Continuous automated MRS data analysis workflow for applied studies

Helge J. Zöllner[1,2*], Christopher W. Davies-Jenkins[1,2], Erik G. Lee[3,4], Timothy J. Hendrickson[3,4], Richard A. E. Edden[1,2], Aaron T. Gudmundson[1,2], Georg Oeltzschner[1,2]

[1] *Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, MD, United States*

[2] *F. M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD, United States*

[3] *Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, MN, United States*

[4] *Informatics Institute, University of Minnesota, Minneapolis, MN, United States.*

***Corresponding author:**

Helge Jörn Zöllner, Ph.D.

Russell H. Morgan Department of Radiology and Radiological Science

The Johns Hopkins University School of Medicine

600 N Wolfe St, Baltimore, MD 21287

hzoelln2@jhmi.edu

Running title: Continuous automated MRS data analysis

Word count: 2,651

Keywords: magnetic resonance spectroscopy; BIDS; NIfTI-MRS; reproducibility; Osprey; linear-combination modeling

Abstract

Purpose
Magnetic resonance spectroscopy (MRS) can non-invasively measure levels of endogenous metabolites in living tissue and is of great interest to neuroscience and clinical research. To this day, MRS data analysis workflows differ substantially between groups, frequently requiring many manual steps to be performed on individual datasets, e.g., data renaming/sorting, manual execution of analysis scripts, and manual assessment of success/failure. Manual analysis practices are a substantial barrier to wider uptake of MRS. They also increase the likelihood of human error and prevent deployment of MRS at large scale. Here, we demonstrate an end-to-end workflow for fully automated data uptake, processing, and quality review.

Methods
The proposed continuous automated MRS analysis workflow integrates several recent innovations in MRS data and file storage conventions. They are efficiently deployed by a directory monitoring service that automatically triggers the following steps upon arrival of a new raw MRS dataset in a project folder: (1) conversion from proprietary manufacturer file formats into the universal format NIfTI-MRS; (2) consistent file system organization according to the data accumulation logic standard BIDS-MRS; (3) executing a command-line executable of our open-source end-to-end analysis software Osprey; (4) e-mail delivery of a quality control summary dashboard for all analysis steps.

Results
The automated architecture successfully completed for a demonstration dataset. The only manual step required is to copy a raw data folder into a monitored directory.

Conclusion
Continuous automated analysis of MRS data can reduce the burden of manual data analysis and quality control, particularly for non-expert users and large-scale studies.

Introduction

Magnetic resonance spectroscopy (MRS) provides access to measuring levels of endogenous biochemicals in vivo with a clinical MRI scanner. Approximately 15 different metabolites can be resolved, e.g., markers of neuronal integrity and cell proliferation, neurotransmitters, and antioxidants. MRS is therefore a method of great interest to clinical research and neuroscience. However, although MRS predates other modern MRI modalities (fMRI, DTI) by a decade, it is not used as widely (1). MRS analysis workflows are commonly not automated and require local spectroscopy expertise to be executed. Everyday use of MRS has therefore remained restricted to a few specialized research centers.

One barrier to efficient inclusion of MRS into multimodal imaging protocols is the number of manual steps required. Applied research studies with MRS in the acquisition protocol require manual data renaming and sorting, manual amendment of analysis scripts, manual execution of the analysis pipeline, and manual assessment of processing success on individual datasets. Depending on the amount of data and complexity of the protocol, these tasks can take a substantial amount of time. Manual analysis labor is therefore a strongly rate-limiting factor that prevents the deployment of MRS at scale.

Because MRS data are less familiar than MRI (and because spectra are not brain images), it can be challenging for non-expert users to distinguish between 'successful' and 'unsuccessful' acquisitions at the scanner. It is therefore regrettably common for data to be lost due to procedural error (e.g., a technologist uses an incorrect protocol version, or accidentally places an MRS voxel in the wrong hemisphere) or as a consequence of hardware and software updates (e.g., a software update introduces a new default exam parameter that overrides an acquisition-critical one). Regular data inspection is therefore necessary to capture sudden unintended acquisition protocol changes – while our recommendation to users is "check after every scan", this places a large burden on research staff, with the result that it is not consistently followed. Finally, the requirement for manual data analysis increases the likelihood of human error, e.g., copying the wrong files, using incorrect filenames, or executing the wrong analysis script.

In this manuscript, we describe a fully automated data review and end-to-end analysis workflow to support non-expert users of MRS conducting local research studies. This procedure makes use of several recent innovations in MRS file formats and large-scale neuroimaging data management, which make it much easier to deploy repeated analyses. The workflow can be customized to accommodate local project protocols and IT infrastructure.

Methods

The primary goal of this work is to reduce the manual overhead involved in checking data regularly, and to make modern MRS analysis easier. The secondary goal is to demonstrate the

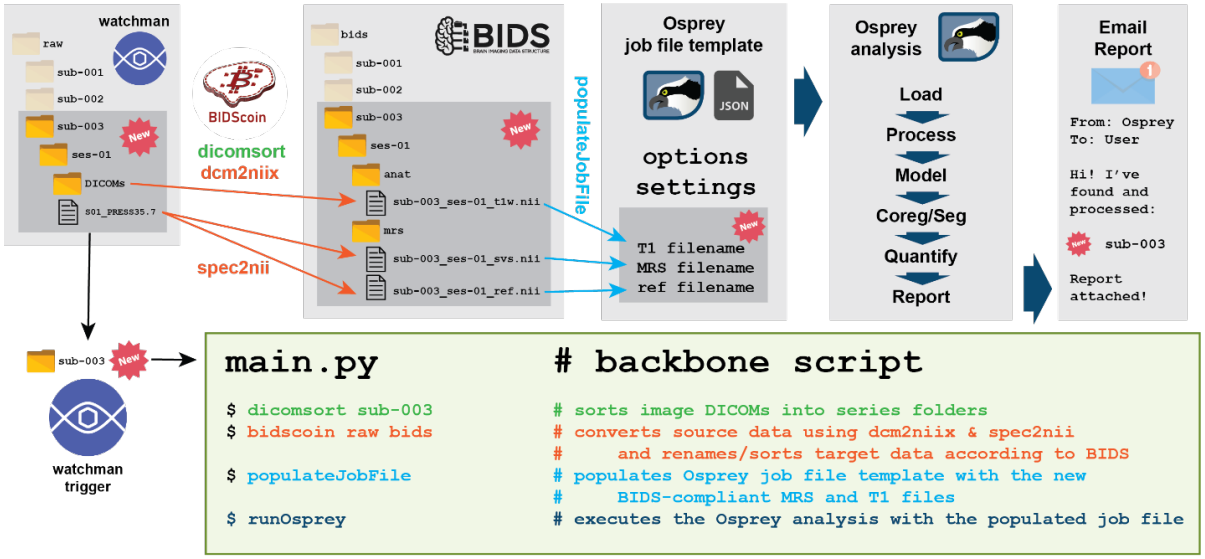

Figure 1: Illustration of the automated MRS data analysis workflow. The *watchman* service notices a new folder in the raw data directory that it monitors. It triggers the backbone script, which first uses *dicomsort* to sort image DICOM into series folders, then executes *bidscoin* which uses the *dcm2niix* and *spec2nii* programs to convert the raw data to NIfTI/NIfTI-MRS/BIDS format. The script then populates a template job file with paths to the new data and passes the job file on to an *Osprey* analysis instance. The last *Osprey* step sends a confirmation e-mail to specified addresses.

benefits of automated analysis workflows for the deployment of MRS methods at scale. Reducing manual analysis effort is necessary to open MRS methods up to be a part of multi-site consortium projects or continuous long-time local initiatives.

First, a command-line job scheduling or file-watching utility is deployed at regular intervals to monitor the content of the project folder. If the service detects previously unencountered raw MRS data, a script is launched to convert them into a standardized folder and file structure according to a pre-defined study-specific template. Next, a new analysis job is automatically set up and executed. After completion of the analysis, the tool sends the user an e-mail with a dashboard summarizing quantitative results, quality control criteria, and analysis diagnostics (warnings, errors). This effectively removes the need to check data manually after every scan – if no email is received, the user is alerted to check whether data export from the scanner occurred as intended. Equally, if the data exported are incomplete (i.e., do not agree with the expected data specified in the study template), the email report will signal the need to repeat data export.

The workflow (**Figure 1**) integrates several aspects and innovations:

- scripted raw data format conversion from proprietary manufacturer file formats into the universal storage format NIfTI-MRS
- consistent organization and metadata annotation according to the data accumulation logic standard BIDS

- a compiled command-line executable version of our open-source end-to-end analysis software Osprey
- regular monitoring of a raw data project folder for automation of the above steps when a new dataset arrives
- e-mail notification upon completion of a new automated analysis.

Example data

We implemented and tested this workflow on raw data from a single site with a 3T General Electric scanner from the Big GABA dataset (site ID 'G1'), specifically the image DICOMs of a 3D anatomical T1-weighted MP-RAGE scan, and a short-TE (TE = 35 ms) PRESS scan in .7 format (2).

Backbone script

The 'chain of commands' (raw data conversion, analysis execution, reporting) is strung together in a small Python script (*main.py*).

NIfTI-MRS & BIDS-MRS conversion

This standardized MRS file format convention was developed to overcome longstanding issues with different raw data formats (3). Each manufacturer uses several proprietary file types (Siemens DAT and RDA; Philips SDAT/SPAR, DATA/LIST and SIN/LAB/RAW; GE .7 P-file; multiple vendor flavors of DICOM). These contain different non-standardized data and header information and are mutually incompatible. NIfTI-MRS is based on the NIfTI-2 standard that is widely used in neuroimaging. Raw data is stored in well-defined array dimensions and accompanied by standardized header information fields saving all metadata necessary to reconstruct the spectra. The Brain Imaging Data Structure (BIDS) is a neuroimaging data organization standard (4). File and folder naming conventions create a standardized hierarchy (study, subject, session, modality) that benefits the automation of analysis pipelines and data sharing. BIDS further catalogues required metadata on technical details and demographics. BIDS-MRS is an MRS-specific extension of BIDS currently in development (5).

Several tools have been developed to automate the conversion and sorting of raw DICOM image source data into BIDS-compliant NIfTI, e.g., HeuDiConv (6), dcm2bids (7), bidskit (8), etc. They use heuristics based on DICOM metadata and file paths to recognize image types and determine 'maps', i.e., well-defined procedures for BIDS-compliant file naming and metadata annotation. While very flexible and powerful, their learning curves can be steep. We decided to use BIDScoin (9) for two reasons:

1) its user-friendly GUI allows researchers to define appropriate mapping strategies after the software has made intelligent guesses.
2) it includes a plugin to run spec2nii, a command-line conversion tool from vendor-proprietary formats into NIfTI-MRS.

The backbone script first calls *dicomsort*, a command-line utility included in BIDScoin that sorts unsorted image DICOMs into folders named for the image series they were acquired under, e.g., "T1-MPRAGE". Using BIDScoin's GUI (*bidsmapper raw bids*) and an example dataset (*sub-001*), we created a bidsmap file to automatically recognize the source MRS dataset and the source anatomical image DICOMs in these image series folders. Using *bidseditor raw*, we defined output naming heuristics, including identifying the DICOMs as a structural T1-weighted image, and adding BIDS-MRS labels *vox-pcc* to denote the voxel location and `acq-press` to highlight that these data were acquired with PRESS localization. These translation heuristics, from source to BIDS-compliant output, are stored in a YAML-formatted file (the *bidsmap*) that serves as a template for any subsequently acquired subject folders (*sub-002*, *sub-003*, etc.).

Osprey

Osprey is our free modular open-source software written in MATLAB that completes all consensus-recommended modern MRS data analysis steps, i.e., pre-processing of raw data, modeling pre-processed data, and converting model parameters into metabolite concentration measures (10). For the present work, we compiled Osprey v2.4.0 into standalone executables for Windows 10 and MacOS using the MATLAB Compiler. We further created an installation routine that automatically downloads and installs the necessary MATLAB Runtime libraries.

Osprey requires a JSON-formatted 'job file' specifying the analysis options, settings, and the complete paths to the NIfTI-MRS/BIDS-formatted data. This workflow automatically generates this file by populating a template job file with the newly encountered BIDS-formatted data. Once the job file has been written, the workflow calls the Osprey executable, with the job file as the only input argument.

Reporting

Once Osprey has completed the requested analysis, a standardized reporting page (in HTML format) is packaged into a zip file and attached to an e-mail sent to a set of addresses specified by the user in the job file. The report contains visualizations of the different analysis stages (raw data loading, pre-processing, modeling, co-registration, segmentation) and quality control (FWHM, SNR, fit error, frequency drift).

Source data directory monitoring

The backbone script *main.py* can easily be triggered manually after a new dataset has been added to the source data folder. We further increased the degree of analysis automation with an open-source file watching service (*watchman*) to continuously monitor the source data directory for changes. When *watchman* detects an added subject-level folder in the source data directory, it performs a quick check whether this new folder fulfills source data formatting requirements. If it does, watchman triggers *main.py*.

Results

The workflow presented here was successfully deployed to process the test dataset. After the initial configuration of the bidsmap, Osprey job file templates, and *watchman* scripts, no further user intervention was required. All seven single-subjects datasets were automatically pushed through the entire pipeline (DICOM sorting, NIfTI-MRS/BIDS conversion, Osprey analysis). Osprey output files included the summary report sent via e-mail. The routine file system output generated by Osprey (PDF visualizations, tab-separated values files holding metabolite estimates for further analysis, binary voxel masks, etc.) is automatically saved into a *derivatives* folder as specified by the parent BIDS.

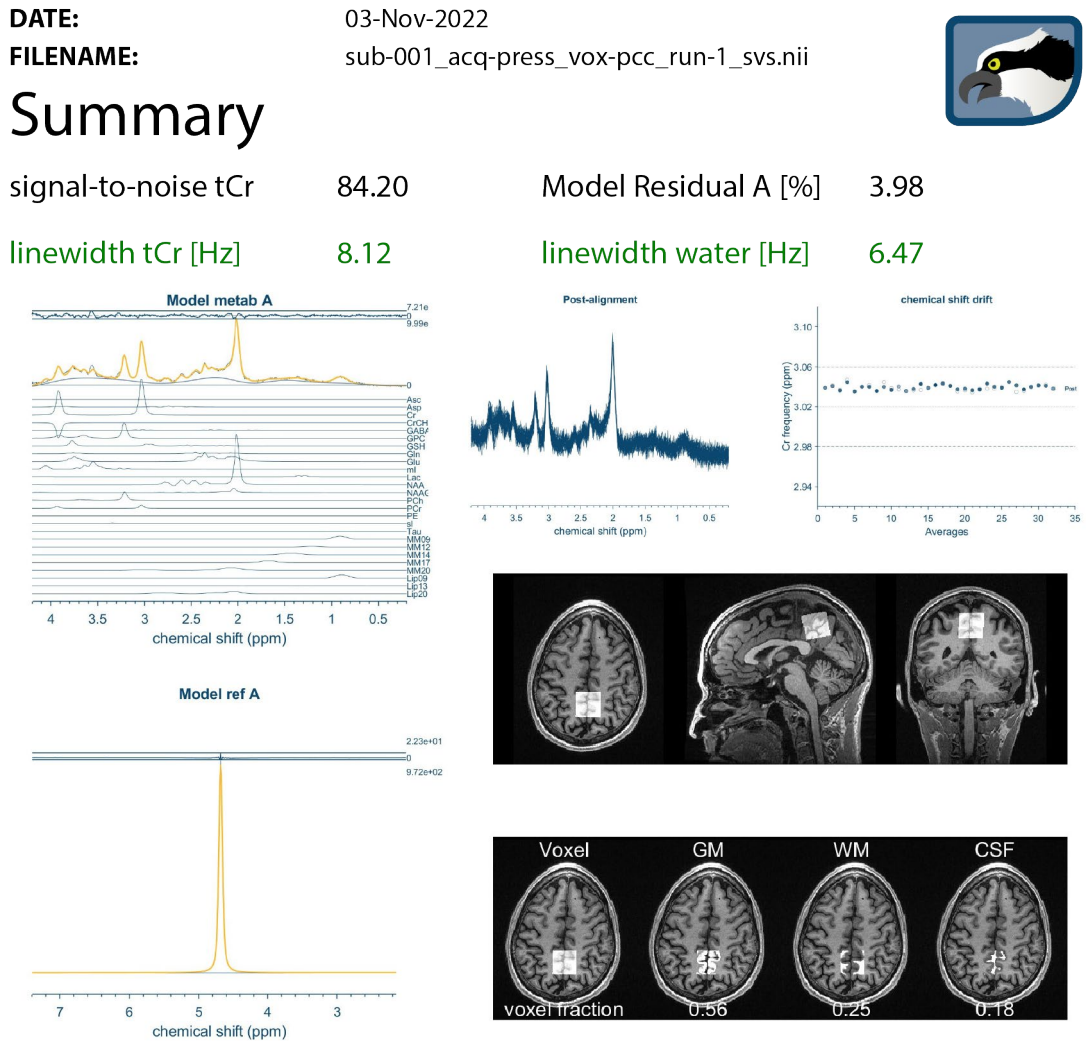

Figure 2: Summary dashboard reporting on all key steps of the automated MRS data analysis workflow (lightly edited from HTML output to minimize whitespace for presentation purposes). The HTML report is automatically e-mailed to a user-specified address after completion of the Osprey analysis. The summary includes quantitative quality control metrics (linewidth, SNR, model fit error) and visualizations of the processing, modeling, and co-registration/segmentation steps of the automated Osprey pipeline.

Discussion

We have demonstrated an example of an automated workflow for project folder organization, data analysis and reporting for applied MRS studies. Automation of MRS data analysis can improve reproducibility and efficiency, particularly at sites that historically have not established local workflows. The immediate feedback helps users identify sudden fatal changes to the hardware or software setup, for example after scanner upgrades. Lastly, automated data analysis can lower the entry-level threshold to engage with MRS in general, and provide much-needed scale.

Standardized workflows and data storage conventions also increase reusability of existing analysis code at various levels. Once set up for a particular study, the setup described herein can easily be ported to a new study protocol, simply by adjusting the bidsmaps and Osprey job file templates. This increases 'horizontal' efficiency for the individual researcher, between members of a research group, and between locations of multi-site consortia, but also 'longitudinal' efficiency for incoming personnel taking over responsibility for existing projects.

This automated analysis workflow may also serve as a template for the integration of MRS into large-scale multi-center multi-modal neuroimaging studies. MRS data acquisition and analysis have not been easy scalable in the past since they were limited by the amount of manual effort they demand. As a result, even the largest MRS datasets may only include a few hundred subjects – orders of magnitude less than the most powerful databases of structural, functional, or diffusion imaging. BIDS has transformed the way in which large-scale neuroimaging datasets are made accessible to other researchers, and the anticipated publication of the BIDS-MRS extension will open up many existing neuroimaging repositories for use with MRS data.

Alternative strategies

Neuroimaging centers may already have solutions to transform source imaging data into BIDS format in place, e.g., for fMRI, PET, or EEG data. The BIDScoin part of the workflow presented here may easily be swapped for existing `BIDS-ification` methods, although it offers the advantage of already offering a work-in-progress MRS integration via a *spec2nii* plugin.

We demonstrate the workflow with Osprey, our own software for MRS pre-processing, modeling, and quantification. Researchers can, of course, swap Osprey for any other MRS analysis software. Note, however, that full BIDS compliance requires NIfTI-MRS as the core file storage format. NIfTI-MRS is supported by recently published open-source software packages (FSL-MRS (11), spant (12)) and the latest versions of some seasoned ones (Vespa (13), jMRUI (14)). LCModel (15), on the contrary, is unlikely to directly interface with NIfTI-MRS unless data are prepared with modular processing packages like FID-A (16) which includes NIfTI-MRS reading and writing routines. Osprey also includes a fully-functioning wrapper to call LCModel. Researchers wishing to maintain vendor-native formats (SDAT/SPAR, RDA, etc.) may construct similar workflows involving directory monitoring and standardized file naming conventions,

although they forfeit many of the advantages that BIDS offers through highly formalized storage logic and existing software solutions.

Finally, continuous directory monitoring is an optional part of the workflow. We selected the *watchman* service because it is free, open-source, and compatible with different OS architectures. Researchers may opt for similar file monitoring services (*fswatch*), scheduled execution (*cron* on Unix-type systems, *Task Scheduler* on Windows) at larger time intervals, or wish to execute the workflow manually after all.

Limitations and future perspectives

Integration of MRS into BIDS and its software ecosystem is a recent innovation that is continuously evolving. NIfTI-MRS conversion with *spec2nii* is supported for many common formats and sequence implementations, but may require further modification for custom sequences, either directly in the source code or with command line options to specify data dimensions and add metadata. Similarly, BIDScoin currently only supports a single 'type' of MRS data per bidsmap, i.e., two different bidsmaps need to be defined to accommodate study protocols with, e.g., one PRESS and one MEGA-PRESS dataset. In this case, the backbone script *main.py* needs to feature two instances of *bidscoiner* with the respective bidsmaps as input argument.

The mapping heuristics can be tailored to be powerful and versatile thanks to regular expressions in the filters, but still require a relatively consistent study protocol to be executed on the scanner. If sequence parameters, scan sequence names and order, and other data attributes and properties change frequently, the BIDScoin heuristics are likely to fail. This workflow is therefore primarily useful for long-running application studies with a fixed protocol, but less suitable for exploratory methodological piloting, which typically requires custom manual analysis anyway.

Quality control metrics remain a challenge for MRS. Model residuals, Cramer-Rao Lower Bounds, spectral linewidth, and signal-to-noise ratio all provide valuable information, but they often fail to reliably indicate the presence of major artefacts like lipid contamination or out-of-voxel echoes. Visual inspection of individual spectra therefore continues to be a key benchmark by which data quality is judged. This practice is clearly not scalable for large-scale studies or even single-site spectroscopic imaging. CSI/MRSI produces thousands of spectra per dataset, and the output is still often judged by the visual appearance of the final color map. Several quantitative metrics for quality assessment have been proposed (17–20), but have not been tested or implemented in mainstream MRS analysis. Automated scalable workflows like the one described here may facilitate uptake and benchmarking of these quantitative QC metrics.

Conclusion

In summary, we present an open-source end-to-end workflow for continuous automated analysis of MRS data, making use of recent advances in standardized file format and data

storage conventions (NIfTI-MRS, BIDS-MRS). Reduced manual analysis overhead can help simplify the integration of MRS into large-scale multi-modal imaging studies and clinical trials.

Acknowledgements

This work has been supported by NIH grants R00 AG062230, R21 EB033516, R01 EB016089, R01 EB023963, and P41 EB031771.

References

1. Soher BJ, Clarke WT, Wilson M, Near J, Oeltzschner G. Community-Organized Resources for Reproducible MRS Data Analysis. Magnetic Resonance in Medicine 2022;88:1959–1961 doi: 10.1002/mrm.29387.

2. Mikkelsen M, Barker PB, Bhattacharyya PK, et al. Big GABA: Edited MR spectroscopy at 24 research sites. NeuroImage 2017;159:32–45 doi: 10.1016/j.neuroimage.2017.07.021.

3. Clarke WT, Bell TK, Emir UE, et al. NIfTI-MRS: A standard data format for magnetic resonance spectroscopy. Magnetic Resonance in Medicine 2022;88:2358–2370 doi: 10.1002/mrm.29418.

4. Gorgolewski KJ, Auer T, Calhoun VD, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 2016;3:160044 doi: 10.1038/sdata.2016.44.

5. BIDS Extension Proposal 22 (BEP022): MRS (Magnetic Resonance Spectroscopy). Google Docs. https://docs.google.com/document/d/1pWCb02YNv5W-UZZja24fZrdXLm4X7knXMiZI7E2z7mY/edit?usp=sharing&usp=embed_facebook. Accessed October 27, 2022.

6. Halchenko Y, Goncalves M, Velasco P, et al. nipy/heudiconv: v0.11.4. 2022 doi: 10.5281/zenodo.7126719.

7. UNF Montreal. dcm2bids. https://github.com/UNFmontreal/Dcm2Bids. Accessed November 3rd 2022.

8. Tyszka M. BIDSKIT. https://github.com/jmtyszka/bidskit. Accessed November 3rd 2022.

9. Zwiers MP, Moia S, Oostenveld R. BIDScoin: A User-Friendly Application to Convert Source Data to Brain Imaging Data Structure. Frontiers in Neuroinformatics 2022;15 doi: 10.3389/fninf.2021.770608.

10. Oeltzschner G, Zöllner HJ, Hui SCN, et al. Osprey: Open-source processing, reconstruction & estimation of magnetic resonance spectroscopy data. Journal of Neuroscience Methods 2020;343:108827 doi: 10.1016/j.jneumeth.2020.108827.

11. Clarke WT, Stagg CJ, Jbabdi S. FSL-MRS: An end-to-end spectroscopy analysis package. Magnetic Resonance in Medicine 2021;85:2950–2964 doi: 10/gk5b34.

12. Wilson M. spant: An R package for magnetic resonance spectroscopy analysis. Journal of Open Source Software 2021;6:3646 doi: 10/gnkrv5.

13. Soher BJ, Semanchuk P, Todd D, Steinberg J, Young K. VeSPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis. In: Proc Int Soc Magn Reson Med. Vol. 19. ; 2011. p. 1410.

14. Stefan D, Cesare FD, Andrasescu A, et al. Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package. Meas. Sci. Technol. 2009;20:104035 doi: 2009090503131100.

15. Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magnetic resonance in medicine 1993;30:672–9 doi: 10.1002/mrm.1910300604.

16. 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. Magnetic Resonance in Medicine 2017;77:23–33 doi: 10.1002/mrm.26091.

17. Pedrosa de Barros N, McKinley R, Knecht U, Wiest R, Slotboom J. Automatic quality control in clinical 1H MRSI of brain cancer. NMR in Biomedicine 2016;29:563–575 doi: 10.1002/nbm.3470.

18. Pedrosa de Barros N, McKinley R, Wiest R, Slotboom J. Improving labeling efficiency in automatic quality control of MRSI data. Magnetic Resonance in Medicine 2017;78:2399–2405 doi: 10.1002/mrm.26618.

19. Pedrosa de Barros N, Slotboom J. Quality management in in vivo proton MRS. Analytical Biochemistry 2017;529:98–116 doi: 10.1016/j.ab.2017.01.017.

20. Tensaouti F, Desmoulin F, Gilhodes J, et al. Quality control of 3D MRSI data in glioblastoma: Can we do without the experts? Magnetic Resonance in Medicine 2022;87:1688–1699 doi: 10.1002/mrm.29098.