to structural and functional development resulting from synaptic formation and increases in energy demand. An increase in glutamate, Glx, and GABA neurotransmitters' concentrations are observed early in development that reflect increase in neuronal activity and the demands of the glutamate-glutamine-GABA neurotransmitter cycles. In contrast, mI decreases rapidly within the first 3 months (94, 96) while Cho decreases gradually over time becoming stable within childhood (96, 98), corresponding to the ongoing myelination and white matter development (99). By adolescence and adulthood, most metabolites reach stable concentrations for a period before decreasing slowly with advancing age. In aging, reductions in neuronal density and activity, reduced mitochondrial activity, and decreased membrane synthesis, are reflected by lower regional levels of NAA, ATP, and phosphomonoester, respectively (100). Aging is also associated with lower glutamate concentrations (101). However, it is challenging to separate voxel composition effects from changes in metabolite tissue-concentration. For example, decreases in GABA levels with age are a result of atrophy/tissue loss and it is likely that the concentration of GABA in the remaining tissue is not actually decreased (82). Age-dependent changes in the T_2 relaxation times - for example, those seen for NAA and Cr from adolescence with advancing age - will also impact metabolite level estimates (77, 80, 81). Small sample populations, the use of limited and/or partially overlapping age ranges, and the non-linear trajectories of metabolite changes over a lifespan (90, 96, 102) will prove challenging for harmonization. It will be essential that harmonization strategies evaluate the extent to which the treatment of age - either as a covariate or as a main variable of interest - is amenable to general linear models, or whether more sophisticated approaches are necessary.

Sex

There is substantial evidence of sex differences in brain structure and function in adults, which may be present throughout the entire lifespan or emerge during development (103–107). In contrast, sex differences in metabolite levels are not clearly understood. Within the current literature, sex effects are divergent and conflicting, due in part to small sample sizes,

differences in anatomical region and acquisition strategy, as well as the absence of absolute metabolite quantification.

In the adult population, both SVS and MRSI studies have reported sex differences in regional cortical gray matter and/or white matter for both NAA and myo-inositol (93, 108–113). However, other work shows no differences in metabolite levels between sexes in the healthy adult population (114–118). Similarly, sex differences in regional neurotransmitter concentrations have also been reported, with inconsistent findings (119, 120).

In the pediatric population, one study of children aged 6–15 years, reported no difference in metabolite levels (NAA, Cr, Cho, mI, lipid, and lactate) between males and females across multiple averaged regions including the frontal white matter, bilateral basal ganglia, bilateral hippocampi, and cerebellum (121). Similar results were also reported in a prior short-TE single-voxel PRESS study in children with cerebral palsy and normal development (122).

A number of studies also reported a significant interaction between age and sex, with differences in NAA/Cho during typical development and aging, with age-dependent alterations occurring earlier in males (123). Similarly, differences in NAA, Cr, and Cho levels (males < female) have been reported in the cingulate gyrus and supraventricular white matter in subjects aged 60–90 years using intermediate echo time (TE = 135 ms) MRSI (124–126). In addition, regional and sex variations in neurotransmitters appear to be affected by age, with faster rates of regional GABA decline in females (127) and faster rates of regional glutamate decline in males (128).

Given the considerable heterogeneity of metabolite concentration across all regions of the brain, which depend on tissue type (gray vs. white matter), subject age, and sex, it is imperative that comparisons between subjects or groups have comparative information for the same region, age, and sex, or at least covary for age and sex in the statistical analysis. Recent studies in neuropsychiatric disorders and diseases like Alzheimer's disease show substantial sex differences in prevalence and there are increasing reports of metabolic sex differences in these disorders that might also show in MRS studies (109, 129, 130).

Pathologic condition

A major goal of harmonization efforts will be toward establishing whether a pathologic condition is associated with different metabolism compared to the non-affected (healthy) brain. Therefore, factors related to both the pathology-affected and control cohort must be considered. Given the inability to cover all neurological conditions, we can use traumatic brain injury (TBI) as an example where multiple injury severities and mechanisms of injury fall within the TBI spectrum. Perhaps unsurprisingly, metabolite differences are seen across

the TBI spectrum and even within single categories. Moreover, these differences depend on other TBI-specific factors such as time-since-injury. Importantly, the definition of the control group may have substantial effects. For example, in mild TBI or concussion studies, non-head-injured controls are often described but this does not always exclude previous head-injury (either explicitly stated or not). Alternatively, to more specifically characterize the brain effects of concussion, orthopedic injury controls are increasingly being used as the control group. A recent study suggests that structural changes seen were not unique to concussion but common across concussion and orthopedic injury when comparing both to non-injured controls (131). Alternatively, local compared to whole brain differences are not always clear. Finally, in some conditions such as multiple sclerosis, or cancer, it is common to make measurements in the abnormal tissue and compare to a contralateral control region; however, the contralateral tissue may not be "normal" compared to a typical population either by the disease process or from treatment (i.e., chemotherapy, radiation therapy). Therefore, while this information is specific and relevant to individual studies, it will have an impact on appropriate use of data aggregation.

Approaches for harmonization

The factors relevant for the harmonization of MRS data have been presented above. In some cases, these factors can be considered and accounted for, particularly in prospective designs. However, the fact remains that for MRS, differences between scanners exist and some level of harmonization is required when aggregating multisite data within the framework of prospective or retrospective multi-scanner study designs.

Meta-analysis

Meta-analysis, which synthesizes summary statistics from previously published single site studies, has been used to aggregate and generalize MRS findings in an attempt to improve accuracy in drawing conclusions about a particular population. For example, while not comprehensive, meta-analyses of MRS data have been used to identify patterns of metabolite alterations in patients with multiple sclerosis (132), mild cognitive impairment, and Alzheimer's Disease (133), to understand variable results in traumatic brain injury (134) and evaluate the diagnostic potential of MRS in CNS tumors (135–137), characterize regional glutamate-glutamine and/or GABA in major depressive disorders (138, 139), schizophrenia (140), substance abuse (141), pain (142), and HIV (143, 144).

On the most basic level, in a meta-analysis effect sizes are aggregated and weighted by the sample size to provide an overall result across all studies included. Although meta-analyses are

typically conducted in adherence with the Preferred Reporting Items for Systematic Reviews and Meta-analysis statement (PRISMA) (145), MRS data are affected by methodological heterogeneity (both technical and biological), variations in data quality, and the statistical approach reported in the original papers. Recently, Peek et al. (142) developed a quality assessment tool, MRS-Q, to assess the quality of MRS data included in meta-analyses but data was not excluded for studies not meeting all criteria, as is common with quality assessments within a meta-analysis. While data quality may impact results, excluding studies that meet *a priori* defined inclusion criteria will also create bias.

In addition to aggregating primary effects, analyses such as meta-regressions and moderator analyses can provide insight as to the relevance of additional factors of interest within each effect. For MRS, these additional approaches provide an opportunity to assess whether technical or biological factors influence the results. For example, in the recent meta-analysis by Joyce et al. (134) metabolite differences with TBI were affected by year of publication, age and voxel location (anatomy) indicating multiple technical and biological factors impact MRS results in TBI. Thus, MRS findings from a meta-analysis can do more than provide aggregated evidence of metabolite effects, they can provide new insight to factors that should be considered and controlled for in future multisite studies whether prospective or retrospective.

Mega-analysis

The methodological diversity seen in meta-analysis can be minimized through the analysis of de-identified individual participant data (IPD) using an agreed-upon processing strategy. The use of IPD either as summary statistics or the raw data itself provides unique opportunities for data harmonization, defining processing pipelines and implementing a common analytical strategy. When raw IPD is accessible, single stage statistical mega-analyses can be performed by the coordinating facility. This is the design used in most ENIGMA studies with two key differences compared to traditional meta-analysis; (1) use of standardized analysis for all data which includes standardized quality control for data, and (2) use of random-effects models rather than fixed-effects to allow for variance across cohort, depending on the biology and demographics of the cohorts. This framework means that the effects across studies are considered to be from a distribution rather than simply weighting by sample size and the heterogeneity of effects can also be measured and reported. Statistical analyses using IPD offer several other advantages including improved consistency in inclusion criteria across sites, improved statistical power, and reduced biases (146). A number of comparison studies suggest more acceptable false positive rates and higher statistical power in mega-analysis with IPD when compared to meta-analyses

(146, 147). While mega-analyses are an attractive alternative to meta-analyses to harmonize analysis methodology, specific challenges to MRS remain. As above, significant differences in data can exist based on differences in acquisitions. Furthermore, depending on the data exported by different sites, some components of the analysis pipeline may not be feasible, for example retrospective frequency-and-phase correction assumes availability of individual transients whereas some data formats only include an average spectrum. A related issue is fitting and metabolite quantification can be impacted by data quality, in particular metabolites with low concentrations may have biased results. To address this issue, one approach is to include SNR and/or linewidth as covariates in the analysis.

Linear modeling

One of the most common approaches for raw data harmonization is to include the factors that need to be harmonized in the statistical model, for example controlling for site, vendor, age as covariates in an analysis of covariance model. The goal is to control for the variance introduced by each of these factors similarly as if they were effects of interest. The advantage of this harmonization approach is that it does not require additional tools and, in some cases, can be performed on summary data (i.e., individual metabolite levels) rather than requiring the raw data for reprocessing or IPD. While it also allows for the inclusion of multiple factors to harmonize such as site, vendor, field strength, age, sex etc. depending on the sample size, power limitation will exist and there must be sufficient sample sizes from each group (factor to harmonize across) as well as across all aggregated data. The disadvantage is a risk of collinearities, particularly in retrospective data aggregation (e.g., if different sites collected data in different voxel locations) and the risk of variance being partitioned in an unexpected manner resulting in incomplete or inefficient data harmonization.

A specific application of harmonization that uses linear regression is to harmonize white matter and gray matter across the whole brain data for MRSI data. An advantage of MRSI is the multi-voxel data can be leveraged to yield partial volume-corrected mean gray matter and white matter concentrations (148). Thus, it is reasonable to strive to integrate this considerable strength in efforts to harmonize MRSI data. Using various approaches based on linear regression or least-squares optimization, average gray matter and white matter concentrations per metabolite can be obtained (148) and these global gray matter or white matter metabolite concentrations may assist to harmonize data rather than these different tissues influencing results. The main caveat of the approach is that it assumes relative homogeneity of metabolite levels in the gray matter and white matter within the MRSI volume and, by extension, only changes that occur diffusely in either (or both) tissue types will be detected.

ComBat

More recently, alternate harmonization procedures have been proposed in imaging and one that is gaining popularity is ComBat. ComBat is an empirical Bayesian method in which (a) factors to harmonize (e.g., sites) and (b) variance to preserve (e.g., biological group differences) are defined (149–151). Briefly, terms to preserve and to harmonize are defined in a multivariate linear mixed-effects regression. This linear model then estimates the mean and the variance for these terms. Using Bayesian methods, factors of interest along with model parameters are then estimated with the removal of effects from factors to harmonize. Removal of unwanted variance, such as that from site effects, is thought to improve power to detect differences of interest, for example biological differences between groups. ComBat was originally developed to harmonize gene expression data (149) but has shown efficacy in removing unwanted variance in MR data, specifically structural analyses of both diffusion tensor data (150), cortical thickness (151–153), and functional data (154). ComBat has recently shown promise for MRS data harmonization and it offers the advantage of maintaining meaningful quantification of metabolite levels (155). While ComBat is an increasingly common approach to harmonize MRI data, and is therefore expected to gain use in MRS, it is largely limited to the assumption that linear or quadratic models are sufficient to explain data variance, though extensions for higher order distributions are being developed.

Artificial intelligence

Deep Learning (DL) has been used for brain MRI analysis to minimize the influence of the varying acquisition scanner/parameters, etc., known as data/domain shift or batch effects (156) and therefore AI holds unexplored potential for MRS data harmonization. DL models learn the relevant features to solve a problem from the raw data (i.e., MR images). Still, its black-box nature means it can easily pick up on nuances of the data that are not present across data points in multi-scanner, multi-vendor studies. In DL, Domain Adaptation (DA) strategies have been proposed for data harmonization, allowing the generalization of DL models across different datasets by avoiding domain specific decision making (157). DL models with consistent results across different imaging types, different groups of subjects (varying age ranges, sex, and pathology), and acquisition parameters are more reliable, allowing broader usage and facilitating the translation to clinical practice. DA strategies can unlearn not just batch effects but also other potential confounders in the data, such as sex and age (158), leading to domain-invariant features and consequent data harmonization (159).

Another common DA strategy is to use adversarial models (160) that can learn to transform samples into a harmonized reference space (161). While most current DL work on data harmonization in the medical imaging literature focuses on MR imaging, these DA strategies can be translated to MRS. The most significant challenges of deploying DA strategies in MRS are the necessity of having sufficient data, which often needs to be labeled to train these models, and hyper-parameter tuning of such models, which can be challenging and time-consuming (162). For retrospective studies, the data availability constraint is less critical, but for prospective studies, it is crucial. One potential solution to harmonize data prospectively is to leverage legacy data from past studies to initialize the corresponding DL models and use zero-shot learning DA techniques (163).

Discussion/interpretations

Aggregating imaging data has provided great insight along with new directions of research. Given these successes, and the abundance of spectroscopy data available, aggregation of MRS data is a relatively unexplored area with great potential as there is every reason to expect similar advances can be achieved by aggregating spectroscopy data as seen with other neuroimaging modalities. However, there is a lack of established approaches for effective harmonization beyond the traditional meta-analysis approach used in systematic reviews. While valuable, meta-analysis approaches come with many limitations. For example, the impact of non-biological sources of variance (e.g., scanner, site, vendor, etc.) are generally not investigated, though recent work shows technical or non-biological factors effects can be identified (134) suggesting more sophisticated data harmonization is required to most accurately aggregate data for informative analyses.

One prospective approach is to minimize differences in acquisition between sites, which was successfully demonstrated by Deelchand et al. (164). However, there is broad acceptance in the field that technical differences exist between scanners in detail even when using nominally comparable acquisition approaches. This has been confirmed in recent multi-site/multi-scanner studies (9, 165) that aimed to mitigate site effects by using common acquisition parameters (to the extent possible) but still show substantial site effects. Given the significant site-effects reported in these studies, it remains unclear for prospective study designs whether the most suitable approach is to harmonize all acquisition parameters, selecting a lowest common denominator approach in terms of acquisition technologies, or whether each site should use their "best" methodologies available to have the most precise measures while accepting different sites will have different data quality and measurement precision. The solution to this challenge will be study-dependent and will also depend on factors such as the study objectives and cohorts being studied, the diversity of scanners and hardware, study duration and timeline including the potential for what upgrades are likely to occur

during the study. We expect in most cases, investigators will opt for a hybrid approach (harmonize some hardware such as field strength and head coil but optimize acquisition parameters) aiming for a balance between maximum data quality while minimizing the risk of not being able to effectively combine data.

Beyond prospective harmonization, retrospective harmonization remains an area with little exploration for MRS. To our knowledge, the only work in this area compared ComBat and residualization to harmonize MRS data collected in controls across multiple sites and showed these methods are effective in removing non-biological sources of variance and may reveal biological factors with the augmented power achieved with large sample sizes that have been harmonized (155). The development and validation of effective retrospective harmonization procedures has multiple applications; first to augment the success of data aggregation from prospective study designs and to facilitate the aggregation of data collected through independent studies. There has been an increase in prospective multisite studies that include MRS (9, 166–173) which will likely benefit from additional data harmonization. Of great potential are the analyses and applications of aggregated data from previous studies, including mega-analyses of data from clinical groups, aggregation of normative data to best control for multi-scanner effects and aggregating data for artificial intelligence approaches. Included in this data aggregation is the possibility of developing fully automated, open access analysis pipelines that are cloud-based, or containerized, to facilitate consistent analyses across multiple users, scanners and sites.

In conclusion, MRS research, and by extension the clinical conditions being studied, has the potential to greatly benefit from data aggregation and harmonization. As outlined here, we define data harmonization to occur either prospectively at the time of study design or retrospectively, though retrospective approaches can be applied to prospectively acquired multi-scanner studies. For effective MRS data harmonization, we have defined two broad categories that impact data harmonization:

(1) Technical/methodological factors that consider the scanner, its hardware and software as well as acquisition and data processing.
(2) Biological factors that depend on the participants, such as their age, sex and anatomy being studied.

Overall, harmonization of MRS data facilitates the development of large sample sizes possibly yielding normative metabolite levels for different brain regions and various diseases that enables investigation of multiple interacting factors and the detection of subtle effects. This unlocked potential will further assist the transition of MRS from a predominantly research-based tool to clinical routine.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Thompson PM, Jahanshad N, Ching CRK, Salminen LE, Thomopoulos SI, Bright J, et al. ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl Psychiatry.* (2020) 10:1–28. doi: 10.1016/j.biopsych.2020.02.167

2. Thompson PM, Jahanshad N, Schmaal L, Turner JA, Winkler AM, Thomopoulos SI, et al. The enhancing neuroimaging genetics through meta-analysis consortium: 10 years of global collaborations in human brain mapping. *Hum Brain Mapp.* (2021) 43:15–22. doi: 10.1002/hbm.25672

3. van Erp TGM, Walton E, Hibar DP, Schmaal L, Jiang W, Glahn DC, et al. Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. *Biol Psychiatry.* (2018) 84:644–54. doi: 10.1016/j.biopsych.2018.04.023

4. Ching CRK, Hibar DP, Gurholt TP, Nunes A, Thomopoulos SI, Abé C, et al. What we learn about bipolar disorder from large-scale neuroimaging: Findings and future directions from the ENIGMA Bipolar Disorder Working Group. *Hum Brain Mapp.* (2022) 43:56–82. doi: 10.1002/hbm.25098

5. Schmaal L, Hibar DP, Sämann PG, Hall GB, Baune BT, Jahanshad N, et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol Psychiatry.* (2017) 22:900–9. doi: 10.1038/mp.2016.60

6. Boedhoe PSW, Schmaal L, Abe Y, Alonso P, Ameis SH, Anticevic A, et al. Cortical abnormalities associated with pediatric and adult obsessive-compulsive disorder: findings from the ENIGMA obsessive-compulsive disorder working group. *AJP.* (2018) 175:453–62. doi: 10.1176/appi.ajp.2017.17050485

7. Sun D, Ching CRK, Lin A, Forsyth JK, Kushan L, Vajdi A, et al. Large-scale mapping of cortical alterations in 22q112 deletion syndrome: Convergence with idiopathic psychosis and effects of deletion size. *Mol Psychiatry.* (2020) 25:1822–34. doi: 10.1038/s41380-018-0078-5

8. Hui SCN, Mikkelsen M, Zöllner HJ, Ahluwalia V, Alcauter S, Baltusis L, et al. Frequency drift in MR spectroscopy at 3T. *Neuroimage.* (2021) 241:118430. doi: 10.1016/j.neuroimage.2021.118430

9. PovaŽan M, Mikkelsen M, Berrington A, Bhattacharyya PK, Brix MK, Buur PF, et al. Comparison of Multivendor Single-Voxel MR Spectroscopy Data Acquired in Healthy Brain at 26 Sites. *Radiology.* (2020) 295:171–80. doi: 10.1148/radiol.2020191037

10. Kreis R, Boer V, Choi I-Y, Cudalbu C, de Graaf RA, Gasparovic C, et al. Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: Background and experts' consensus recommendations. *NMR Biomed.* (2020) 34:e4347. doi: 10.1002/nbm.4347

11. Near J, Harris AD, Juchem C, Kreis R, Marjańska M, Öz G, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR Biomed.* (2021) 34:e4257. doi: 10.1002/nbm.4257

12. Oz G, Alger JR, Barker PB, Bartha R, Bizzi A, Boesch C, et al. Clinical proton MR spectroscopy in central nervous system disorders. *Radiology.* (2014) 270:658–79. doi: 10.1148/radiol.13130531

13. Tomiyasu M, Harada M. In vivo human MR spectroscopy using a clinical scanner: development, applications, and future prospects. *Magn Reson Med Sci.* (2022) 21:235–52. doi: 10.2463/mrms.rev.2021-0085

14. Lin A, Tran T, Bluml S, Merugumala S, Liao H-J, Ross BD. Guidelines for acquiring and reporting clinical neurospectroscopy. *Semin Neurol.* (2012) 32:432–53. doi: 10.1055/s-0032-1331814

15. Cecil KM. Proton magnetic resonance spectroscopy: technique for the neuroradiologist. *Neuroimaging Clin N Am.* (2013) 23:381–92. doi: 10.1016/j.nic.2012.10.003

16. Frahm J, Bruhn H, Gyngell ML, Merboldt KD, Hänicke W, Sauter R. Localized proton NMR spectroscopy in different regions of the human brain in vivo. Relaxation times and concentrations of cerebral metabolites. *Magn Reson Med.* (1989) 11:47–63. doi: 10.1002/mrm.1910110105

17. Bottomley PA. Spatial localization in NMR spectroscopy in vivo. *Ann N Y Acad Sci.* (1987) 508:333–48. doi: 10.1111/j.1749-6632.1987.tb32915.x

18. Mlynárik V, Gambarota G, Frenkel H, Gruetter R. Localized short-echo-time proton MR spectroscopy with full signal-intensity acquisition. *Magn Reson Med.* (2006) 56:965–70. doi: 10.1002/mrm.21043

19. Garwood M, DelaBarre L. The return of the frequency sweep: designing adiabatic pulses for contemporary NMR. *J Magn Reson.* (2001) 153:155–77. doi: 10.1006/jmre.2001.2340

20. Scheenen TWJ, Klomp DWJ, Wijnen JP, Heerschap A. Short echo time 1H-MRSI of the human brain at 3T with minimal chemical shift displacement errors using adiabatic refocusing pulses. *Magn Reson Med.* (2008) 59:1–6. doi: 10.1002/mrm.21302

21. Öz G, Deelchand DK, Wijnen JP, Mlynárik V, Xin L, Mekle R, et al. Advanced single voxel 1H magnetic resonance spectroscopy techniques in humans: Experts' consensus recommendations. *NMR Biomed.* (2020) 34:e4236. doi: 10.1002/nbm.4236

22. Deelchand DK, Kantarci K, Öz G. Improved localization, spectral quality, and repeatability with advanced MRS methodology in the clinical setting. *Magn Reson Med.* (2018) 79:1241–50. doi: 10.1002/mrm.26788

23. Maudsley AA, Andronesi OC, Barker PB, Bizzi A, Bogner W, Henning A, et al. Advanced magnetic resonance spectroscopic neuroimaging: Experts' consensus recommendations. *NMR Biomed.* (2021) 34:e4309. doi: 10.1002/nbm.4309

24. Posse S, DeCarli C, Le Bihan D. Three-dimensional echo-planar MR spectroscopic imaging at short echo times in the human brain. *Radiology.* (1994) 192:733–8. doi: 10.1148/radiology.192.3.8058941

25. Lin F-H, Tsai S-Y, Otazo R, Caprihan A, Wald LL, Belliveau JW, et al. Sensitivity-encoded (SENSE) proton echo-planar spectroscopic imaging (PEPSI) in the human brain. *Magn Reson Med.* (2007) 57:249–57. doi: 10.1002/mrm.21119

26. Adalsteinsson E, Star-Lack J, Meyer CH, Spielman DM. Reduced spatial side lobes in chemical-shift imaging. *Magn Reson Med.* (1999) 42:314–23. doi: 10.1002/(SICI)1522-2594(199908)42:2<314::AID-MRM14>3.0.CO;2-X

27. Schirda CV, Zhao T, Andronesi OC, Lee Y, Pan JW, Mountz JM, et al. In vivo brain rosette spectroscopic imaging (RSI) with LASER excitation, constant gradient strength readout, and automated LCModel quantification for all voxels. *Magn Reson Med.* (2016) 76:380–90. doi: 10.1002/mrm.25896

28. Furuyama JK, Wilson NE, Thomas MA. Spectroscopic imaging using concentrically circular echo-planar trajectories in vivo. *Magn Reson Med.* (2012) 67:1515–22. doi: 10.1002/mrm.23184

29. Vidya Shankar R, Chang JC, Hu HH, Kodibagkar VD. Fast data acquisition techniques in magnetic resonance spectroscopic imaging. *NMR Biomed.* (2019) 32:e4046. doi: 10.1002/nbm.4046

30. Bogner W, Otazo R, Henning A. Accelerated MR spectroscopic imaging-a review of current and emerging techniques. *NMR Biomed.* (2021) 34:e4314. doi: 10.1002/nbm.4314

31. Sabati M, Sheriff S, Gu M, Wei J, Zhu H, Barker PB, et al. Multivendor implementation and comparison of volumetric whole-brain echo-planar MR spectroscopic imaging. *Magn Reson Med.* (2015) 74:1209–20. doi: 10.1002/mrm.25510

32. Versluis MJ, Kan HE, van Buchem MA, Webb AG. Improved signal to noise in proton spectroscopy of the human calf muscle at 7 T using localized B1 calibration. *Magn Reson Med.* (2010) 63:207–11. doi: 10.1002/mrm.22195

33. Juchem C, Cudalbu C, de Graaf RA, Gruetter R, Henning A, Hetherington HP, et al. B0 shimming for in vivo magnetic resonance spectroscopy: Experts' consensus recommendations. *NMR Biomed.* (2021) 34:e4350. doi: 10.1002/nbm.4350

34. Tkáč I, Deelchand D, Dreher W, Hetherington H, Kreis R, Kumaragamage C, et al. Water and lipid suppression techniques for advanced 1 H MRS and MRSI of the human brain: Experts' consensus recommendations. *NMR Biomed.* (2021) 34:e4459. doi: 10.1002/nbm.4459

35. Dreher W, Leibfritz D. New method for the simultaneous detection of metabolites and water in localized in vivo 1H nuclear magnetic resonance spectroscopy. *Magn Reson Med.* (2005) 54:190–5. doi: 10.1002/mrm.20549

36. Ogg RJ, Kingsley RB, Taylor JS, WET. a T1- and B1-Insensitive Water-Suppression Method for in Vivo Localized 1H NMR Spectroscopy. *J Magn Reson.* (1994) 104:1–10. doi: 10.1006/jmrb.1994.1048

37. Tkáč I, Starcuk Z, Choi IY, Gruetter R. In vivo 1H NMR spectroscopy of rat brain at 1 ms echo time. *Magn Reson Med.* (1999) 41:649–56. doi: 10.1002/(SICI)1522-2594(199904)41:4<649::AID-MRM2>3.0.CO;2-G

38. Wilson M, Andronesi O, Barker PB, Bartha R, Bizzi A, Bolan PJ, et al. Methodological consensus on clinical proton MRS of the brain: Review and recommendations. *Magn Reson Med.* (2019) 82:527–50. doi: 10.1002/mrm.27742

39. Cudalbu C, Behar KL, Bhattacharyya PK, Bogner W, Borbath T, de Graaf RA, et al. Contribution of macromolecules to brain 1 H MR spectra: Experts' consensus recommendations. *NMR Biomed.* (2021) 34:e4393. doi: 10.1002/nbm.4393

40. Ross B, Bluml S. Magnetic resonance spectroscopy of the human brain. *Anat Rec.* (2001) 265:54–84. doi: 10.1002/ar.1058

41. Choi C, Ganji SK, DeBerardinis RJ, Hatanpaa KJ, Rakheja D, Kovacs Z, et al. 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. *Nat Med.* (2012) 18:624–9. doi: 10.1038/nm.2682

42. Berrington A, Voets NL, Larkin SJ, de Pennington N, Mccullagh J, Stacey R, et al. A comparison of 2-hydroxyglutarate detection at 3 and 7 T with long-TE semi-LASER. *NMR Biomed.* (2018) 31:e3886. doi: 10.1002/nbm.3886

43. Andronesi OC, Bhattacharyya PK, Bogner W, Choi I-Y, Hess AT, Lee P, et al. Motion correction methods for MRS: experts' consensus recommendations. *NMR Biomed.* (2021) 34:e4364. doi: 10.1002/nbm.4364

44. Vanhamme L, van den Boogaart A, Van Huffel S. Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge. *J Magn Reson.* (1997) 129:35–43. doi: 10.1006/jmre.1997.1244

45. Wright AJ, Arús C, Wijnen JP, Moreno-Torres A, Griffiths JR, Celda B, et al. Automated quality control protocol for MR spectra of brain tumors. *Magn Reson Med.* (2008) 59:1274–81. doi: 10.1002/mrm.21533

46. Hernández-Villegas Y, Ortega-Martorell S, Arús C, Vellido A, Julià-Sapé M. Extraction of artefactual MRS patterns from a large database using non-negative matrix factorization. *NMR Biomed.* (2019) 34:e4193. doi: 10.1002/nbm.4193

47. Menze BH, Kelm BM, Weber M-A, Bachert P, Hamprecht FA. Mimicking the human expert: pattern recognition for an automated assessment of data quality in MR spectroscopic images. *Magn Reson Med.* (2008) 59:1457–66. doi: 10.1002/mrm.21519

48. Pedrosa de Barros N, McKinley R, Knecht U, Wiest R, Slotboom J. Automatic quality control in clinical (1)H MRSI of brain cancer. *NMR Biomed.* (2016) 29:563–75. doi: 10.1002/nbm.3470

49. Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn Reson Med.* (1993) 30:672–9. doi: 10.1002/mrm.1910300604

50. Stefan D, Di Cesare F, Andrasescu A, Popa E, Lazariev A, Vescovo E, et al. Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package. *Meas Sci Technol.* (2009) 20:104035. doi: 10.1088/0957-0233/20/10/104035

51. Clarke WT, Stagg CJ, Jbabdi S, FSL-MRS. An end-to-end spectroscopy analysis package. *Magn Reson Med.* (2021) 85:2950–64. doi: 10.1002/mrm.28630

52. Oeltzschner G, Zöllner HJ, Hui SCN, Mikkelsen M, Saleh MG, Tapper S, et al. Osprey: Open-source processing, reconstruction & estimation of magnetic resonance spectroscopy data. *J Neurosci Methods.* (2020) 343:108827. doi: 10.1016/j.jneumeth.2020.108827

53. Soher BJ, Semanchuk P, Todd D, Steinberg J, Young K. VeSPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis. In: *Proceedings International Social Magnetic Resonance Medicine* (2011).

54. Bhogal AA, Schür RR, Houtepen LC, van de Bank B, Boer VO, Marsman A, et al. 1 H-MRS processing parameters affect metabolite quantification: The urgent need for uniform and transparent standardization. *NMR Biomed.* (2017) 30:e3804. doi: 10.1002/nbm.3804

55. Clarke WT, Bell TK, Emir UE, Mikkelsen M, Oeltzschner G, Shamaei A, et al. NIfTI-MRS: A standard data format for magnetic resonance spectroscopy. *Magn Reson Med.* (2022) 88:2358–70. doi: 10.1002/mrm.29418

56. Soher BJ, Clarke WT, Wilson M, Near J, Oeltzschner G. Community-organized resources for reproducible MRS data analysis. *Magn Reson Med.* (2022) 88:1959–61. doi: 10.1002/mrm.29387

57. Wilson M, Reynolds G, Kauppinen RA, Arvanitis TN, Peet AC. A constrained least-squares approach to the automated quantitation of in vivo ^{1}H magnetic resonance spectroscopy data. *Magn Reson Med.* (2011) 65:1–12. doi: 10.1002/mrm.22579

58. Wilson M. Adaptive baseline fitting for 1 H MR spectroscopy analysis. *Magn Reson Med.* (2021) 85:13–29. doi: 10.1002/mrm.28385

59. Poullet J-B, Sima DM, Simonetti AW, De Neuter B, Vanhamme L, Lemmerling P, et al. An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES. *NMR Biomed.* (2007) 20:493–504. doi: 10.1002/nbm.1112

60. Chong DGQ, Kreis R, Bolliger CS, Boesch C, Slotboom J. Two-dimensional linear-combination model fitting of magnetic resonance spectra to define the macromolecule baseline using FiTAID, a Fitting Tool for Arrays of Interrelated Datasets. *MAGMA.* (2011) 24:147–64. doi: 10.1007/s10334-011-0246-y

61. Fuchs A, Boesiger P, Schulte RF, Henning A. ProFit revisited. *Magn Reson Med.* (2014) 71:458–68. doi: 10.1002/mrm.24703

62. Zöllner HJ, Považan M, Hui SCN, Tapper S, Edden RAE, Oeltzschner G. Comparison of different linear-combination modeling algorithms for short-TE proton spectra. *NMR Biomed.* (2021) 34:e4482. doi: 10.1002/nbm.4482

63. Craven AR, Bhattacharyya PK, Clarke WT, Dydak U, Edden RAE, Ersland L, et al. Comparison of seven modelling algorithms for γ-aminobutyric acid-edited proton magnetic resonance spectroscopy. *NMR Biomed.* (2022) 35:e4702. doi: 10.1002/nbm.4702

64. Marjańska M, Deelchand DK, Kreis R. 2016 ISMRM MRS Study Group Fitting Challenge Team. Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM. *Magn Reson Med.* (2022) 87:11–32. doi: 10.1002/mrm.28942

65. Lin A, Andronesi O, Bogner W, Choi I-Y, Coello E, Cudalbu C, et al. Minimum Reporting Standards for in vivo Magnetic Resonance Spectroscopy (MRSinMRS): Experts' consensus recommendations. *NMR Biomed.* (2021) 34:e4484. doi: 10.1002/nbm.4484

66. Smith SA, Levante TO, Meier BH, Ernst RR. Computer Simulations in Magnetic Resonance. An Object-Oriented Programming Approach. *J Magn Reson A.* (1994) 106:75–105. doi: 10.1006/jmra.1994.1008

67. Simpson R, Devenyi GA, Jezzard P, Hennessy TJ, Near J. Advanced processing and simulation of MRS data using the FID appliance (FID-A)-An open source, MATLAB-based toolkit. *Magn Reson Med.* (2017) 77:23–33. doi: 10.1002/mrm.26091

68. Zhang Y, An L, Shen J. Fast computation of full density matrix of multispin systems for spatially localized in vivo magnetic resonance spectroscopy. *Med Phys.* (2017) 44:4169–78. doi: 10.1002/mp.12375

69. Landheer K, Swanberg KM, Juchem C. Magnetic resonance Spectrum simulator (MARSS), a novel software package for fast and computationally efficient basis set simulation. *NMR Biomed.* (2021) 34:e4129. doi: 10.1002/nbm.4129

70. Cudalbu C, Mlynarik V, Xin L, Gruetter R. Comparison of two approaches to model the macromolecule spectrum for the quantification of short TE ^{1}H MRS spectra. In: *2008 IEEE International Workshop on Imaging Systems and Techniques.* (2008). p. 311–314. doi: 10.1109/IST.2008.4659991

71. Giapitzakis I-A, Avdievich N, Henning A. Characterization of macromolecular baseline of human brain using metabolite cycled semi-LASER at 94T. *Magn Reson Med.* (2018) 80:462–73. doi: 10.1002/mrm.27070

72. Mlynárik V, Gruber S, Moser E. Proton T (1) and T (2) relaxation times of human brain metabolites at 3 Tesla. *NMR Biomed.* (2001) 14:325–31. doi: 10.1002/nbm.713

73. Ethofer T, Mader I, Seeger U, Helms G, Erb M, Grodd W, et al. Comparison of longitudinal metabolite relaxation times in different regions of the human brain at 15 and 3 Tesla. *Magn Reson Med.* (2003) 50:1296–301. doi: 10.1002/mrm.10640

74. Träber F, Block W, Lamerichs R, Gieseke J, Schild HH. 1H metabolite relaxation times at 30 tesla: Measurements of T1 and T2 values in normal brain and determination of regional differences in transverse relaxation. *J Magn Reson Imaging.* (2004) 19:537–45. doi: 10.1002/jmri.20053

75. Wyss PO, Bianchini C, Scheidegger M, Giapitzakis IA, Hock A, Fuchs A, et al. In vivo estimation of transverse relaxation time constant (T2) of 17 human brain metabolites at 3T. *Magn Reson Med.* (2018) 80:452–61. doi: 10.1002/mrm.27067

76. Murali-Manohar S, Borbath T, Wright AM, Soher B, Mekle R, Henning A. T2 relaxation times of macromolecules and metabolites in the human brain at 94 T. *Magn Reson Med.* (2020) 84:542–58. doi: 10.1002/mrm.28174

77. Kirov II, Fleysher L, Fleysher R, Patil V, Liu S, Gonen O. Age dependence of regional proton metabolites T2 relaxation times in the human brain at 3 T. *Magn Reson Med.* (2008) 60:790–5. doi: 10.1002/mrm.21715

78. Gasparovic C, Neeb H, Feis DL, Damaraju E, Chen H, Doty MJ, et al. Quantitative spectroscopic imaging with in situ measurements of tissue water T1, T2, and density. *Magn Reson Med.* (2009) 62:583–90. doi: 10.1002/mrm.22060

79. Ongür D, Prescot AP, Jensen JE, Rouse ED, Cohen BM, Renshaw PF, et al. T2 relaxation time abnormalities in bipolar disorder and schizophrenia. *Magn Reson Med.* (2010) 63:1–8. doi: 10.1002/mrm.22148

80. Marjańska M, Emir UE, Deelchand DK, Terpstra M. Faster metabolite (1)H transverse relaxation in the elder human brain. *PLoS ONE.* (2013) 8:e77572. doi: 10.1371/journal.pone.0077572

81. Jiru F, Skoch A, Wagnerova D, Dezortova M, Viskova J, Profant O, et al. The age dependence of T2 relaxation times of N-acetyl aspartate, creatine and choline in the human brain at 3 and 4T. *NMR Biomed.* (2016) 29:284–92. doi: 10.1002/nbm.3456

82. Porges EC, Woods AJ, Lamb DG, Williamson JB, Cohen RA, Edden RAE, et al. Impact of tissue correction strategy on GABA-edited MRS findings. *Neuroimage.* (2017) 162:249–56. doi: 10.1016/j.neuroimage.2017.08.073

83. DeMayo MM, McGirr A, Selby B, MacMaster F, Debert CT, Harris AD. Investigating metabolite regional dependencies in the frontal lobe: Overlapping small and large voxels. *OSF Preprints.* (2022). doi: 10.31219/osf.io/3a42z

84. Park YW, Deelchand DK, Joers JM, Hanna B, Berrington A, Gillen JS, et al. AutoVOI: real-time automatic prescription of volume-of-interest for single voxel spectroscopy. *Magn Reson Med.* (2018) 80:1787–98. doi: 10.1002/mrm.27203

85. Bai X, Harris AD, Gong T, Puts NAJ, Wang G, Schär M, et al. Voxel Placement Precision for GABA-Edited Magnetic Resonance Spectroscopy. *Open J Radiol.* (2017) 7:35–44. doi: 10.4236/ojrad.2017.71004

86. Kimura H, Fujii Y, Itoh S, Matsuda T, Iwasaki T, Maeda M, et al. Metabolic alterations in the neonate and infant brain during

development: evaluation with proton MR spectroscopy. *Radiology.* (1995) 194:483–9. doi: 10.1148/radiology.194.2.7529934

87. Chang L, Ernst T, Poland RE, Jenden DJ. In vivo proton magnetic resonance spectroscopy of the normal aging human brain. *Life Sci.* (1996) 58:2049–56. doi: 10.1016/0024-3205(96)00197-X

88. Baker EH, Basso G, Barker PB, Smith MA, Bonekamp D, Horská A. Regional apparent metabolite concentrations in young adult brain measured by (1)H MR spectroscopy at 3 Tesla. *J Magn Reson Imaging.* (2008) 27:489–99. doi: 10.1002/jmri.21285

89. Shimizu M, Suzuki Y, Yamada K, Ueki S, Watanabe M, Igarashi H, et al. Maturational decrease of glutamate in the human cerebral cortex from childhood to young adulthood: a 1H-MR spectroscopy study. *Pediatr Res.* (2017) 82:749–52. doi: 10.1038/pr.2017.101

90. Porges EC, Jensen G, Foster B, Edden RA, Puts NA. The trajectory of cortical GABA across the lifespan, an individual participant data meta-analysis of edited MRS studies. *Elife.* (2021) 10:e62575. doi: 10.7554/eLife.62575

91. Kirov II, Sollberger M, Davitz MS, Glodzik L, Soher BJ, Babb JS, et al. Global brain volume and N-acetyl-aspartate decline over seven decades of normal aging. *Neurobiol Aging.* (2021) 98:42–51. doi: 10.1016/j.neurobiolaging.2020.10.024

92. Kreis R, Ernst T, Ross BD. Development of the human brain: in vivo quantification of metabolite and water content with proton magnetic resonance spectroscopy. *Magn Reson Med.* (1993) 30:424–37. doi: 10.1002/mrm.1910300405

93. Pouwels PJ, Brockmann K, Kruse B, Wilken B, Wick M, Hanefeld F, et al. Regional age dependence of human brain metabolites from infancy to adulthood as detected by quantitative localized proton MRS. *Pediatr Res.* (1999) 46:474–85. doi: 10.1203/00006450-199910000-00019

94. Dezortova M, Hajek M. 1H MR spectroscopy in pediatrics. *Eur J Radiol.* (2008) 67:240–9. doi: 10.1016/j.ejrad.2008.02.035

95. Panigrahy A, Nelson MD, Blüml S. Magnetic resonance spectroscopy in pediatric neuroradiology: clinical and research applications. *Pediatr Radiol.* (2010) 40:3–30. doi: 10.1007/s00247-009-1450-z

96. Blüml S, Wisnowski JL, Nelson MD Jr, Paquette L, Gilles FH, Kinney HC, et al. Metabolic maturation of the human brain from birth through adolescence: insights from in vivo magnetic resonance spectroscopy. *Cereb Cortex.* (2013) 23:2944–55. doi: 10.1093/cercor/bhs283

97. Bültmann E, Nägele T, Lanfermann H, Klose U. Changes of brain metabolite concentrations during maturation in different brain regions measured by chemical shift imaging. *Neuroradiology.* (2017) 59:31–41. doi: 10.1007/s00234-016-1763-1

98. Holmes MJ, Robertson FC, Little F, Randall SR, Cotton MF, van der Kouwe AJW, et al. Longitudinal increases of brain metabolite levels in 5-10 year old children. *PLoS ONE.* (2017) 12:e0180973. doi: 10.1371/journal.pone.0180973

99. Lebel C, Gee M, Camicioli R, Wieler M, Martin W, Beaulieu C. Diffusion tensor imaging of white matter tract evolution over the lifespan. *Neuroimage.* (2012) 60:340–52. doi: 10.1016/j.neuroimage.2011.11.094

100. Schmitz B, Wang X, Barker PB, Pilatus U, Bronzlik P, Dadak M, et al. Effects of Aging on the Human Brain: A Proton and Phosphorus MR Spectroscopy Study at 3T. *J Neuroimaging.* (2018) 28:416–21. doi: 10.1111/jon.12514

101. Merritt K, McGuire PK, Egerton A, Aleman A, Block W, Bloemen OJ, et al. Association of age, antipsychotic medication, and symptom severity in schizophrenia with proton magnetic resonance spectroscopy brain glutamate level: a mega-analysis of individual participant-level data. *JAMA Psychiatry.* (2021) 78:667–81. doi: 10.1001/jamapsychiatry.2021.0380

102. Raininko R, Mattsson P. Metabolite concentrations in supraventricular white matter from teenage to early old age: A short echo time 1H magnetic resonance spectroscopy (MRS) study. *Acta Radiol.* (2010) 51:309–15. doi: 10.3109/02841850903476564

103. Luders E, Kurth F. Structural differences between male and female brains. *Handb Clin Neurol.* (2020) 175:3–11. doi: 10.1016/B978-0-444-64123-6.00001-1

104. Peper JS, Burke SM, Wierenga LM. Sex differences and brain development during puberty and adolescence. *Handb Clin Neurol.* (2020) 175:25–54. doi: 10.1016/B978-0-444-64123-6.00003-5

105. Williams CM, Peyre H, Toro R, Ramus F. Sex differences in the brain are not reduced to differences in body size. *Neurosci Biobehav Rev.* (2021) 130:509–11. doi: 10.1016/j.neubiorev.2021.09.015

106. Hodgetts S, Hausmann M. Sex/gender differences in brain lateralisation and connectivity. In: *Current Topics in Behavioral Neurosciences.* Berlin, Heidelberg: Springer. (2022) doi: 10.1007/7854_2022_303

107. Salminen LE, Tubi MA, Bright J, Thomopoulos SI, Wieand A, Thompson PM. Sex is a defining feature of neuroimaging phenotypes in major brain disorders. *Hum Brain Mapp.* (2022) 43:500–42. doi: 10.1002/hbm.25438

108. Wilkinson ID, Paley MN, Miszkiel KA, Hall-Craggs MA, Kendall BE, Chinn RJ, et al. Cerebral volumes and spectroscopic proton metabolites on MR: is sex important? *Magn Reson Imaging.* (1997) 15:243–8. doi: 10.1016/S0730-725X(96)00334-7

109. Grachev ID, Apkarian AV. Chemical heterogeneity of the living human brain: a proton MR spectroscopy study on the effects of sex, age, and brain region. *Neuroimage.* (2000) 11:554–63. doi: 10.1006/nimg.2000.0557

110. Braun CMJ, Boulanger Y, Labelle M, Khiat A, Dumont M, Mailloux C. Brain metabolic differences as a function of hemisphere, writing hand preference, and gender. *Laterality.* (2002) 7:97–113. doi: 10.1080/13576500143000212

111. Tayoshi S'ya, Sumitani S, Taniguchi K, Shibuya-Tayoshi S, Numata S, Iga J-I, et al. Metabolite changes and gender differences in schizophrenia using 3-Tesla proton magnetic resonance spectroscopy (1H-MRS). *Schizophr Res.* (2009) 108:69–77. doi: 10.1016/j.schres.2008.11.014

112. Maudsley AA, Domenig C, Govind V, Darkazanli A, Studholme C, Arheart K, et al. Mapping of brain metabolite distributions by volumetric proton MR spectroscopic imaging (MRSI). *Magn Reson Med.* (2009) 61:548–59. doi: 10.1002/mrm.21875

113. García Santos JM, Fuentes LJ, Vidal JB, Antequera M, Torres Del Río S, Antúnez C, et al. Regional effects of age and sex in magnetic resonance spectroscopy. *Radiologia.* (2010) 52:342–50. doi: 10.1016/S2173-5107(10)70023-9

114. Charles HC, Lazeyras F, Krishnan KR, Boyko OB, Patterson LJ, Doraiswamy PM, et al. Proton spectroscopy of human brain: effects of age and sex. *Prog Neuropsychopharmacol Biol Psychiatry.* (1994) 18:995–1004. doi: 10.1016/0278-5846(94)90125-2

115. Komoroski RA, Heimberg C, Cardwell D, Karson CN. Effects of gender and region on proton MRS of normal human brain. *Magn Reson Imaging.* (1999) 17:427–33. doi: 10.1016/S0730-725X(98)00186-6

116. Nagae-Poetscher LM, Bonekamp D, Barker PB, Brant LJ, Kaufmann WE, Horská A. Asymmetry and gender effect in functionally lateralized cortical regions: a proton MRS imaging study. *J Magn Reson Imaging.* (2004) 19:27–33. doi: 10.1002/jmri.10429

117. Jung RE, Haier RJ, Yeo RA, Rowland LM, Petropoulos H, Levine AS, et al. Sex differences in N-acetylaspartate correlates of general intelligence: an 1H-MRS study of normal human brain. *Neuroimage.* (2005) 26:965–72. doi: 10.1016/j.neuroimage.2005.02.039

118. Chiu P-W, Mak HK-F, Yau KK-W, Chan Q, Chang RC-C, Chu L-W. Metabolic changes in the anterior and posterior cingulate cortices of the normal aging brain: proton magnetic resonance spectroscopy study at 3 T. *Age.* (2014) 36:251–64. doi: 10.1007/s11357-013-9545-8

119. O'Gorman RL, Michels L, Edden RA, Murdoch JB, Martin E. In vivo detection of GABA and glutamate with MEGA-PRESS: reproducibility and gender effects. *J Magn Reson Imaging.* (2011) 33:1262–7. doi: 10.1002/jmri.22520

120. Goryawala MZ, Sheriff S, Maudsley AA. Regional distributions of brain glutamate and glutamine in normal subjects. *NMR Biomed.* (2016) 29:1108–16. doi: 10.1002/nbm.3575

121. Cichocka M, Kozub J, Karcz P, Urbanik A. Sex differences in brain metabolite concentrations in healthy children - proton magnetic resonance spectroscopy study (1HMRS). *Pol J Radiol.* (2018) 83:e24–31. doi: 10.5114/pjr.2018.74536

122. Kulak W, Sobaniec W, Smigielska-Kuzia J, Kubas B, Walecki J. Metabolite profile in the basal ganglia of children with cerebral palsy: a proton magnetic resonance spectroscopy study. *Dev Med Child Neurol.* (2006) 48:285–9. doi: 10.1017/S0012162206000612

123. Kadota T, Horinouchi T, Kuroda C. Development and aging of the cerebrum: assessment with proton MR spectroscopy. *AJNR Am J Neuroradiol.* (2001) 22:128–35.

124. Sijens PE, Oudkerk M, de Leeuw FE, de Groot JC, Achten E, Heijboer R, et al. 1H chemical shift imaging of the human brain at age 60-90 years reveals metabolic differences between women and men. *Magn Reson Med.* (1999) 42:24–31. doi: 10.1002/(SICI)1522-2594(199907)42:1<24::AID-MRM5>3.0.CO;2-3

125. Sijens PE, Den Heijer T, De Leeuw FE, De Groot JC, Achten E, Heijboer RJ, et al. Human brain chemical shift imaging at age 60 to 90: analysis of the causes of the observed sex differences in brain metabolites. *Invest Radiol.* (2001) 36:597–603. doi: 10.1097/00004424-200110000-00005

126. Sijens PE, den Heijer T, Origgi D, Vermeer SE, Breteler MMB, Hofman A, et al. Brain changes with aging: MR spectroscopy at supraventricular plane shows differences between women and men. *Radiology.* (2003) 226:889–96. doi: 10.1148/radiol.2263011937

127. Gao F, Edden RAE Li M, Puts NAJ, Wang G, Liu C, Zhao B, et al. Edited magnetic resonance spectroscopy detects an age-related decline in brain GABA levels. *Neuroimage.* (2013) 78:75–82. doi: 10.1016/j.neuroimage.2013.04.012

128. Sailasuta N, Ernst T, Chang L. Regional variations and the effects of age and gender on glutamate concentrations in the human brain. *Magn Reson Imaging.* (2008) 26:667–75. doi: 10.1016/j.mri.2007.06.007

129. Wickens MM, Bangasser DA, Briand LA. Sex differences in psychiatric disease: a focus on the glutamate system. *Front Mol Neurosci.* (2018) 11:197. doi: 10.3389/fnmol.2018.00197

130. Pigoni A, Delvecchio G, Squarcina L, Bonivento C, Girardi P, Finos L, et al. Sex differences in brain metabolites in anxiety and mood disorders. *Psychiat Res Neuroimag.* (2020) 305:111196. doi: 10.1016/j.pscychresns.2020.111196

131. Ware AL, Yeates KO, Tang K, Shukla A, Onicas AI, Guo S, et al. Longitudinal white matter microstructural changes in pediatric mild traumatic brain injury: An A-CAP study. *Hum Brain Mapp.* (2022) 43:3809–23. doi: 10.1002/hbm.25885

132. Caramanos Z, Narayanan S, Arnold DL. 1H-MRS quantification of tNA and tCr in patients with multiple sclerosis: a meta-analytic review. *Brain.* (2005) 128:2483–506. doi: 10.1093/brain/awh640

133. Song T, Song X, Zhu C, Patrick R, Skurla M, Santangelo I, et al. Mitochondrial dysfunction, oxidative stress, neuroinflammation, and metabolic alterations in the progression of Alzheimer's disease: A meta-analysis of in vivo magnetic resonance spectroscopy studies. *Ageing Res Rev.* (2021) 72:101503. doi: 10.1016/j.arr.2021.101503

134. Joyce JM, La PL, Walker R, Harris AD. Magnetic resonance spectroscopy of traumatic brain injury and subconcussive hits: a systematic review and meta-analysis. *J Neurotrauma.* (2022) 39:1455–1476. doi: 10.1089/neu.2022.0125

135. Suh CH, Kim HS, Jung SC, Choi CG, Kim SJ. 2-Hydroxyglutarate MR spectroscopy for prediction of isocitrate dehydrogenase mutant glioma: a systemic review and meta-analysis using individual patient data. *Neuro Oncol.* (2018) 20:1573–83. doi: 10.1093/neuonc/noy113

136. Suh CH, Kim HS, Jung SC, Park JE, Choi CG, Kim SJ, et al. as a diagnostic biomarker for differentiating primary central nervous system lymphoma from glioblastoma: A systematic review and meta-analysis. *J Magn Reson Imaging.* (2019) 50:560–72. doi: 10.1002/jmri.26602

137. Wang L, Chen G, Dai K. Hydrogen proton magnetic resonance spectroscopy (MRS) in differential diagnosis of intracranial tumors: a systematic review. *Contrast Media Mol Imaging.* (2022) 2022:7242192. doi: 10.1155/2022/7242192

138. Kantrowitz JT, Dong Z, Milak MS, Rashid R, Kegeles LS, Javitt DC, et al. Ventromedial prefrontal cortex/anterior cingulate cortex Glx, glutamate, and GABA levels in medication-free major depressive disorder. *Transl Psychiatry.* (2021) 11:419. doi: 10.1038/s41398-021-01541-1

139. Truong V, Cheng PZ, Lee H-C, Lane TJ, Hsu T-Y, Duncan NW. Occipital gamma-aminobutyric acid and glutamate-glutamine alterations in major depressive disorder: An MRS study and meta-analysis. *Psychiatry Res Neuroimag.* (2021) 308:111238. doi: 10.1016/j.pscychresns.2020.111238

140. Smucny J, Carter CS, Maddock RJ. Medial Prefrontal Cortex Glutamate Is Reduced in Schizophrenia and Moderated by Measurement Quality: A Meta-analysis of Proton Magnetic Resonance Spectroscopy Studies. *Biol Psychiatry.* (2021) 90:643–51. doi: 10.1016/j.biopsych.2021.06.008

141. Chen T, Tan H, Lei H, Li X, Wu Q, Xu X, et al. Nature of glutamate alterations in substance dependence: A systematic review and meta-analysis of proton magnetic resonance spectroscopy studies. *Psychiatry Res Neuroimaging.* (2021) 315:111329. doi: 10.1016/j.pscychresns.2021.111329

142. Peek AL, Rebbeck T, Puts NA, Watson J, Aguila M-ER, Leaver AM. Brain GABA and glutamate levels across pain conditions: A systematic literature review and meta-analysis of 1H-MRS studies using the MRS-Q quality assessment tool. *Neuroimage.* (2020) 210:116532. doi: 10.1016/j.neuroimage.2020.116532

143. Chelala L, O'Connor EE, Barker PB, Zeffiro TA. Meta-analysis of brain metabolite differences in HIV infection. *Neuroimage Clin.* (2020) 28:102436. doi: 10.1016/j.nicl.2020.102436

144. Dahmani S, Kaliss N, VanMeter JW, Moore DJ, Ellis RJ, Jiang X. Alterations of Brain Metabolites in Adults With HIV: A Systematic Meta-analysis of Magnetic Resonance Spectroscopy Studies. *Neurology.* (2021) 97:e1085–96. doi: 10.1212/WNL.0000000000012394

145. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* (2015) 4:1. doi: 10.1186/2046-4053-4-1

146. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ.* (2010) 340:c221. doi: 10.1136/bmj.c221

147. Belias M, Rovers MM, Reitsma JB, Debray TPA, IntHout J. Statistical approaches to identify subgroups in meta-analysis of individual participant data: a simulation study. *BMC Med Res Methodol.* (2019) 19:183. doi: 10.1186/s12874-019-0817-6

148. Tal A, Kirov II, Grossman RI, Gonen O. The role of gray and white matter segmentation in quantitative proton MR spectroscopic imaging. *NMR Biomed.* (2012) 25:1392–400. doi: 10.1002/nbm.2812

149. Johnson WE Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* (2006) 8:118–27. doi: 10.1093/biostatistics/kxj037

150. Fortin J-P, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage.* (2017) 161:149–70. doi: 10.1016/j.neuroimage.2017.08.047

151. Fortin J-P, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage.* (2018) 167:104–20. doi: 10.1016/j.neuroimage.2017.11.024

152. Radua J, Vieta E, Shinohara R, Kochunov P, Quidé Y, Green MJ, et al. Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. *Neuroimage.* (2020) 218:116956. doi: 10.1016/j.neuroimage.2020.116956

153. Sun D, Rakesh G, Haswell CC, Logue M, Lexi Baird C, O'Leary BM, et al. A comparison of methods to harmonize cortical thickness measurements across scanners and sites. *Neuroimage.* (2021) 261:119509. doi: 10.1101/2021.09.22.461242

154. Yu M, Linn KA, Cook PA, Phillips ML, McInnis M, Fava M, et al. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum Brain Mapp.* (2018) 39:4213–27. doi: 10.1002/hbm.24241

155. Bell TK, Godfrey KJ, Ware AL, Yeates KO, Harris AD. Harmonization of multi-site MRS data with ComBat. *Neuroimage.* (2022) 257:119330. doi: 10.1016/j.neuroimage.2022.119330

156. Bento M, Fantini I, Park J, Rittner L, Frayne R. Deep Learning in Large and Multi-Site Structural Brain MR Imaging Datasets. *Front Neuroinform.* (2021) 15:805669. doi: 10.3389/fninf.2021.805669

157. Guan H, Liu M. Domain adaptation for medical image analysis: a survey. *IEEE Trans Biomed Eng.* (2022) 69:1173–85. doi: 10.1109/TBME.2021.3117407

158. Dinsdale NK, Jenkinson M, Namburete AIL. Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *Neuroimage.* (2021) 228:117689. doi: 10.1016/j.neuroimage.2020.117689

159. Moyer D, Ver Steeg G, Tax CMW, Thompson PM. Scanner invariant representations for diffusion MRI harmonization. *Magn Reson Med.* (2020) 84:2174–89. doi: 10.1002/mrm.28243

160. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. *Med Image Anal.* (2019) 58:101552. doi: 10.1016/j.media.2019.101552

161. Armanious K, Jiang C, Fischer M, Küstner T, Hepp T, Nikolaou K, et al. MedGAN: Medical image translation using GANs. *Comput Med Imaging Graph.* (2020) 79:101684. doi: 10.1016/j.compmedimag.2019.101684

162. Saat P, Nogovitsyn N, Hassan MY, Ganaie MA, Souza R, Hemmati H. A domain adaptation benchmark for T1-weighted brain magnetic resonance image segmentation. *Front Neuroinform.* (2022) 16:919779. doi: 10.3389/fninf.2022.919779

163. Bian C, Yuan C, Ma K, Yu S, Wei D, Zheng Y. Domain adaptation meets zero-shot learning: an annotation-efficient approach to multi-modality medical image segmentation. *IEEE Trans Med Imaging.* (2022) 41:1043–56. doi: 10.1109/TMI.2021.3131245

164. Deelchand DK, Adanyeguh IM, Emir UE, Nguyen T-M, Valabregue R, Henry P-G, et al. Two-site reproducibility of cerebellar and brainstem neurochemical profiles with short-echo, single-voxel MRS at 3T. *Magn Reson Med.* (2015) 73:1718–25. doi: 10.1002/mrm.25295

165. Mikkelsen M, Barker PB, Bhattacharyya PK, Brix MK, Buur PF, Cecil KM, et al. Big GABA: Edited MR spectroscopy at 24 research sites. *Neuroimage.* (2017) 159:32–45. doi: 10.1016/j.neuroimage.2017.07.021

166. Sturrock A, Laule C, Wyper K, Milner RA, Decolongon J, Dar Santos R, et al. A longitudinal study of magnetic resonance spectroscopy Huntington's disease biomarkers. *Mov Disord.* (2015) 30:393–401. doi: 10.1002/mds.26118

167. van de Bank BL, Emir UE, Boer VO, van Asten JJA, Maas MC, Wijnen JP, et al. Multi-center reproducibility of neurochemical profiles in the human brain at 7 T. *NMR Biomed.* (2015) 28:306–16. doi: 10.1002/nbm.3252

168. Yeates KO, Beauchamp M, Craig W, Doan Q, Zemek R, Bjornson B, et al. Advancing Concussion Assessment in Pediatrics (A-CAP): a prospective, concurrent cohort, longitudinal study of mild traumatic brain injury in children: protocol study. *BMJ Open.* (2017) 7:e017012. doi: 10.1136/bmjopen-2017-017012

169. Deelchand DK, Berrington A, Noeske R, Joers JM, Arani A, Gillen J, et al. Across-vendor standardization of semi-LASER for single-voxel MRS at 3T. *NMR Biomed.* (2019) 34:e4218. doi: 10.1002/nbm.4218

170. Graf C, MacMillan EL, Fu E, Harris T, Traboulsee A, Vavasour IM, et al. Intra- and inter-site reproducibility of human brain single-voxel proton MRS at 3 T. *NMR Biomed.* (2019) 32:e4083. doi: 10.1002/nbm.4083

171. Alosco ML, Mariani ML, Adler CH, Balcer LJ, Bernick C, Au R, et al. Developing methods to detect and diagnose chronic traumatic encephalopathy during life: rationale, design, and methodology for the DIAGNOSE CTE Research Project. *Alzheimers Res Ther.* (2021) 13:136. doi: 10.1186/s13195-021-00872-x

172. Prescot A, Huber R, Kanekar S, Kondo D, Prisciandaro J, Ongur D, et al. Effect of moderate altitude on human cerebral metabolite levels: A preliminary, multi-site, proton magnetic resonance spectroscopy investigation. *Psychiat Res Neuroimag.* (2021) 314:111314. doi: 10.1016/j.pscychresns.2021.111314

173. Wisnowski JL, Bluml S, Panigrahy A, Mathur AM, Berman J, Chen P-SK, et al. Integrating neuroimaging biomarkers into the multicentre, high-dose erythropoietin for asphyxia and encephalopathy (HEAL) trial: rationale, protocol and harmonisation. *BMJ Open.* (2021) 11:e043852. doi: 10.1136/bmjopen-2020-043852