MASARYKOVA UNIVERZITA

Lékařská fakulta

# Optimalizace akvizice a metod analýzy obrazů magnetické rezonance

Habilitační práce

Mgr. Ing. Marek Dostál, Ph.D.

Brno 2024

MUNI
MED

# Obsah

# 1. Úvod

Od roku 1977, kdy vznikla první celotělová jaderná magnetická rezonance (MR) pro zobrazování lidské tkáně, uběhlo již téměř padesát let. Za tuto dobu se stalo vyšetření MR standardní radiologickou metodou. Tato zobrazovací modalita má své nezastupitelné místo při vyšetření centrální nervové soustavy (CSN), muskuloskeletálního systému (MSK), vyšetření pánve a dalších anatomických oblastí. Díky své fyzikální podstatě, rozvoji hardwaru a narůstající výpočetní síle nabízí širokou škálu modifikací, jak detekovat, rekonstruovat a analyzovat signál. Každým rokem přibývají nové zobrazovací sekvence, rekonstrukční algoritmy nebo analytické metody. I přes tento vývoj má ovšem metoda stále své limity a ty se většinou projevují jako různé artefakty v obraze. Větší část autorova výzkumu se týká difuzně váženého zobrazování (DWI), proto je teoretická část zaměřena na vznik a potlačení nejzávažnějších artefaktů právě z pohledu DWI. Následně navazuje vědecká část, kde autor prokazuje aktivní činnost v oblasti optimalizace vyšetřovacího protokolu pro dosažení co nejkvalitnějšího a reprodukovatelného zobrazení v oblasti krční míchy. Navazuje přehled analytických metod, jež autor buď vytvořil, či ověřuje jejich přesnost pro statistické vyhodnocení zobrazení CNS. V poslední kapitole jsou uvedeny vědecké práce klinického charakteru, na kterých se autor podílel a kde je aplikována některá z autorových metod. Předložená práce je koncipována jako soubor komentovaných publikací.

## 2. Seznam původních publikací autora

I. KOPŘIVOVÁ, Tereza, Miloš KEŘKOVSKÝ, Tomáš JŮZA, Václav VYBÍHAL, Tomáš ROHAN, Michal KOZUBEK a **Marek DOSTÁL**. Possibilities of Using Multi-b-value Diffusion Magnetic Resonance Imaging for Classification of Brain Lesions. Academic Radiology [online]. 2024, 31(1), 261–272. ISSN 1076-6332. Dostupné z: doi:10.1016/j.acra.2023.10.002

II. COHEN-ADAD, Julien, Eva ALONSO-ORTIZ, Mihael ABRAMOVIC, Carina ARNEITZ, Nicole ATCHESON, Laura BARLOW, Robert L. BARRY, Markus BARTH, Marco BATTISTON, Christian BÜCHEL, Matthew BUDDE, Virginie CALLOT, Anna J. E. COMBES, Benjamin DE LEENER, Maxime DESCOTEAUX, Paulo Loureiro DE SOUSA, **Marek DOSTÁL**, Julien DOYON, Adam DVORAK, Falk EIPPERT, Karla R. EPPERSON, Kevin S. EPPERSON, Patrick FREUND, Jürgen FINSTERBUSCH, Alexandru FOIAS, Michela FRATINI, Issei FUKUNAGA, Claudia A. M. Gandini WHEELER-KINGSHOTT, Giancarlo GERMANI, Guillaume GILBERT, Federico GIOVE, Charley GROS, Francesco GRUSSU, Akifumi HAGIWARA, Pierre-Gilles HENRY, Tomáš HORÁK, Masaaki HORI, James JOERS, Kouhei KAMIYA, Haleh KARBASFOROUSHAN, Miloš KEŘKOVSKÝ, Ali KHATIBI, Joo-Won KIM, Nawal KINANY, Hagen KITZLER, Shannon KOLIND, Yazhuo KONG, Petr KUDLIČKA, Paul KUNTKE, Nyoman D. KURNIAWAN, Slawomir KUSMIA, René LABOUNEK, Maria Marcella LAGANÀ, Cornelia LAULE, Christine S. LAW, Christophe LENGLET, Tobias LEUTRITZ, Yaou LIU, Sara LLUFRIU, Sean MACKEY, Eloy MARTINEZ-HERAS, Loan MATTERA, Igor NESTRASIL, Kristin P. O'GRADY, Nico PAPINUTTO, Daniel PAPP, Deborah PARETO, Todd B. PARRISH, Anna PICHIECCHIO, Ferran PRADOS, Àlex ROVIRA, Marc J. RUITENBERG, Rebecca S. SAMSON, Giovanni SAVINI, Maryam SEIF, Alan C. SEIFERT, Alex K. SMITH, Seth A. SMITH, Zachary A. SMITH, Elisabeth SOLANA, Yuichi SUZUKI, George TACKLEY, Alexandra TINNERMANN, Jan VALOŠEK, Dimitri VAN DE VILLE, Marios C. YIANNAKAS, Kenneth A. WEBER, Nikolaus WEISKOPF, Richard G. WISE, Patrik O. WYSS a Junqian XU. Generic acquisition protocol for quantitative MRI of the spinal cord. Nature Protocols [online]. 2021, 16(10), 4611–4632. ISSN 1750-2799. Dostupné z: doi:10.1038/s41596-021-00588-0

III. COHEN-ADAD, Julien, Eva ALONSO-ORTIZ, Mihael ABRAMOVIC, Carina ARNEITZ, Nicole ATCHESON, Laura BARLOW, Robert L. BARRY, Markus BARTH, Marco BATTISTON, Christian BÜCHEL, Matthew BUDDE, Virginie CALLOT, Anna J. E. COMBES, Benjamin DE LEENER, Maxime DESCOTEAUX, Paulo Loureiro DE SOUSA, **Marek DOSTÁL,** Julien DOYON, Adam DVORAK, Falk EIPPERT, Karla R. EPPERSON, Kevin S. EPPERSON, Patrick FREUND, Jürgen FINSTERBUSCH, Alexandru FOIAS, Michela FRATINI, Issei FUKUNAGA, Claudia A. M. GANDINI WHEELER-KINGSHOTT, Giancarlo GERMANI, Guillaume GILBERT, Federico GIOVE, Charley GROS, Francesco GRUSSU, Akifumi HAGIWARA, Pierre-Gilles HENRY, Tomáš HORÁK, Masaaki HORI, James JOERS, Kouhei KAMIYA, Haleh KARBASFOROUSHAN, Miloš KEŘKOVSKÝ, Ali KHATIBI, Joo-Won KIM, Nawal KINANY, Hagen H. KITZLER, Shannon KOLIND, Yazhuo KONG, Petr

KUDLIČKA, Paul KUNTKE, Nyoman D. KURNIAWAN, Slawomir KUSMIA, René LABOUNEK, Maria Marcella LAGANÀ, Cornelia LAULE, Christine S. LAW, Christophe LENGLET, Tobias LEUTRITZ, Yaou LIU, Sara LLUFRIU, Sean MACKEY, Eloy MARTINEZ-HERAS, Loan MATTERA, Igor NESTRASIL, Kristin P. O'GRADY, Nico PAPINUTTO, Daniel PAPP, Deborah PARETO, Todd B. PARRISH, Anna PICHIECCHIO, Ferran PRADOS, Àlex ROVIRA, Marc J. RUITENBERG, Rebecca S. SAMSON, Giovanni SAVINI, Maryam SEIF, Alan C. SEIFERT, Alex K. SMITH, Seth A. SMITH, Zachary A. SMITH, Elisabeth SOLANA, Y. SUZUKI, George TACKLEY, Alexandra TINNERMANN, Jan VALOŠEK, Dimitri VAN DE VILLE, Marios C. YIANNAKAS, Kenneth A. WEBER II, Nikolaus WEISKOPF, Richard G. WISE, Patrik O. WYSS a Junqian XU. Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers. Scientific Data [online]. 2021, 8(1), 219. ISSN 2052-4463. Dostupné z: doi:10.1038/s41597-021-00941-8

IV. COHEN-ADAD, Julien, Eva ALONSO-ORTIZ, Stephanie ALLEY, Maria Marcella LAGANA, Francesca BAGLIO, Signe Johanna VANNESJO, Haleh KARBASFOROUSHAN, Maryam SEIF, Alan C. SEIFERT, Junqian XU, Joo-Won KIM, René LABOUNEK, Lubomír VOJTÍŠEK, **Marek DOSTÁL**, Jan VALOŠEK, Rebecca S. SAMSON, Francesco GRUSSU, Marco BATTISTON, Claudia A. M. GANDINI WHEELER-KINGSHOTT, Marios C. YIANNAKAS, Guillaume GILBERT, Torben SCHNEIDER, Brian JOHNSON a Ferran PRADOS. Comparison of multicenter MRI protocols for visualizing the spinal cord gray matter. Magnetic Resonance in Medicine [online]. 2022, 88(2), 849–859. ISSN 1522-2594. Dostupné z: doi:10.1002/mrm.29249

V. **DOSTÁL, Marek**, Miloš KEŘKOVSKÝ, Eva KORIŤÁKOVÁ, Eva NĚMCOVÁ, Jakub STULÍK, Monika STAŇKOVÁ a Vladan BERNARD. Analysis of diffusion tensor measurements of the human cervical spinal cord based on semiautomatic segmentation of the white and gray matter. Journal of magnetic resonance imaging: JMRI [online]. 2018. ISSN 1522-2586. Dostupné z: doi:10.1002/jmri.26166

VI. **DOSTÁL, Marek**, Miloš KEŘKOVSKÝ, Erik STAFFA, Josef BEDNAŘÍK, Andrea ŠPRLÁKOVÁ-PUKOVÁ a Marek MECHL. Voxelwise analysis of diffusion MRI of cervical spinal cord using tract-based spatial statistics. Magnetic Resonance Imaging [online]. 2020 [vid. 2020-07-28]. ISSN 0730-725X. Dostupné z: doi:10.1016/j.mri.2020.07.008

VII. PATI, Sarthak, Ujjwal BAID, Brandon EDWARDS, Micah SHELLER, Shih-Han WANG, G. Anthony REINA, Patrick FOLEY, Alexey GRUZDEV, Deepthi KARKADA, Christos DAVATZIKOS, Chiharu SAKO, Satyam GHODASARA, Michel BILELLO, Suyash MOHAN, Philipp VOLLMUTH, Gianluca BRUGNARA, Chandrakanth J. PREETHA, Felix SAHM, Klaus MAIER-HEIN, Maximilian ZENK, Martin BENDSZUS, Wolfgang WICK, Evan CALABRESE, Jeffrey RUDIE, Javier VILLANUEVA-MEYER, Soonmee CHA, Madhura INGALHALIKAR, Manali JADHAV, Umang PANDEY, Jitender SAINI, John GARRETT, Matthew LARSON,

Robert JERAJ, Stuart CURRIE, Russell FROOD, Kavi FATANIA, Raymond Y. HUANG, Ken CHANG, Carmen BALAÑA, Jaume CAPELLADES, Josep PUIG, Johannes TRENKLER, Josef PICHLER, Georg NECKER, Andreas HAUNSCHMIDT, Stephan MECKEL, Gaurav SHUKLA, Spencer LIEM, Gregory S. ALEXANDER, Joseph LOMBARDO, Joshua D. PALMER, Adam E. FLANDERS, Adam P. DICKER, Haris I. SAIR, Craig K. JONES, Archana VENKATARAMAN, Meirui JIANG, Tiffany Y. SO, Cheng CHEN, Pheng Ann HENG, Qi DOU, Michal KOZUBEK, Filip LUX, Jan MICHÁLEK, Petr MATULA, Miloš KEŘKOVSKÝ, Tereza KOPŘIVOVÁ, **Marek DOSTÁL**, Václav VYBÍHAL, Michael A. VOGELBAUM, J. Ross MITCHELL, Joaquim FARINHAS, Joseph A. MALDJIAN, Chandan Ganesh Bangalore YOGANANDA, Marco C. PINHO, Divya REDDY, James HOLCOMB, Benjamin C. WAGNER, Benjamin M. ELLINGSON, Timothy F. CLOUGHESY, Catalina RAYMOND, Talia OUGHOURLIAN, Akifumi HAGIWARA, Chencai WANG, Minh-Son TO, Sargam BHARDWAJ, Chee CHONG, Marc AGZARIAN, Alexandre Xavier FALCÃO, Samuel B. MARTINS, Bernardo C. A. TEIXEIRA, Flávia SPRENGER, David MENOTTI, Diego R. LUCIO, Pamela LAMONTAGNE, Daniel MARCUS, Benedikt WIESTLER, Florian KOFLER, Ivan EZHOV, Marie METZ, Rajan JAIN, Matthew LEE, Yvonne W. LUI, Richard MCKINLEY, Johannes SLOTBOOM, Piotr RADOJEWSKI, Raphael MEIER, Roland WIEST, Derrick MURCIA, Eric FU, Rourke HAAS, John THOMPSON, David Ryan ORMOND, Chaitra BADVE, Andrew E. SLOAN, Vachan VADMAL, Kristin WAITE, Rivka R. COLEN, Linmin PEI, Murat AK, Ashok SRINIVASAN, J. Rajiv BAPURAJ, Arvind RAO, Nicholas WANG, Ota YOSHIAKI, Toshio MORITANI, Sevcan TURK, Joonsang LEE, Snehal PRABHUDESAI, Fanny MORÓN, Jacob MANDEL, Konstantinos KAMNITSAS, Ben GLOCKER, Luke V. M. DIXON, Matthew WILLIAMS, Peter ZAMPAKIS, Vasileios PANAGIOTOPOULOS, Panagiotis TSIGANOS, Sotiris ALEXIOU, Ilias HALIASSOS, Evangelia I. ZACHARAKI, Konstantinos MOUSTAKAS, Christina KALOGEROPOULOU, Dimitrios M. KARDAMAKIS, Yoon Seong CHOI, Seung-Koo LEE, Jong Hee CHANG, Sung Soo AHN, Bing LUO, Laila POISSON, Ning WEN, Pallavi TIWARI, Ruchika VERMA, Rohan BAREJA, Ipsa YADAV, Jonathan CHEN, Neeraj KUMAR, Marion SMITS, Sebastian R. VAN DER VOORT, Ahmed ALAFANDI, Fatih INCEKARA, Maarten M. J. WIJNENGA, Georgios KAPSAS, Renske GAHRMANN, Joost W. SCHOUTEN, Hendrikus J. DUBBINK, Arnaud J. P. E. VINCENT, Martin J. VAN DEN BENT, Pim J. FRENCH, Stefan KLEIN, Yading YUAN, Sonam SHARMA, Tzu-Chi TSENG, Saba ADABI, Simone P. NICLOU, Olivier KEUNEN, Ann-Christin HAU, Martin VALLIÈRES, David FORTIN, Martin LEPAGE, Bennett LANDMAN, Karthik RAMADASS, Kaiwen XU, Silky CHOTAI, Lola B. CHAMBLESS, Akshitkumar MISTRY, Reid C. THOMPSON, Yuriy GUSEV, Krithika BHUVANESHWAR, Anousheh SAYAH, Camelia BENCHEQROUN, Anas BELOUALI, Subha MADHAVAN, Thomas C. BOOTH, Alysha CHELLIAH, Marc MODAT, Haris SHUAIB, Carmen DRAGOS, Aly ABAYAZEED, Kenneth KOLODZIEJ, Michael HILL, Ahmed ABBASSY, Shady GAMAL, Mahmoud MEKHAIMAR, Mohamed QAYATI, Mauricio REYES, Ji Eun PARK, Jihye YUN, Ho Sung KIM, Abhishek MAHAJAN, Mark MUZI, Sean BENSON, Regina G. H. BEETS-TAN, Jonas TEUWEN, Alejandro HERRERA-TRUJILLO, Maria TRUJILLO, William ESCOBAR,

Ana ABELLO, Jose BERNAL, Jhon GÓMEZ, Joseph CHOI, Stephen BAEK, Yusung KIM, Heba ISMAEL, Bryan ALLEN, John M. BUATTI, Aikaterini KOTROTSOU, Hongwei LI, Tobias WEISS, Michael WELLER, Andrea BINK, Bertrand POUYMAYOU, Hassan F. SHAYKH, Joel SALTZ, Prateek PRASANNA, Sampurna SHRESTHA, Kartik M. MANI, David PAYNE, Tahsin KURC, Enrique PELAEZ, Heydy FRANCO-MALDONADO, Francis LOAYZA, Sebastian QUEVEDO, Pamela GUEVARA, Esteban TORCHE, Cristobal MENDOZA, Franco VERA, Elvis RÍOS, Eduardo LÓPEZ, Sergio A. VELASTIN, Godwin OGBOLE, Mayowa SONEYE, Dotun OYEKUNLE, Olubunmi ODAFE-OYIBOTHA, Babatunde OSOBU, Mustapha SHU'AIBU, Adeleye DORCAS, Farouk DAKO, Amber L. SIMPSON, Mohammad HAMGHALAM, Jacob J. PEOPLES, Ricky HU, Anh TRAN, Danielle CUTLER, Fabio Y. MORAES, Michael A. BOSS, James GIMPEL, Deepak Kattil VEETTIL, Kendall SCHMIDT, Brian BIALECKI, Sailaja MARELLA, Cynthia PRICE, Lisa CIMINO, Charles APGAR, Prashant SHAH, Bjoern MENZE, Jill S. BARNHOLTZ-SLOAN, Jason MARTIN a Spyridon BAKAS. Federated learning enables big data for rare cancer boundary detection. Nature Communications [online]. 2022, 13, 7346. ISSN 2041-1723. Dostupné z: doi:10.1038/s41467-022-33407-5

VIII. KEŘKOVSKÝ, Miloš, Jakub STULÍK, Ivana OBHLÍDALOVÁ, Petra PRAKSOVÁ, Josef BEDNARIK, **Marek DOSTAL**, Matyáš KUHN, Andrea SPRLÁKOVÁ a Marek MECHL. State-of-the-Art MRI Techniques for Multiple Sclerosis. Česká a slovenská neurologie a neurochirurgie [online]. 2017, 80/113, 647–659. Dostupné z: doi:10.14735/amcsnn2017647

IX. **DOSTÁL, Marek**, Miloš KEŘKOVSKÝ, Jakub STULÍK, Josef BEDNAŘÍK, Petra PRAKSOVÁ, Monika HULOVÁ, Yvonne BENEŠOVÁ, Eva KORIŤÁKOVÁ, Andrea ŠPRLÁKOVÁ-PUKOVÁ a Marek MECHL. MR Diffusion Properties of Cervical Spinal Cord as a Predictor of Progression to Multiple Sclerosis in Patients with Clinically Isolated Syndrome. Journal of Neuroimaging [online]. 2021, 31(1), 108–114. ISSN 1552-6569. Dostupné z: doi:https://doi.org/10.1111/jon.12808

X. STULÍK, Jakub, Miloš KEŘKOVSKÝ, Matyáš KUHN, Monika SVOBODOVÁ, Yvonne BENEŠOVÁ, Josef BEDNAŘÍK, Andrea ŠPRLÁKOVÁ-PUKOVÁ, Marek MECHL a **Marek DOSTÁL**. Evaluating Magnetic Resonance Diffusion Properties Together with Brain Volumetry May Predict Progression to Multiple Sclerosis. Academic Radiology [online]. 2022 [vid. 2022-06-14]. ISSN 1076-6332. Dostupné z: doi:10.1016/j.acra.2021.12.015

# 3. Techniky zobrazování CNS pomocí difuzně vážených obrazů

## 3.1. Teoretické principy DWI

Zobrazování magnetickou rezonancí (MR) využívá mírně odlišných magnetických vlastností jader vodíku v závislosti na okolním magnetickém poli, kde například vodíky navázané na uhlovodíkových řetězcích tuků mají rozdílnou T1 relaxivitu[1] v porovnání s vodíky navázanými na kyslíku ve vodě. Pomocí vnějších magnetických impulzů lze vychylovat magnetické momenty jader z rovnovážných poloh, přičemž cívky detekují signál, který vzniká při jejich návratu do rovnovážné polohy. Velikost signálu je dána počtem vodíkových jader v objemu (protonová hustota) a dvěma odlišnými relaxačními ději, kdy rychlost jejich relaxace je charakterizována T1 a T2 relaxačními časy. Prostorovou distribuci tohoto signálu můžeme prostřednictvím šedotónové stupnice vykreslit do obrazu a podle toho, jaký děj se na vzniklém signálu dominantně podílí, mluvíme pak o T1 nebo T2 váženém obraze.

Časový průběh koncentrace látky *c* v důsledku jejího difundování nám popisuje druhý Fickův zákon, jehož diferenciální tvar je:

$$\frac{d\,c}{d\,t} = D\Delta c \tag{1}$$

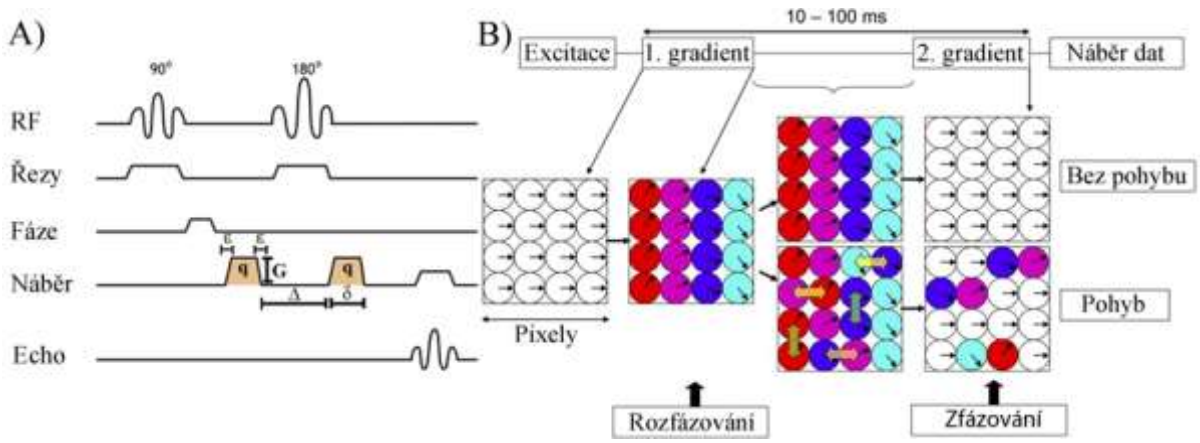
Je přímo úměrný koeficientu difuze *D*, který je přímo závislý na teplotě a nepřímo na viskozitě a velikosti částic, dle Stokesovy-Einsteinovy rovnice:

$$D = \frac{k_b T}{6\pi\eta r} \tag{2}$$

Signál indukovaný detekčními cívkami je sumou signálů všech jader vodíku v objemu okolo cívky. Aby bylo možné jej rozdělit podle prostorových souřadnic (x, y, z) a tím vytvořit obraz, je třeba aplikovat relativně komplikovaný sled gradientních[2] magnetických impulzů, které prostorové souřadnice zakódují pomocí mírně odlišných frekvencí a fází precesního pohybu. Tento proces je nazýván pozičním kódováním a je nedílnou součástí tvorby každého obrazu MR.

Použijeme-li další dva gradientní impulzy po vzoru Stejskal-Tennerovy metody [1], můžeme signál z jednotlivých voxelů nejprve řízeně rozfázovat a po uplynutí určité doby opět zfázovat. Jestli dojde v průběhu tohoto procesu k pohybu vodíkových jader, výsledek zpětného zfázování nebude totožný s původním stavem, což bude mít za následek pokles výsledného signálu v porovnání s obrazem bez difuzního vážení (bez aplikace difuzních gradientních pulsů). Tento proces je zachycen na obrázku 1.

[1] Rychlost návratu vektoru magnetizace do rovnovážné polohy poté, co z této rovnovážné polohy byl vektor magnetizace vychýlen.
[2] V prostoru proměnlivých.




Obr. 1. A) Schéma zapínání gradientních a RF impulzů pro difuzně vážené obrazy. Gradientní impulz odpovědný za difuzní vážení je charakterizován amplitudou G, délkou trvání δ, časem mezi difuzními impulzy Δ a strmostí zapnutí a vypnutí impulzu ε. Převzato z [2]. B) Ukázka změny signálu způsobená difuzí molekul. První difuzní gradient vede k rozfázování precese v konkrétním směru (v tomto případě zprava doleva). Poté je ponechán prostor pro samotnou difuzi a 180° refokusační impulz. Molekuly jsou následně zfázovány druhým difuzním gradientem. Jestliže k difuzi nedochází, pozice jednotlivých molekul je stejná jako před prvním difuzním gradientem a v obraze nedojde ke změně signálu. Pokud k difuzi dochází, molekuly pohybující se ve směru difuzního gradientu (zprava doleva) budou mít po druhém difuzním gradientu rozdílnou fázi a díky tomu dojde ke změně signálu. Molekuly, které se pohybují ve směru kolmém na difuzní gradient, ke změně signálu nepřispívají, a jejich pohyb tudíž není detekován. Převzato z [3].

Míru difuzního vážení v obraze popisuje tzv. b-hodnota neboli b-faktor [s.mm$^{-2}$] (rovnice 3), jež zjednodušeně ukazuje, kolik času necháme molekulám vodíku k difuzi a jak moc rozfázujeme precesní pohyb. Je-li b-hodnota nízká (10 – 100 s.mm$^{-2}$), na úbytku signálu v obraze se podílejí pouze ta jádra, která se pohybují rychle. Čím větší je b-hodnota, tím dáváme více času jádrům k pohybu, čímž jsme schopni charakterizovat jiné pohyby. Teoreticky můžeme volit libovolnou velikost b-hodnot, ale v praxi jsme limitováni primárně kvalitou gradientního systému přístroje MR. Při zanedbání prostorových gradientů můžeme b-hodnotu určit dle vztahu [4]:

$$b = \gamma G^2 \left[\delta^2 \left(\Delta - \frac{\delta}{3}\right) + \frac{\varepsilon^2}{30} - \frac{\delta\varepsilon^2}{6}\right] \tag{3}$$

*γ* představuje gyromagnetický poměr, *G* amplitudu gradientního pulzu, *Δ* čas mezi gradientními pulzy, *δ* dobu a *ε* rychlost sepnutí gradientního pulzu (obr. 1 – A). Následují různé varianty, jak difuzní signál dále analyzovat. Nejjednodušší a v klinické praxi etablovaný přístup je popis difuze pomocí monoexponenciální funkce (rovnice 4), kde $S_{b0}$ je velikost signálu bez difuzního vážení, $S_b$ je velikost signálu s difuzním vážením o velikosti *b* a v exponentu vystupuje míra difuzního vážení *b* a koeficient zdánlivé difuze (*ADC*)[3]:

$$S_b = S_{b0} e^{-b*ADC} \tag{4}$$

Vypočítané hodnoty *ADC* pro jednotlivé voxely se zobrazují jako tzv. ADC mapy a jejich hodnoty mohou vypovídat o typu postižení [5, 6]. Dle zobrazované patologie či anatomie existují obecná doporučení,

[3] Z angl. apparent diffusion coeficient.

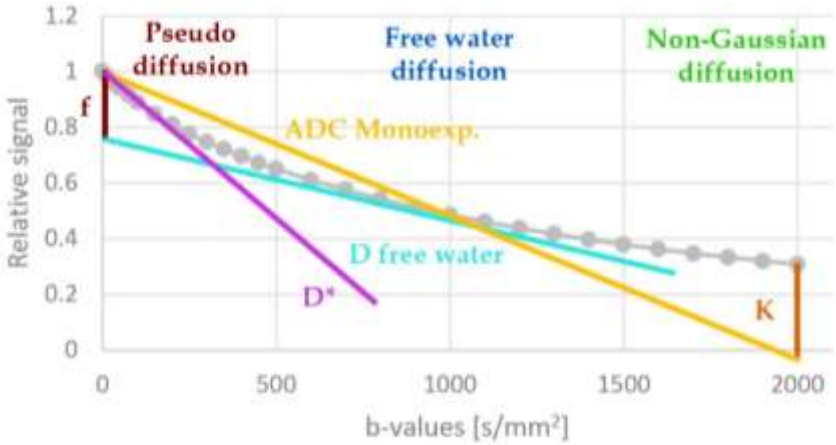
jakou míru difuzního vážení volit (vyšetření mozku b = 1000 s.mm$^{-2}$, prostaty b > 1500 s.mm$^{-2}$ atp.), ale neexistuje jednoznačná hodnota, protože míra difuze je ovlivněna mnoha faktory [7].

Monoexponenciální model není použitelný pro vysoké hodnoty b-faktoru (b > 2000 s.mm$^{-2}$), kdy již neodpovídá skutečnosti, protože difuze se v této oblasti nechová dle Gaussova rozdělení [8]. Proto se může tento model upravit o difuzně kvadratický člen s parametrem kurtozy (špičatosti) *K*, který je bezrozměrný a charakterizuje odklon od monoexponenciáloního modelu v oblastech vysokých b-hodnot.

$$S_b = S_{b0} e^{(-bD+(bD)^2 K/6)} \tag{5}$$

V opačném případě se tento model také nedá použít, a to pro velmi nízké hodnoty b-faktoru (b < 200 s.mm$^{-2}$), kdy signál obsahuje informaci o rychlých difuzních dějích. Jedním ze zdrojů rychlého pohybu vodíkových jader je perfuze krve, což není difuzní děj. Proto se někdy o oblasti nízkých b-hodnot mluví jako o pseudodifuzi a od difuze ji můžeme oddělit pomocí modelu zvaného intravoxel incoherent motion (IVIM). Ten je popsán biexponenciální rovnicí (6), kde *f* je koeficient perfuzní frakce, *D** je pseudodifuzní koeficient (v některých studiích označován také jako rychlá difuze) a *D* je difuzní koeficient [9]. Obrázek 2 souhrnně zobrazuje rozdíly jednotlivých modelů.

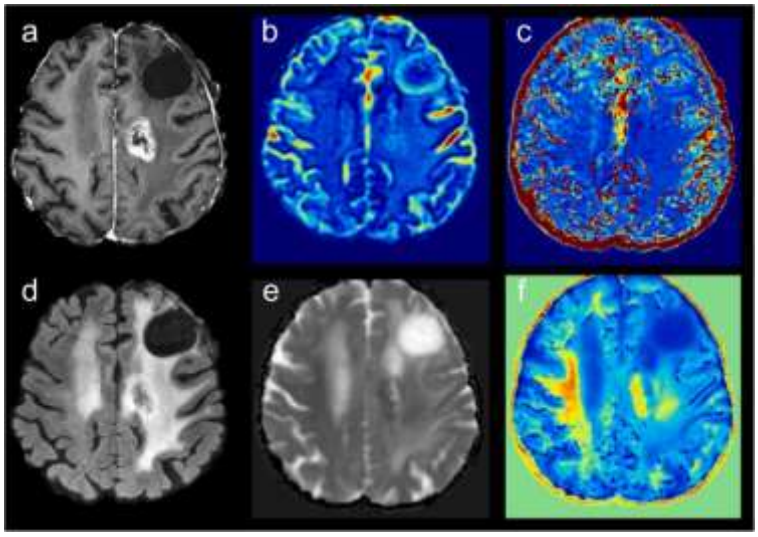
$$S_b = S_{b0}(f e^{-b*D^*} + (1-f)e^{-b*D}) \tag{6}$$




Obr. 2. Komplexní ukázka různých modelů analýzy difuzně váženého obrazu. Šedá křivka znázorňuje DWI signál. Žlutá křivka (ADC Monoexp.) znázorňuje model dle rovnice 4. Oranžová úsečka (K) znázorňuje odklon ADC modelu od reálného signálu v oblasti vysokých b-hodnot dle rovnice 5. Červená úsečka (f), fialová (D*) a tyrkysová (D free water) jsou výstupem modelu IVIM dle rovnice 6, přičemž D* charakterizuje část signálu z difuze vody v krevním řečišti, D charakterizuje část signálu bez signálu z krevního řečiště a f charakterizuje odklon parametru D od monoexponenciálního modelu ADC.

O možnostech a limitacích těchto jednotlivých modelů pojednával z pohledu klasifikace mozkových tumorů autorův přehledový článek **I. (Kopřivová et al. 2024)**. V tomto textu autor rozebírá a obrazově demonstruje možnosti aplikace modelů pro diferenciaci mozkových tumorů od netumorózních ložisek,

glioblastomů od primárních CNS lymfomů nebo od metastáz. Taktéž zkoumá možnosti gradingu gliomů (low-grade vs high-grade) či odlišení progrese od pseudoprogrese při pooperačním sledování pacientů (obr. 3).

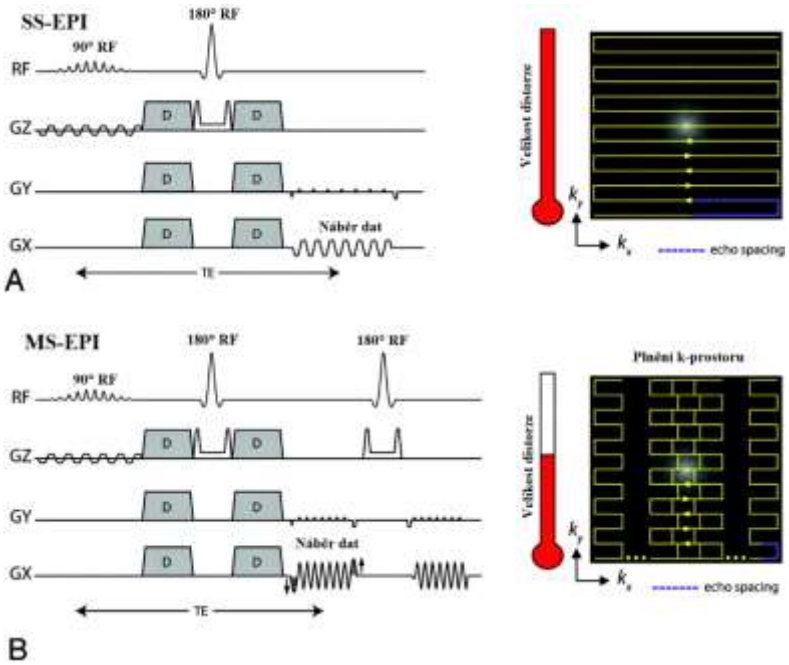



Obr. 3. Postkontrastně sytící se léze v bílé hmotě levé mozkové hemisféry odpovídající pseudoprogresi u pacienta s glioblastomem (grade IV, divoký typ IDH) po chirurgické resekci a radioterapii. A) - T1-w postkontrastní axiální obraz, b) - f mapa, c) - D* mapa, d) - FLAIR axiální obraz, e) - D mapa, f) - K mapa. Sytící se léze (a) se nachází dorzálně od nesytící se pooperační oblasti v levém frontálním laloku; rozsáhlá leukoencefalopatie je patrná v obraze FLAIR (d). FLAIR, fluid-attenuated inversion recovery; IDH – isocitrátdehydrogenáza.

Další možností, jak můžeme přistupovat k difuznímu signálu, je sledovat jeho prostorovou orientaci. Difuze je obecně náhodný děj. To ale platí pouze v případě volné tekutiny, což do určité míry můžeme pozorovat například v mozkomíšním moku o větším objemu (postranní komory mozku). V tkáni je difuze omezená přítomností buněk, které mají různý tvar a prostorovou orientaci. Máme-li pravidelnou strukturu (např. vlákna bílé hmoty), má voda tendenci difundovat ve směru menšího odporu, takže podél vláken. Menší difuze je pak ve směru kolmém na tato vlákna. Toto anizotropní chování difuze vody se dá kvantifikovat pomocí difuzně vážených obrazů. Je ovšem nutné přídavné difuzní gradientní pulzy aplikovat minimálně v 6 rozdílných směrech. Pak je třeba dopočítat tenzor difuze a jeho skalární parametry, jako je například střední difuzivita (MD) nebo frakční anizotropie (FA). Když v tkáni dojde k narušení pravidelné vláknité struktury (např. v případě bílé hmoty k demyelinizaci), anizotropie difuze se změní. Poklesne totiž její odpor ve směru kolmém na vlákna, tudíž voda začne ve větší míře difundovat i v tomto předtím nepreferovaném směru (dojde k poklesu FA). Jestli používáme velikost difuzního vážení okolo b = 1000 $s.mm^{-2}$, pak o tomto přístupu mluvíme coby o zobrazení tenzoru difuze (DTI). Pokud používáme i vyšší b hodnoty, tak odklon DTI parametrů od Gaussova rozdělení můžeme charakterizovat, a to obdobně jako v předchozích odstavcích pomocí kurtozy (špičatosti). Hovoříme tedy o zobrazení kurtozy tenzoru difuze (DKI) [3].

## 3.2. Artefakty DWI

Zobrazování difuze pomocí MR přináší spoustu úskalí. Kromě již výše zmiňovaných složitějších analytických přístupů je nutné věnovat pozornost samotné zobrazovací sekvenci. Jelikož se snažíme detekovat velmi rychlý děj, tak i obraz musí být vytvořen rychle (naplnění k-prostoru). Proto se v klinické praxi standardně využívá tzv. single-shot echo-planar imaging (SS-EPI) snímání dat. Tato metoda je single-shot, což znamená, že během jedné excitace je vytvořen jeden obraz. EPI zaplňování k-prostoru probíhá pomocí „šňůry“ velmi rychlých gradientních pulzů (obr. 5 – A). Takto rychlý náběr avšak způsobuje prostorové distorze (deformace) obrazu (obr. 6), tvorbu vířivých proudů a klade vysoké nároky na gradientní a shimovací[4] vybavení stroje MR [11, 12].

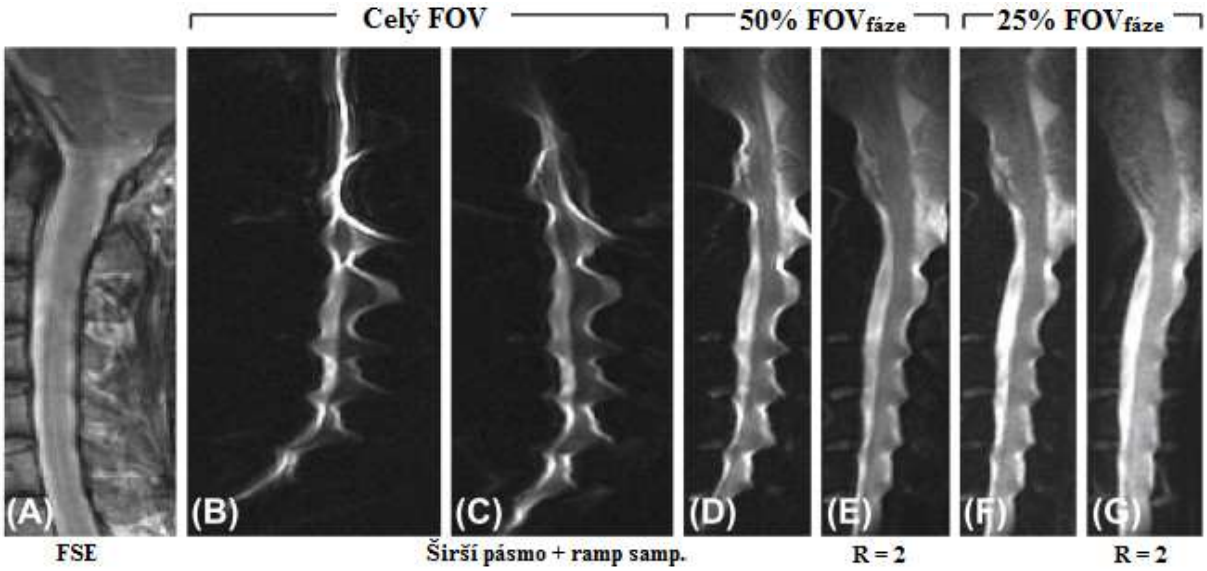



Obr. 5. Ukázka časové souslednosti jednotlivých impulzů v případě A) single-shot (SS) a B) multi-shot (MS) echo-planar imaging (EPI) sekvence (levá část) a průběh zaplňovaní k-prostoru (pravá část). RF – radiofrekvenční. Převzato z [13].

Existuje několik přístupů, jak tyto distorzní artefakty potlačit, nicméně nikdy se jich zcela nezbavíme. Jednou z možností je využití multi-shot přístupu, kdy k-prostor není zaplněn během jedné, ale více excitací (obr. 5 – B). Dalšími variantami je zkrácení času mezi jednotlivými snímáními. Toho se může dosáhnout pomocí zvětšení šířky frekvenčního pásma či ramp samplingem (k zaplňování k-prostoru dochází již při náběhu čtecího gradientu). Dále k tomu také lze využít paralelní techniky, zmenšení obrazu (FOV) ve směru fázového kódování nebo můžeme zkombinovat více metod. Ovšem při zmenšování fázového FOV vzniká artefakt překlopení. Tento problém v posledních letech byl vyřešen novým způsobem excitace tkáně, kdy se na místo jednoho excitačního pulzu použijí dva prostorově

[4] Shimming slouží k zvýšení homogenity magnetického pole v zobrazované oblasti [10].

selektivní radiofrekvenční pulzy (2D RF excitace). Kromě eliminace takového artefaktu jsou rovněž efektivně potlačeny artefakty distorzí. Výsledky jednotlivých metod jsou demonstrovány na sagitálním zobrazení krční míchy (obr. 6, převzato z [14]).




Obr. 6. Ukázka distorze obrazu krční míchy při různých nastaveních single-shot echo-planar (SS-EPI) sekvence. A) referenční obraz anatomie krční míchy pořízený sekvencí rychlého spinového echa (FSE). B) SS-EPI obraz bez korekcí distorze. Na obraze snímaném při širším frekvenčním pásmu a ramp samplingu (C) je patrný drobný úbytek artefaktů. Potlačení artefaktů je velmi dobré už jen při zmenšení fázového FOV na polovinu (D, E) či čtvrtinu (F, G). K potlačení dojde také, když k tomu ještě využijeme paralelní techniku (E, G). Převzato z [14].

## 3.3. Snímání DWI

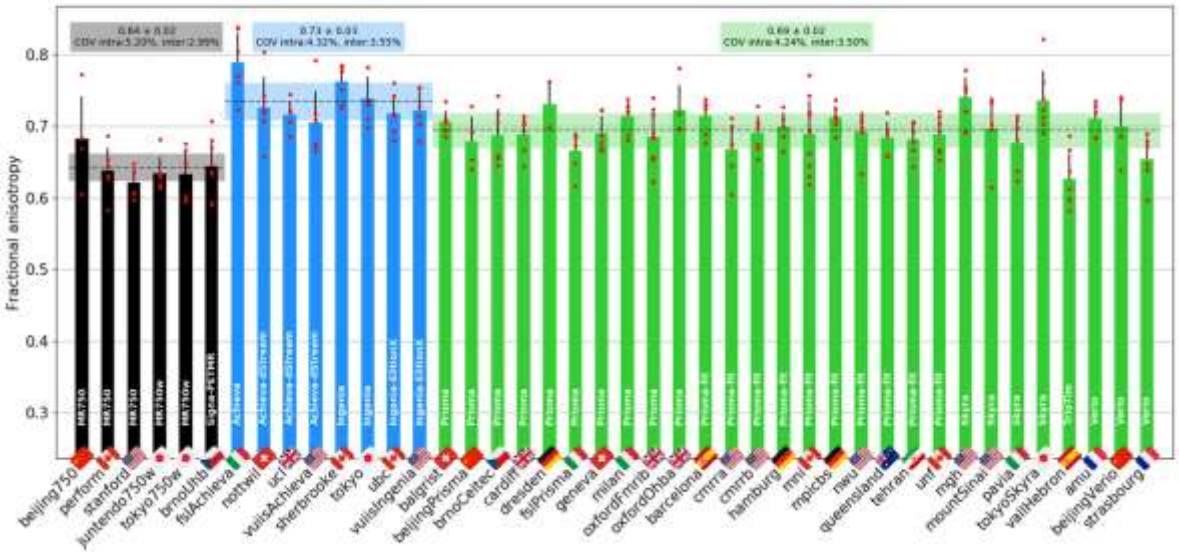
Krční mícha je obecně pro zobrazení MR komplikovanou strukturou už jen proto, že je malá (je třeba dobré prostorové rozlišení a poměr signálu k šumu). K tomu je obklopena kostmi a blízko ní je i vzduch z dýchacích cest. Všechny tyto struktury mají rozdílnou magnetickou susceptibilitu, čímž je snížena homogenita magnetického pole, což zvětšuje distorzní artefakty. Do obrazu se také promítají artefakty z dýchání, polykání nebo pulzace mozkomíšního moku.

Z tohoto důvodu se autor aktivně podílel na optimalizaci a testování multicentrického protokolu pro zobrazování krční míchy, kde došlo k optimalizaci anatomického (T1, T2, T2* atp.) i difuzně váženého zobrazení (DTI). Výstup z tohoto projektu byl publikován v mezinárodním časopise **II. (Cohen-Adad et al. 2021b).** Pro tři hlavní výrobce strojů MR jsou doporučené protokoly volně k dispozici na online databázi (GitHub[5]). Výsledkem této práce je sada doporučení, jak správně polohovat pacienta při ukládání do stroje, jak ne/rotovat akviziční roviny, jak umístit tzv. shim box a nastavit jednotlivé

[5] https://github.com/spine-generic/protocols

parametry. Důraz je kladen na ne/výhody jednotlivých sekvencí, jejich klinické využití a přehled častých problémů a jejich řešení.

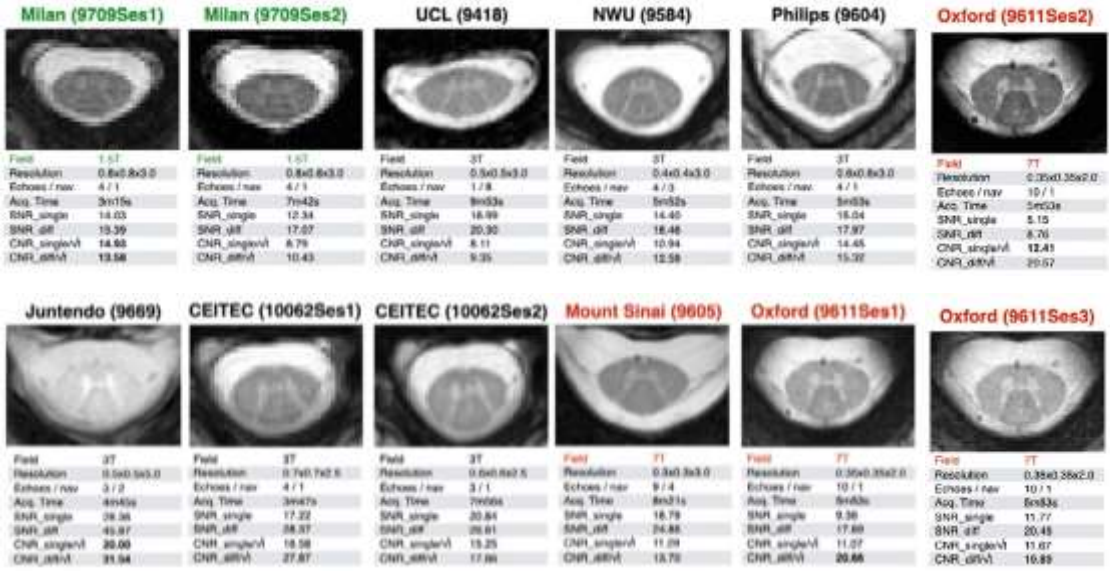
Jak bylo uvedeno v části zabývající se analýzou difuzně vážených obrazů, výstupem bývají různé koeficienty difuze. Díky komplexnosti snímání MR je otázka reprodukovatelnosti kvantitativního vyšetření zcela na místě [15, 16, 17 s. 202]. Na kvalitu detekovaného signálu má vliv celá řada proměnných, které se liší nejenom pacient od pacienta, či pracoviště od pracoviště, ale taky mezi jednotlivými výrobci i mezi stroji MR jednoho výrobce (použité detekční cívky, nastavení parametrů snímání, gradientní a shimovací systém, typ excitace, rozdíly v časování a amplitudách jednotlivých pulzů, uložení pacienta, pohyb a dech pacienta, odstínění vyšetřovny od rušivých signálů atp.). Proto se autor aktivně účastnil jak akvizice, tak analýzy multicentrického souboru dat vyšetření krční míchy pomocí optimalizovaného protokolu. Chtěl prokázat míru reprodukovatelnosti tohoto zobrazovacího protokolu **III. (Cohen-Adad et al. 2021a)**. Z výsledků měření stejného subjektu na 19 různých přístrojích MR od 3 výrobců a se stejným vyšetřovacím protokolem je patrné, že již měření plochy odpovídajících si etáží míchy se různí mezi výrobci. Koeficient variability (CoV) u stejného výrobce se pohybuje mezi 0,9 % - 2,3 %. Pokud se budeme zabývat měřením plochy šedé hmoty, tak reprodukovatelnost klesá a koeficient variability vzroste na 2,5 % - 3,4 %. Reprodukovatelnost poměru magnetizačního transferu byla nejnižší ze sledovaných parametrů (CoV 3,6 % - 8,0 %) naopak reprodukovatelnost DTI měření charakterizovaného frakční anizotropií byla relativně vysoká (CoV 0,8 % - 4,5 %). Výsledky měření DTI na různých subjektech jsou demonstrovány na obrázku 7, kde je patrný statisticky významný rozdíl mezi výrobci ($p < 0,01$). Inter CoV byl menší než 5,21 % a intra CoV byl menší než 3,56 %, což se dá považovat za velmi dobrou reprodukovatelnost.




Obr. 7. Srovnání hodnot frakční anizotropie mezi třemi výrobci (GE – černá (vlevo), Philips – modrá (střed) a Siemens – zelená (vpravo)) na základě měření pěti zdravých dobrovolníků. Průměrné hodnoty, +- směrodatná

odchylka a odpovídající koeficienty variace (CoV) pro inter a intra reprodukovatelnost jsou uvedeny v horní části grafů u jednotlivých výrobců.

Výše zmíněná práce byla zaměřena na optimalizaci obecného zobrazovacího protokolu. V současnosti nejpoužívanější sekvencí pro anatomické zobrazení krční míchy v transversální rovině je gradientní T2* vážená sekvence s rekonstrukcí obrazu při různých echo časech. Tyto obrazy se posléze mohou kombinovat (sčítat, průměrovat…) pro zvýšení poměrů signálu a kontrastu k šumu. Volba echo časů a jejich počtu není triviální. Přináší s sebou různé komplikace (větší zastoupení artefaktů, delší vyšetřovací čas atp.) a nastavení je omezeno hardwarovým vybavením konkrétního stroje MR. V této oblasti se autor podílel na jedné z největších multicentrických studií **IV. (Cohen-Adad et al. 2022)**, kde je sledován vliv výrobce a velikosti statického magnetického pole stroje MR (obr. 8) na výslednou míru signálu a kontrastu obrazu. Na tyto dva parametry má vliv také počet snímaných ech T2* vážené sekvence, a proto bylo třeba nalézt jejich optimální počet. Výsledky jsou porovnány s teoretickým modelem i fantomovým měřením.




Obr. 8. Ukázka výstupů z různých strojů MR o velikosti magnetické indukce 1,5, 3 a 7 T společně s vyhodnocením kvality obrazu (poměr signálu (SNR) a kontrastu (CNR) vůči šumu, oboje čím vyšší, tím lepší).

Na základě výše uvedeného je patrné, že zobrazování krční míchy ať už morfologické, či difuzně vážené je celosvětově komplikovanou záležitostí. Autor habilitační práce se aktivně podílí na tvorbě a verifikaci mezinárodního zobrazovacího protokolu, který by měl přinášet optimální obrazové informace uniformně napříč stroji MR různých výrobců, magnetických polí a vybavení.

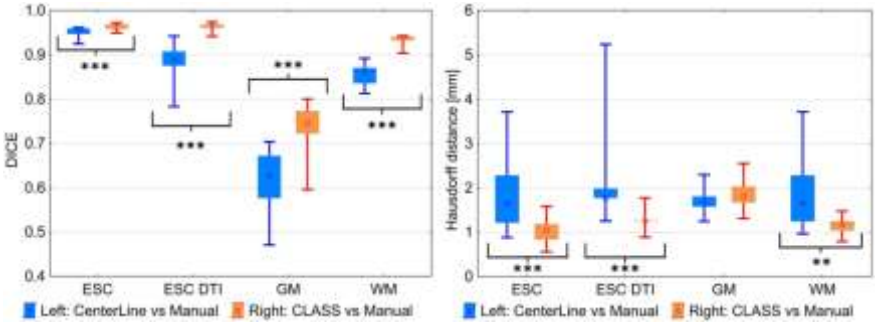
# 4. Analýza obrazu

Když máme kvalitně vyšetřené pacienty, tak je potřeba provést nezbytné kroky před statistickým hodnocením difuzních charakteristik. U většiny vyšetření musíme nejprve od sebe diferencovat různé

struktury. Nejčastěji se v případě CNS odlišuje mozkomíšní mok, bílá a šedá hmota a případně patologické struktury jako nekrotická tkáň, sytící se oblast po aplikaci kontrastní látky nebo demyelinizační léze. Tomuto procesu se říká segmentace nebo také labeling a existuje celá řada algoritmů [18]. Ovšem pro velmi hrubé dělení můžeme použít tři základní kategorie.

První je sice nejpracnější, ale často nejpřesnější. Jedná se o manuální značení jednotlivých voxelů, vrstvu po vrstvě, což v případě snímků o malých tloušťkách je velmi zdlouhavý proces. Ve většině případů je manuální segmentace zkušeným hodnotitelem (většinou radiologem) považována za zlatý standard, vůči kterému se porovnávají zbylé dvě metody. Druhou variantou je mírné zapojení automatizace s určitým manuální vstupem, kdy můžeme mluvit o semi-automatických segmentačních algoritmech. Tyto metody nejen že práci zrychlují a ulehčují, ale také minimalizují do určité míry subjektivní vlivy jednotlivých hodnotitelů. Poslední možností jsou již plně automatické segmentační algoritmy, při nichž je uživatelský vstup téměř nulový. Do této kategorie mimo jiné spadá aktuálně velmi oblíbené strojové učení.

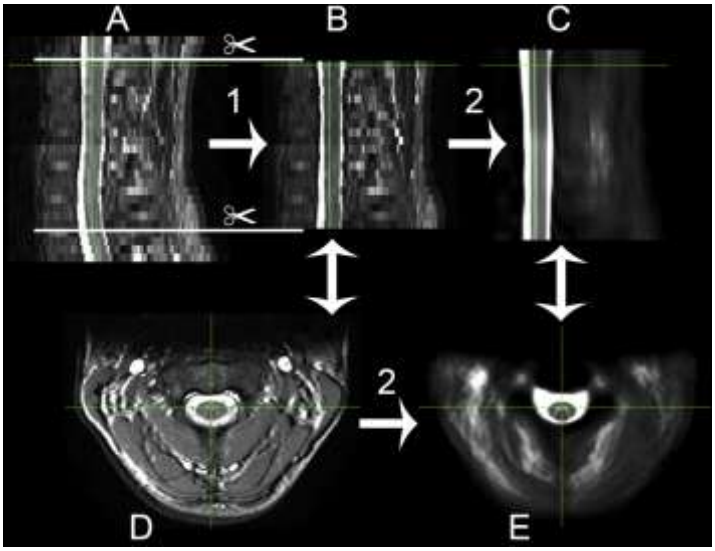
K segmentaci mozku bez větších patologií (např. při Alzheimerově chorobě) je aktuálně k dispozici celá řada jak výzkumných, tak i certifikovaných algoritmů pro klinické použití [19, 20]. V případě segmentace krční míchy je nicméně množství algoritmů výrazně nižší [21]. Studium šedé hmoty v oblasti krční míchy je užitečné například u onemocnění amyotrofickou laterální sklerózou a při kvantifikaci rozsahu lézí u roztroušené sklerózy. Dále také zlepšuje interpretaci funkční magnetické rezonance [22, 23]. Proto autor vytvořil metodiku jak segmentace celé krční míchy, tak i separace na šedou a bílou hmotu. Tento postup využívá semi-automatickou segmentaci, kterou autor následně aplikoval na analýzu difuzních dat krční míchy, a výsledky publikoval v mezinárodním časopise **V. (Dostál et al. 2018)**. V této práci autor nejprve důkladně krok za krokem popisuje metodiku segmentace T2* vážených obrazů, jež pro tyto potřeby vytvořil. Následně je ověřována na obrazech 20 neurologicky zdravých dobrovolnících, kde je přesnost srovnávána jednak s manuální segmentací provedenou zkušenými radiology a autorem a jednak s veřejně dostupnou plně automatizovanou metodou. Byla ověřena intra i inter reprodukovatelnost jednotlivých přístupů, kdy byly čtyřikrát provedeny všechny segmentační postupy na sedmi náhodných subjektech. Tímto přístupen vznikly stovky parametrů metriky podobnostní (Diceho koeficient) a vzdálenostní (Hausdorffova vzdálenost), které autor statisticky vyhodnotil. Při segmentaci celé míchy, šedé i bílé hmoty byla shoda (DICE) vyvinuté metodiky a manuální segmentace statisticky výrazně vyšší než v případě automatické metodiky a manuální segmentace. Obdobného výsledku bylo dosaženo i při analýze vzdálenostní metriky, kdy maximální chyba mezi porovnávanými metodami byla statisticky významně nižší ve všech tkáních kromě šedé hmoty, u níž rozdíl nebyl statisticky významný. Přehledná ukázka výsledků je na obrázku 9.



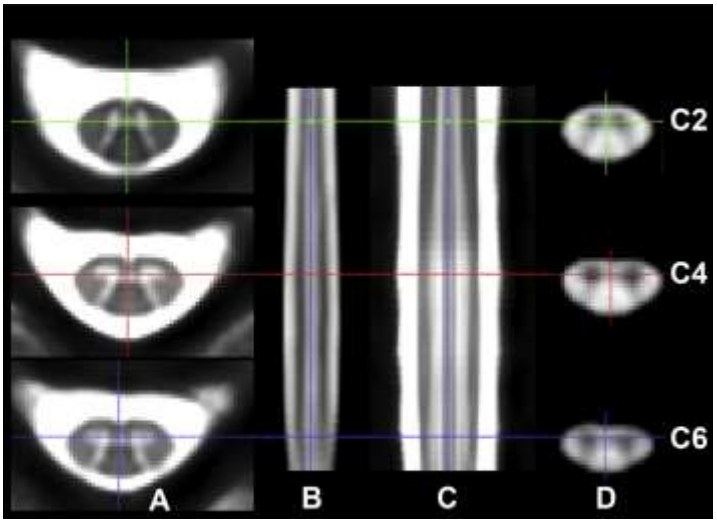

Obr. 9. Porovnání výsledků semi-automatické (CLASS) a automatické (CenterLine) segmentace s manuální segmentací celé krční míchy na anatomických (ESC) a na difuzně vážených (ESC DTI) obrazech, na šedé (GM) a bílé (WM) hmotě. Diceho koeficient (DICE, levý graf) charakterizuje podobnost (vyšší je lepší) a Hausdorffova vzdálenost (pravý graf) charakterizuje maximální neshodu v mm (nižší je lepší). Krabicový graf představuje postupně od spodu minimum, 1. kvartil, medián (čtvereček), 3. kvartil, maximum. Statisticky významné rozdíly jsou označeny ** - $p < 0{,}01$; *** - $p < 0{,}001$.

Výsledné segmentační masky byly registrovány do prostoru difuzně vážených obrazů. Také se zjistilo, jaký vliv má metoda, hodnotitel a tkáň na mediánové hodnoty frakční anizotropie. Statisticky se prokázalo, že jsou rozdíly mezi metodami a tkáněmi, nikoliv mezi hodnotiteli. Jak se dalo očekávat, plně automatizovaná metoda dosáhla v tomto případě nejnižších hodnot koeficientu variace. Prezentovaná metodika měla o něco vyšší hodnoty a nejméně reprodukovatelnou metodou z pohledu hodnot frakční anizotropie byla manuální segmentace.

Statistický přístup je jednou z alternativních analytických metod pro studium difuzních charakteristik bílé hmoty bez nutnosti segmentace šedé a bílé hmoty. Není však možné jej provést na jednotlivci, ale na skupině, která musí mít k dispozici DTI data. Detekce bílé hmoty je založena na prostorové orientaci nervových vláken pomocí tract-based spatial statistics (TBSS). Tato metoda se používá pro analýzu mozku [24], ovšem implementace pro studium krční míchy chybí. Přitom to může být užitečným nástrojem pro skupinové analýzy. Autor tuto metodu úspěšně zavedl v oblasti krční míchy, modifikoval pro segmentaci šedé a bílé hmoty a rovněž verifikoval. Vše je popsáno v mezinárodní publikaci **VI. (Dostál et al. 2020)**. Nejprve bylo třeba vytvořit postup pro sjednocení obrazů všech subjektů, aby výsledky nebyly zatíženy rozdílnou délkou krční míchy v kraniokaudálním směru a aby mohly být registrovány do společného prostoru (obr. 10). Tím byl vytvořen atlas jak anatomický (obr. 11 – A, B, C), tak i difuzní (obr. 11 – D).

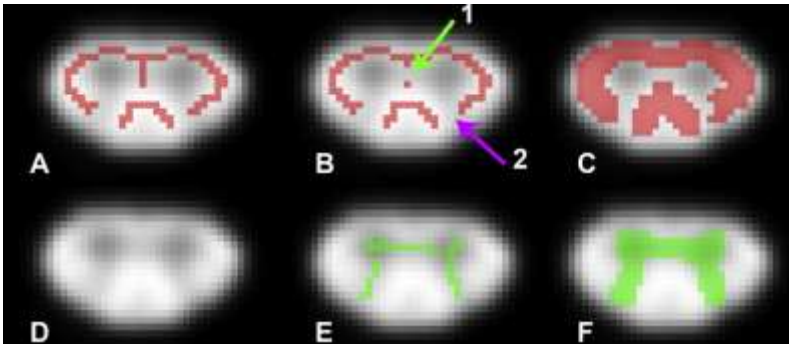



Obr. 10. Ukázka jednotlivých kroků nutných před analýzou. Ořezání originálního obrazu (1) na požadovaný rozsah od disku C1/2 po disk C6/7 (A – B), následováno registrací (2) do společného prostoru v sagitální (C) rovině. Originální (D) a atlasový obraz (E) v axiální rovině.



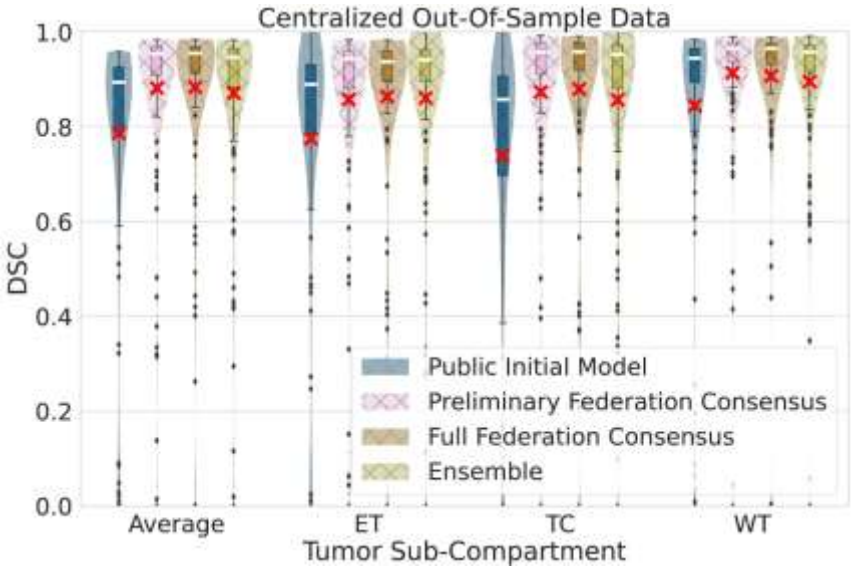

Obr. 11. Ukázka vzniklého anatomického atlasu ve 3 axiálních pozicích (A), dále orientace sagitální (B) a koronální (C). Ukázka difuzního atlasu charakterizovaného skalárním parametrem frakční anizotropie ve 3 axiálních pozicích (D). Barevné přímky odpovídají zobrazeným rovinám.

Jakmile byly takto předzpracovány obrazy všech subjektů, použila se modifikovaná TBSS analýza, jež vytvořila kostru voxelů, které jsou ze statistického pohledu nejpravděpodobněji bílou nebo šedou hmotou napříč všemi subjekty (obr. 12 – A, E). Pro zvětšení počtu voxelů byly tyto kostry prostorově rozšířeny (pomocí dilatace obrazu, obr. 12 – C, F). Ověřili jsme, jaký vliv má toto rozšíření na přesnost segmentace, jaká bude tedy míra nepřesnosti mezi maskou rozšířenou a manuální. Segmentační masky jsou vykresleny na obrazu průměrné frakční anizotropie (obr. 12 – D).




Obr. 12. Ukázka výsledných segmentačních masek bílé (A) a šedé (E) hmoty na obrazech frakční anizotropie (D) před úpravou a po ní (B, C, F).

Jak je uvedeno výše, nejen segmentace šedé a bílé hmoty je nutným krokem před analýzou difuzních dat, ale také segmentace případných patologií (např. tumoru). Lze totiž předpokládat, že nekrotická tkáň bude mít odlišené difuzní charakteristiky, než má edém či sytící se část tumoru [25 – 27]. Strojové učení potřebuje velké množství dat a principiálně mohou být použity libovolné obrazy s danou patologií. Musíme nicméně pamatovat na Obecné nařízení EU o ochraně údajů (GDPR) a poskytování zdravotnické obrazové dokumentace třetím stranám, přestože se jedná o potřeby vědeckého charakteru. K maximalizaci bezpečnosti pacientských dat je možné použít přístup kolaborativního učení (federated learning) [28]. V tomto případě se algoritmus učí z více zdrojů souběžně, přičemž nedochází k transferu samotných dat, ale přenáší se pouze číselné parametry modelu mezi servery centrálním a jednotlivými nemocničními, kde jsou pacientská data bezpečně uložena. Tohoto principu bylo využito při multicentrickém tréninku algoritmu na základech 3D ResUNet pro segmentaci gliomů třídy 3 a 4 (high grade gliomas) na 3 základní tkáně (sytící se část po aplikaci kontrastní látky, edém a nesytící se část, co není edematická). Výstupy byly publikovány v mezinárodním časopise **VII. (Pati et al. 2022)**, ve kterém byla použita data 6314 subjektů ze 71 pracovišť. Takto natrénovaná neuronová síť dosáhla vysoké shody s manuální segmentací, kdy se mediánové hodnoty Diceho koeficientu na testovací skupině (154 subjektů) pohybovaly přes 90 % v závislosti na typu segmentované tkáně (obr. 13).




Obr. 13. Ukázka výsledků značení celého tumoru (WT), edematické tkáně (ET), postkontrastně se sytící části (TC) a průměru všech tkání (Average) pomocí modelu kolaborativního učení dle různých vzorů (barevně odlišeny). Bílá čára nebo červený křížek představují mediánovou nebo průměrnou hodnotu Diceho koeficientu (DSC) (větší je lepší).

## 5. Klinické aplikace DWI při studiu RS

Demyelinizace, gliózy, zánětlivé infiltrace či axonální poškození v různých oblastech CNS jsou charakteristickými znaky roztroušené sklerózy (RS) [29]. Základním radiologickým hodnocením je detekce T2 hyperintenzních ložisek. O iniciálním stádiu demyelinizačního onemocnění mluvíme jako o klinicky izolovaném syndromu (CIS) [30]. Z hlediska predikce vývoje a rozvoje onemocnění do definitivní RS jsou důležité parametry, jako je počet, lokalizace, postkontrastní sycení ložisek a dynamika nálezu v čase, což je shrnuto v McDonaldových kritériích [31]. V minulosti byla pro detekci ložisek RS v mozku zásadní T2 vážená sekvence, ovšem v posledních letech je klíčová T2 vážená sekvence s potlačením signálu volné vody (FLAIR) v 2D či 3D provedení. FLAIR obrazy v oblasti krční míchy nedosahují tak vysoké senzitivity jako v oblasti mozku, proto se zde doporučuje klasická T2 vážená sekvence jak bez saturace tuku, tak i s ní (STIR). Možnosti využití dalších metod jako susceptibilně vážené obrazy (SWI), magnetizační transfer (MT), DTI nebo spektroskopie jsou pojednávány v přehledovém článku autora **VIII. (Keřkovský et al. 2017)**.

Detekce hyperintenzních ložisek mozku a míchy je základní diagnostickou metodou, ale k samotné predikci progrese z CIS do definitivní RS nestačí. Tuto problematiku se autor snažil vyřešit analýzou difuzních charakteristik (DTI) krční míchy u 47 CIS pacientů, kteří byli klinicky i radiologicky sledováni po dobu 2 let, a u 57 neurologicky zdravých dobrovolníků, jež sloužili jako kontrolní skupina. 15 CIS pacientů do 2 let progredovali do definitivní RS. Pomocí semi-automatického algoritmu byla odlišena šedá a bílá hmota míchy, zatímco hyperintenzní ložiska byla značena manuálně. Analýza histogramu

DTI parametrů identifikovala špičatost (kurtozu) frakční anizotropie v oblasti normálně vyhlížející bílé hmoty jako nejsilnější prediktor pro konverzi do definitivní RS s vysokou sensitivitou (93 %) a průměrnou specificitou (73 %). To lze považovat jako zlepšení v porovnání se senzitivitou (80 %) a specificitou (53 %) u McDonaldových kritérií obrazů mozku. Výsledky této práce jsou publikovány v mezinárodním časopise **IX. (Dostál et al. 2021)**.

Při studiu predikčního potenciálu parametrů DTI a měření objemů mozku pro konverzi z CIS do definitivní RS využil autor statistickou metodu TBSS. Během dvouletého sledování 72 CIS pacientů 22 z nich progredovalo do definitivní RS. U pacientů s progresí došlo ke statisticky významnému zmenšení objemu bílé hmoty, což bylo prokázáno jako parametr velmi senzitivní (90,9 %), ale málo specifický (58 %). Mediánová hodnota frakční anizotropie bílé hmoty byla u těchto pacientů také výrazně nižší, avšak tento parametr byl naopak vysoce specifický (90 %) a středně sensitivní (77,3 %). Výsledky studie byly publikovány v mezinárodním časopise **X. (Stulík et al. 2022)**.

Z výše uvedeného je patrné, že kombinace anatomických a difuzně vážených obrazů může vést ke zvýšení predikční schopnosti zobrazování MR v případě předpovědi konverze z klinicky izolovaného syndromu do stádia definitivní roztroušené sklerózy.

# 6. Závěr

V předložené habilitační práci autor nejprve prezentuje základní teoretické principy difuzně váženého zobrazovaní magnetickou rezonancí. Následně navrhuje a ověřuje optimální zobrazovací protokoly krční míchy. Po této akviziční části jsou představeny vědecké aktivity v oblasti zpracování a analýzy anatomických a difuzně vážených obrazů. V závěrečné části je ukázka klinických aplikací v oblasti mozku a krční míchy. Autor podrobněji rozebírá některé ze svých vědeckých publikací, kde demonstruje přijatelnou míru reprodukovatelnosti zobrazení MR, vysoký potenciál této metody pro další vědecké a klinické aplikace, které se v některých situacích jeví přínosněji než standardní diagnostické zobrazování. I když prezentované publikace dosahují dobrých výsledků, ze souhrnného autorova článku z roku 2024 je patrný velký prostor pro další vědeckou práci za použití pokročilejších difuzních modelů. A to je cesta, jakou se autor plánuje dále vědecky ubírat.

# 7. Seznam použité literatury


[1] STEJSKAL, E. O. a J. E. TANNER. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics* [online]. 1965, **42**(1), 288–292. ISSN 0021-9606, 1089-7690. Dostupné z: doi:10.1063/1.1695690

[2] HAMMER, Mark. *Image: MRI Physics: Diffusion-Weighted Imaging - XRayPhysics* [online]. [vid. 2018-04-18]. Dostupné z: https://www.google.com/imgres?imgurl=http%3A%2F%2Fxrayphysics.com%2Fdwi_simple_pulse.png&imgrefurl=http%3A%2F%2Fxrayphysics.com%2Fdwi.html&docid=ZKbnCmFKZ6CERM&tbnid=P5_kSQK-95Yv_M%3A&vet=10ahUKEwjQ5ZyGpsPaAhXLDpoKHfQHBp4QMwg_KAAwAA..i&w=445&h=300&client=firefox-b&bih=1063&biw=1920&q=DWI%20pulse%20schema&ved=0ahUKEwjQ5ZyGpsPaAhXLDpoKHfQHBp4QMwg_KAAwAA&iact=mrc&uact=8

[3] MORI, Susumu a Jiangyang ZHANG. Principles of Diffusion Tensor Imaging and Its Applications to Basic Neuroscience Research. *Neuron* [online]. 2006, **51**(5), 527–539. ISSN 0896-6273. Dostupné z: doi:10.1016/j.neuron.2006.08.012

[4] KINGSLEY, Peter B. Introduction to diffusion tensor imaging mathematics: Part II. Anisotropy, diffusion-weighting factors, and gradient encoding schemes. *Concepts in Magnetic Resonance Part A* [online]. 2006, **28A**(2), 123–154. ISSN 1552-5023. Dostupné z: doi:10.1002/cmr.a.20049

[5] JONES, Derek K. *Diffusion MRI*. B.m.: Oxford University Press, 2010. ISBN 978-0-19-970870-3.

[6] MORITANI, Toshio, Sven EKHOLM a Per-Lennart A. WESTESSON. *Diffusion-Weighted MR Imaging of the Brain*. B.m.: Springer Science & Business Media, 2009. ISBN 978-3-540-78785-3.

[7] OGURA, Akio, Daisuke KOYAMA, Norio HAYASHI, Isamu HATANO, Kohki OSAKABE a Natsumi YAMAGUCHI. Optimal b Values for Generation of Computed High-b-Value DW Images. *American Journal of Roentgenology* [online]. 2016, **206**(4), 713–718. ISSN 0361-803X. Dostupné z: doi:10.2214/AJR.15.14867

[8] MINATI, Ludovico a Władysław P. WĘGLARZ. Physical foundations, models, and methods of diffusion magnetic resonance imaging of the brain: A review. *Concepts in Magnetic Resonance Part A* [online]. 2007, **30A**(5), 278–307. ISSN 1546-6086, 1552-5023. Dostupné z: doi:10.1002/cmr.a.20094

[9] LE BIHAN, Denis. What can we see with IVIM MRI? *NeuroImage* [online]. 2017 [vid. 2018-10-16]. ISSN 1053-8119. Dostupné z: doi:10.1016/j.neuroimage.2017.12.062

[10] BLASCHE, Mathias a Daniel FISCHER. Magnet Homogeneity and Shimming. 2017.

[11] PIERPAOLI, Carlo. Artifacts in Diffusion MRI. In: PhD JONES Derek K., ed. *Diffusion MRI: Theory, Methods, and Applications* [online]. B.m.: Oxford University Press, 2010 [vid. 2024-07-19], s. 0. ISBN 978-0-19-536977-9. Dostupné z: doi:10.1093/med/9780195369779.003.0018

[12] IRFANOGLU, M. Okan, Joelle SARLLS, Amritha NAYAK a Carlo PIERPAOLI. Evaluating corrections for Eddy-currents and other EPI distortions in diffusion MRI: methodology and a

dataset for benchmarking. *Magnetic Resonance in Medicine* [online]. 2019, **81**(4), 2774–2787. ISSN 1522-2594. Dostupné z: doi:10.1002/mrm.27577

[13] HOLDSWORTH, S. J., K. YEOM, S. SKARE, A. J. GENTLES, P. D. BARNES a R. BAMMER. Clinical Application of Readout-Segmented− Echo-Planar Imaging for Diffusion-Weighted Imaging in Pediatric Brain. *American Journal of Neuroradiology* [online]. 2011, **32**(7), 1274–1279. ISSN 0195-6108, 1936-959X. Dostupné z: doi:10.3174/ajnr.A2481

[14] SARITAS, Emine U., Samantha J. HOLDSWORTH a Roland BAMMER. Chapter 2.3 - Susceptibility Artifacts. In: Julien COHEN-ADAD a Claudia A.M. WHEELER-KINGSHOTT, ed. *Quantitative MRI of the Spinal Cord* [online]. San Diego: Academic Press, 2014, s. 91–105. ISBN 978-0-12-396973-6. Dostupné z: doi:10.1016/B978-0-12-396973-6.00007-1

[15] SORACE, Anna G., Chengyue WU, Stephanie L. BARNES, Angela M. JARRETT, Sarah AVERY, Debra PATT, Boone GOODGAME, Jeffery J. LUCI, Hakmook KANG, Richard G. ABRAMSON, Thomas E. YANKEELOV a John VIROSTKO. Repeatability, reproducibility, and accuracy of quantitative mri of the breast in the community radiology setting. *Journal of Magnetic Resonance Imaging* [online]. 2018, **48**(3), 695–707. ISSN 1522-2586. Dostupné z: doi:10.1002/jmri.26011

[16] MCGUIRE, Stephen A., S. Andrea WIJTENBURG, Paul M. SHERMAN, Laura M. ROWLAND, Meghann RYAN, John H. SLADKY a Peter V. KOCHUNOV. Reproducibility of quantitative structural and physiological MRI measurements. *Brain and Behavior* [online]. 2017, **7**(9), e00759. ISSN 2162-3279. Dostupné z: doi:10.1002/brb3.759

[17] WANG, Y., S. TADIMALLA, R. RAI, J. GOODWIN, S. FOSTER, G. LINEY, L. HOLLOWAY a A. HAWORTH. Quantitative MRI: Defining repeatability, reproducibility and accuracy for prostate cancer imaging biomarker development. *Magnetic Resonance Imaging* [online]. 2021, **77**, 169–179. ISSN 1873-5894. Dostupné z: doi:10.1016/j.mri.2020.12.018

[18] MAHMOUD, Amira A., El-Sayed M. EL-RABAIE, Taha E. TAHA, Adel ELFISHAWY, Osama ZAHRAN a Fathi E. ABD EL-SAMIE. Medical Image Segmentation Techniques, a Literature Review, and Some Novel Trends. *Menoufia Journal of Electronic Engineering Research* [online]. 2018, **27**(2), 23–58. ISSN 1687-1189. Dostupné z: doi:10.21608/mjeer.2018.63179

[19] FAWZI, Ali, Anusha ACHUTHAN a Bahari BELATON. Brain Image Segmentation in Recent Years: A Narrative Review. *Brain Sciences* [online]. 2021, **11**(8), 1055. ISSN 2076-3425. Dostupné z: doi:10.3390/brainsci11081055

[20] EJAZ, Khurram, Mohd Shafry Mohd RAHIM, Muhammad ARIF, Diana IZDRUI, Daniela Maria CRACIUN a Oana GEMAN. Review on Hybrid Segmentation Methods for Identification of Brain Tumor in MRI. *Contrast Media & Molecular Imaging* [online]. 2022, **2022** [vid. 2024-07-19]. Dostupné z: doi:10.1155/2022/1541980

[21] GARG, Sheetal a S. R. BHAGYASHREE. Spinal Cord MRI Segmentation Techniques and Algorithms: A Survey. *SN Computer Science* [online]. 2021, **2**(3), 229. ISSN 2661-8907. Dostupné z: doi:10.1007/s42979-021-00618-4

[22] RUGGIERI, Serena, Maria PETRACCA, Laura DE GIGLIO, Francesca DE LUCA, Costanza GIANNÌ, Flavia GURRERI, Nikolaos PETSAS, Silvia TOMMASIN, Carlo POZZILLI a Patrizia PANTANO. A matter of atrophy: differential impact of brain and spine damage on disability worsening in multiple sclerosis. *Journal of Neurology* [online]. 2021, **268**(12), 4698–4706. ISSN 0340-5354. Dostupné z: doi:10.1007/s00415-021-10576-9

[23] WENDEBOURG, Maria Janina, Matthias WEIGEL, Claudia WEIDENSTEINER, Laura SANDER, Eva KESENHEIMER, Nicole NAUMANN, Tanja HAAS, Philipp MADOERIN, Nathalie BRAUN, Christoph NEUWIRTH, Markus WEBER, Kathleen JAHN, Ludwig KAPPOS, Cristina GRANZIERA, Kathi SCHWEIKERT, Michael SINNREICH, Oliver BIERI a Regina SCHLAEGER. Cervical and thoracic spinal cord gray matter atrophy is associated with disability in patients with amyotrophic lateral sclerosis. *European Journal of Neurology* [online]. 2024, **31**(6), e16268. ISSN 1468-1331. Dostupné z: doi:10.1111/ene.16268

[24] SMITH, Stephen M., Mark JENKINSON, Heidi JOHANSEN-BERG, Daniel RUECKERT, Thomas E. NICHOLS, Clare E. MACKAY, Kate E. WATKINS, Olga CICCARELLI, M. Zaheer CADER, Paul M. MATTHEWS a Timothy EJ BEHRENS. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage; Amsterdam* [online]. 2006, **31**(4), 1487–1505. ISSN 10538119. Dostupné z: doi:http://dx.doi.org/10.1016/j.neuroimage.2006.02.024

[25] VAN SANTWIJK, Lusien, Valentina KOUWENBERG, Frederick MEIJER, Marion SMITS a Dylan HENSSEN. A systematic review and meta-analysis on the differentiation of glioma grade and mutational status by use of perfusion-based magnetic resonance imaging. *Insights into Imaging* [online]. 2022, **13**(1), 102. ISSN 1869-4101. Dostupné z: doi:10.1186/s13244-022-01230-7

[26] MOMENI, Farideh, Razzagh ABEDI-FIROUZJAH, Zahra FARSHIDFAR, Nastaran TALEINEZHAD, Leila ANSARI, Ali RAZMKON, Amin BANAEI a Alireza MEHDIZADEH. Differentiating Between Low- and High-grade Glioma Tumors Measuring Apparent Diffusion Coefficient Values in Various Regions of the Brain. *Oman Medical Journal* [online]. 2021, **36**(2), e251. ISSN 1999-768X. Dostupné z: doi:10.5001/omj.2021.59

[27] CSUTAK, Csaba, Paul-Andrei ŞTEFAN, Lavinia Manuela LENGHEL, Cezar Octavian MOROŞANU, Roxana-Adelina LUPEAN, Larisa ŞIMONCA, Carmen Mihaela MIHU a Andrei LEBOVICI. Differentiating High-Grade Gliomas from Brain Metastases at Magnetic Resonance: The Role of Texture Analysis of the Peritumoral Zone. *Brain Sciences* [online]. 2020, **10**(9), 638. ISSN 2076-3425. Dostupné z: doi:10.3390/brainsci10090638

[28] REHMAN, Muhammad Habib ur, Walter HUGO LOPEZ PINAYA, Parashkev NACHEV, James T. TEO, Sebastin OURSELIN a M. Jorge CARDOSO. Federated learning for medical imaging radiology. *The British Journal of Radiology* [online]. 2023, **96**(1150), 20220890. ISSN 0007-1285. Dostupné z: doi:10.1259/bjr.20220890

[29] CONFAVREUX, Christian, Sandra VUKUSIC, Thibault MOREAU a Patrice ADELEINE. Relapses and progression of disability in multiple sclerosis. *New England Journal of Medicine* [online]. 2000, **343**(20), 1430–1438. ISSN 0028-4793. Dostupné z: doi:10.1056/NEJM200011163432001

[30] MILLER, David H., Declan T. CHARD a Olga CICCARELLI. Clinically isolated syndromes. *The Lancet. Neurology* [online]. 2012, **11**(2), 157–169. ISSN 1474-4465. Dostupné z: doi:10.1016/S1474-4422(11)70274-5

[31] THOMPSON, Alan J., Brenda L. BANWELL, Frederik BARKHOF, William M. CARROLL, Timothy COETZEE, Giancarlo COMI, Jorge CORREALE, Franz FAZEKAS, Massimo FILIPPI, Mark S. FREEDMAN, Kazuo FUJIHARA, Steven L. GALETTA, Hans Peter HARTUNG, Ludwig KAPPOS, Fred D. LUBLIN, Ruth Ann MARRIE, Aaron E. MILLER, David H. MILLER, Xavier MONTALBAN, Ellen M. MOWRY, Per Soelberg SORENSEN, Mar TINTORÉ, Anthony L. TRABOULSEE, Maria TROJANO, Bernard M. J. UITDEHAAG, Sandra VUKUSIC, Emmanuelle WAUBANT, Brian G.

WEINSHENKER, Stephen C. REINGOLD a Jeffrey A. COHEN. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet. Neurology* [online]. 2018, **17**(2), 162–173. ISSN 1474-4465. Dostupné z: doi:10.1016/S1474-4422(17)30470-2

# 8. Abstrakt

Optimization of acquisition and analysis methods for magnetic resonance images

Diffusion-weighted brain imaging is nowadays a standard examination method that can detect, differentiate or predict the current state of some disabilities better than methods based on anatomical image analysis. With the development of hardware equipment of the MR systems, the possibilities of using advanced diffusion models such as diffusion tensor imaging (DTI), diffusion kurtosis imaging (DKI) or intravoxel incoherent motion (IVIM) are expanding. These advanced models have great scientific and clinical potential, but also have their technical challenges to overcome, especially in the cervical spinal cord.

In this paper, the author first presents the technical principles and challenges of diffusion-weighted imaging, followed by the optimization of the examination protocol in the cervical spinal cord. In the next section, some aspects of image analysis, especially image segmentation, are presented. In the last part, the author focuses on the clinical applications of diffusion-weighted images in patients with CNS involvement.

This work presents the great potential of advanced diffusion-weighted imaging models for both scientific and clinical purposes, which, however, needs to be thoroughly verified.

Optimalizace akvizice a metod analýzy obrazů magnetické rezonance

Difuzně vážené zobrazování mozku je v dnešní době standardní vyšetřovací metodou, která může detekovat, diferenciovat nebo predikovat aktuální stav postižení lépe než metody založené na analýze anatomických obrazů. S rozvojem hardwarového vybavení strojů MR se rozšiřují i možnosti využití pokročilých difuzních modelů, jakými je zobrazování tenzoru difuze (DTI), kurtozy tenzoru difuze (DKI) nebo intravoxel incoherent motion (IVIM). Tyto pokročilé modely mají velký vědecký i klinický potenciál, ovšem také své technické výzvy, jež je třeba ještě překonat, a to hlavně v oblasti krční míchy.

Autor v této práci nejprve představuje technické principy a výzvy difuzně váženého zobrazování, následované optimalizací vyšetřovacího protokolu v oblasti krční míchy. V další části jsou představeny některé aspekty analýzy obrazů, především segmentace obrazu. V poslední části se autor zaměřuje na klinické aplikace difuzně vážených obrazů u pacientů s postižením CNS.

Tato práce prezentuje velký potenciál pokročilých modelů difuzně váženého zobrazovaní jak pro vědecké, tak pro klinické účely, které je ovšem nutno důkladně verifikovat.

# 9. Původní práce autora

Special Review

# Possibilities of Using Multi-b-value Diffusion Magnetic Resonance Imaging for Classification of Brain Lesions

Tereza Kopřivová, Miloš Keřkovský, Tomáš Jůza, Václav Vybíhal, Tomáš Rohan, Michal Kozubek, Marek Dostál

In contrast to conventional diffusion magnetic resonance imaging (MRI), multi-b-value diffusion MRI methods are able to separate the signal from free water, pseudo-diffusion, and non-Gaussian components of water molecule diffusion. These approaches can then be utilised in so-called intravoxel incoherent motion imaging and diffusion kurtosis imaging. Various parameters provided by these methods can describe additional characteristics of the tissue microstructure and potentially help in the diagnosis and classification of various pathological processes. In this review, we present the basic principles and methods of analysing multi-b-value diffusion imaging data and specifically focus on the known possibilities for its use in the diagnosis of brain lesions. We also suggest possible directions for further research.

**Key Words:** DWI; IVIM; DKI; Brain tumours.



## INTRODUCTION

Classification of brain lesions using conventional magnetic resonance imaging (MRI) is often difficult because many lesions have non-specific patterns and numerous common characteristics. Nevertheless, their accurate differential diagnosis is important for determining correct therapeutic approaches.

Diffusion-weighted imaging (DWI) is an MRI technique that allows for visualising and quantifying the process of water molecule diffusion in tissue. This is highly beneficial in diagnosing a number of pathological conditions, and, in what may be termed its "basic" version founded on use of the Gaussian diffusivity model, it already is a widely accepted part of conventional brain MRI protocols. Recently, however, new approaches have emerged for acquiring and analysing diffusion data using the intravoxel incoherent motion (IVIM) model or diffusion kurtosis imaging (DKI). Because both methods use measurements with many different b-values, they can be referred to collectively as "multi-b-value diffusion imaging". These methods can provide a number of quantitative parameters that, alone or in combination, may have the potential to improve the diagnostic accuracy of MRI and are therefore being investigated in many applications, including non-invasive classification of brain lesions. This review aims to present the basic principles and methods of analysis of multi-b-value diffusion imaging data while focusing on the known possibilities for their use in the diagnosis of brain lesions and suggesting possible directions for further research.

**Acad Radiol 2024; 31:261–272**

From the Department of Radiology and Nuclear Medicine, Faculty of Medicine, Masaryk University, Brno and University Hospital Brno, Jihlavská 20, 625 00, Brno, Czech Republic (T.K., M.K., T.J., T.R., M.D.); Department of Neurosurgery, Faculty of Medicine, Masaryk University Brno and University Hospital Brno, 625 00, Brno, Czech Republic (V.V.); Centre for Biomedical Image Analysis, Faculty of Informatics, Masaryk University, Šumavská, Brno, Czech Republic (M.K.); Department of Biophysics, Faculty of Medicine, Masaryk University, Brno, Czech Republic (T.J., M.D.). Received June 13, 2023; revised September 25, 2023; accepted October 3, 2023. **Address correspondence to:** M.K. e-mail: kerkovsky.milos@fnbrno.cz


https://doi.org/10.1016/j.acra.2023.10.002

## PRINCIPLES OF MULTI-B-VALUE DIFFUSION IMAGING

MRI can quantify the diffusive behaviour of water molecules in vivo. Diffusion is a random translational motion of molecules that originates from their thermal energy. In what may be termed an "unrestricted" environment (e.g. water), the mean path between two collisions of water molecules can be determined. This is on the order of tens of nanometres, and the

molecules' mean velocity can be considered to be around 100 m/s. We can also assume a Gaussian distribution of both quantities. Based on Einstein's equation, we can calculate that the diffusion coefficient in water is around $10^{-9}$ $m^2/s$.

In living tissue, however, water molecules not only diffuse but also flow in the bloodstream. The latter, too, can be seen as a sort of diffusion, which is so-called pseudo-diffusion. Again in accordance with Einstein's equation, we could determine the mean segment length and the mean blood flow velocity and then calculate the pseudo-diffusion coefficient, which is approximately ten times larger ($10^{-8}$ $m^2/s$) than in the case of free water diffusion. The "free" water model is also not entirely realistic in tissues from a DWI perspective because the movement of water molecules is greatly limited by the presence of cell membranes and other structures that interact with them. This results in a departure from the Gaussian statistical distribution and necessitates a change in the model used. Because all three of these effects are usually present within a small region of space (voxel), we can speak of incoherent motion of water molecules within the imaged voxel, or intravoxel incoherent motion (IVIM).(1,2).

DWI pulse sequence uses additional directionally selective gradient pulses whose amplitude and duration can be characterised by the so-called b-value. This value affects the resulting contrast and signal of the image, inasmuch as the magnetic resonance (MR) signal decreases exponentially with increasing b-value.

In clinical practice, the simplest diffusion model, the so-called monoexponential model, is standardly used. In this case, a DWI image is acquired without diffusion weighting ($b = 0$ $s/mm^2$) and with diffusion weighting with a single b-value (typically around 1000 $s/mm^2$) and then the apparent diffusion coefficient is calculated as in Eq. 1.

$$S_b = S_{b0}e^{-b\star ADC}, \tag{1}$$

where $S_{b0}$ is the signal without the application of diffuse gradients, $S_b$ is the signal with the application of diffuse gradients, and $b$ is the diffusion weighting value.

This model is greatly simplified because it cannot distinguish between the pseudo-diffusion component and the non-Gaussian diffusion component, hence the term apparent diffusion coefficient (ADC) is used. Nevertheless, diffusion-weighted images are often very useful in clinical practice and, because of their simplicity (both in terms of data acquisition and calculation), they have become an integral part of diagnostic MR protocols.(3–5) If we want to capture the non-Gaussian behaviour of water diffusion, diffusion-weighting of the images needs to be enhanced. The exact threshold is difficult to determine, but b-values of 2000 $s/mm^2$ or greater are generally considered sufficient. The kurtosis model (Eq. 2) can then be used, which mathematically extends the classical monoexponential relationship (Eq. 1) by a quadratic term:

$$S_b = S_{b0}e^{(-bD+(bD)^2K/6)}, \tag{2}$$

where $D$ is the diffusion coefficient without the non-Gaussian component and $K$ is the kurtosis coefficient expressing the degree of deviation from the Gaussian model.

The application of larger diffusion-weighting, however, significantly reduces the amount of signal in the image, so image averaging must be used, and that leads to longer acquisition time.(6,7) Also, the hardware of MR machines must be adapted (in terms of switching speed and maximum amplitude of the gradient coils) for imaging with such high b-values. Last but not least, mathematical analysis of the signal and interpretation of the results are thereby made more complex.

Because pseudo-diffusion has a diffusion coefficient approximately 10 times larger than that of "free" water, to distinguish its contributions to the total signal, the behaviour at low diffusion weighting ($b < 250$ $mm^2/s$) must be observed. The monoexponential model (Eq. 1) can then be extended to include the pseudo-diffusion component and its fraction:

$$S_b = S_{b0}(fe^{-b\star D^*} + (1-f)e^{-b\star D}), \tag{3}$$

where $f$ is the perfusion fraction coefficient, $D^*$ the pseudo-diffusion coefficient, and $D$ the diffusion coefficient without the pseudo-diffusion component. Thus, there is no need to speak of ADC anymore, but the coefficient $D$ can be seen as the diffusion of water in tissue.

For more accurate separation of individual signals, a larger number of different and lower b-values must be applied.(8,9) An example showing specific parameters of a multi-b-value diffusion imaging sequence used at our department is presented in Table 1. The parameters might represent a compromise between robustness of the data implying the accuracy of subsequent estimations of diffusion parameters and clinically acceptable acquisition time (4:45 min), similarly to as reported by Guo and Jiang.(10) Several parameters of the sequence might be modified (e.g. by adding more acquisitions with different b factors or enhancing the averaging to obtain better signal-to-noise ratios), which would of course impact also on the scanning time. Eq. 4 can be used to roughly estimate the length of the SS-EPI acquisition:

$$t[s] = TR[s]\star dir\star pack\star \sum_b NSA_b, \tag{4}$$

where $TR$ is repletion time in seconds, $dir$ is number of directions (in our case 3 orthogonal), $pack$ is number of packages or acquisition or concatenations (depending on the manufacturer, in our case 1), and the last parameter is the sum of NSA over all b-values (in our case 21). The number of layers does not appear in the equation because this parameter affects the minimum value of TR. Therefore, changing the number of layers will change the value of TR (in the case of a constant TR, the number of packages may change) and thus change the acquisition time.

Although the IVIM technique cannot be perceived as a direct representation of brain tissue perfusion, this method, due to the aforementioned effect of capillary blood flow on

TABLE 1. Sample Parameters of the Multi-b-value DWI Sequence Used on a 1.5T Magnetic Induction Machine at Our Department

| Parameter | Value | Parameter | Value | Parameter | Value |
|---|---|---|---|---|---|
| FOV | 23 * 23 cm | acq.pix. | 1.5* 1.5 mm | Matrix | 154 * 154 |
| Slice thickness | 3 mm | Gap | 1 mm | Number of slices | 27 |
| TR | 4500 ms | TE | 99 ms | FatSat | SPIR |
| WFS | 16.6 pix | SENSE | 2 | Acquisition time | 4:45 |
| b-values (NSA) | | 0 (1), 10 (1), 20(1), 30(1), 50(1), 100 (2), 200 (2), 500 (3), 1000 (3), 2000 (6) | | | |

FOV, field of view; acq.pix, acquisition pixel size; FatSat, type of fat signal suppression method used; NSA, number of signal averages; SENSE, type of parallel acquisition technique used; SPIR, spectral pre-saturation with inversion recovery; TE, echo time; TR, repetition time; WFS, fat and water frequency shift.

the measured diffusivity values, does to some extent reflect the state of tissue perfusion. Therefore, the method has potential to be applied in the study of focal brain lesions, as discussed in the following, as well as in the exploration of diffuse pathologic processes affecting brain perfusion, such as ALS (11), dementias, (12,13), and others. It has been shown that the various IVIM parameters can be related to conventional perfusion imaging (dynamic susceptibility contrast, DSC) parameters such as blood volume (BV), mean transit time (MTT), and blood flow (BF) using relationships shown in Eqs. 5–7: (14).

$$f_{IVIM} = \frac{BV}{f_v}, \tag{5}$$

$$MTT = \frac{Ll}{6D^*}, \tag{6}$$

$$BF = f_{IVIM} D^* \frac{6f_v}{Ll}, \tag{7}$$

where $f_v$ is the fraction of MR visible water, $L$ is total length of the capillary bed, and $l$ is mean length of the capillary segment.

All three models mentioned can be clearly displayed in a single plot (Fig. 1), where the pseudo-diffusion effect is captured in its left part (b ~ 100 s/mm$^2$), the middle part can be considered as the free water diffusion region (b ~ 1000 s/mm$^2$), and the right part captures the departure from the Gaussian characteristic at high b factors (b ~ 2000 s/mm$^2$).

In scientific fields, we often encounter yet another use of DWI, namely monitoring the anisotropic behaviour of the diffusivity using diffusion tensor imaging (DTI) (15) or the anisotropic behaviour of the non-Gaussian component of the signal in the form of kurtosis tensor imaging (DKI) (16). In contrast to the aforementioned approaches, both of these methods exploit the possibility for applying gradient pulses in different (usually more than 16) directions to investigate the spatial behaviour of diffusion or kurtosis, given the same conditions on the choice of b-values. Although these measurements are rather time-consuming, they provide unique insight into the tissue under investigation that cannot be obtained by any other in vivo method. The outputs of the DTI or DKI analysis can be, among other parameters, the mean values of scalar quantities describing the corresponding mean diffusion (MD) and mean kurtosis (MK) tensors. These quantities describe primarily not anisotropy but what can be termed the "average spatial" value of the parameter. Therefore, these are roughly comparable to the $D$ and $K$ parameters from the aforementioned multi-b-value analyses, which also describe not anisotropic behaviour but only the average diffusivity or kurtosis of the tissue.

It has been reported that the MK parameter is associated with histological complexity of the examined tissue, with structurally complex tissues showing higher MK values compared to less complex tissues. This fact can be utilised in characterisation and differential diagnosis of focal brain lesions,(17) as well as in detection of diffuse affections which, unlike focal lesions, cannot be detected on conventional MRI images.(18).

Previously, it is clear that, if we extend conventional DWI imaging to multiple measurements with a larger number and range of b-factor values, we can obtain comprehensive data in a clinically acceptable timeframe that, with the help of more advanced analyses, provide additional parameters beyond the apparent diffusivity measure that may have the potential to more closely characterise tissues' structural changes.

## APPLICATION OF MULTI-B-VALUE DIFFUSION IMAGING IN CLASSIFICATION OF BRAIN LESIONS

### Differentiation of Tumour and Non-tumour Lesions

Considering the obvious implications for further treatment, probably the most important task of MRI diagnostics is to try and distinguish tumorous from non-tumorous lesions.

Especially differential diagnosis of the so-called ring-shaped lesions poses a specific problem. These may have both benign (i.e. non-tumorous) aetiology, represented, for example, by brain abscesses or tumefactive demyelinating lesions (TDLs), and malignant lesions, represented for example by metastases or high-grade gliomas.(19).

The use of conventional DWI in the diagnosis of brain abscesses is already well known and established in routine clinical practice. It has been shown that this method can

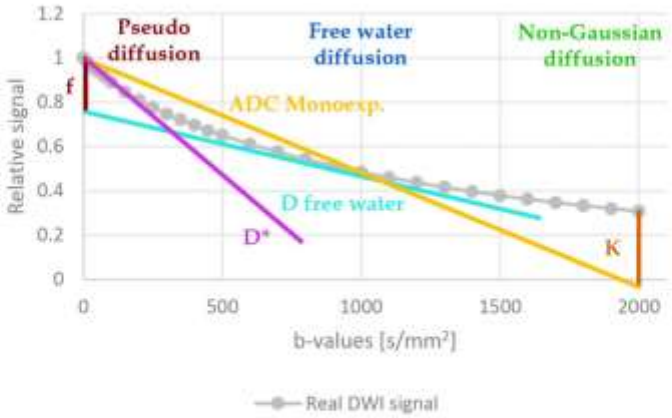



**Figure 1.** Illustration of the dependence of DWI signal on b-value in tissue and graphical representation of the parameters of each model. ADC based on the classical monoexponential model (Eq. 1, calculation from two b-values, namely 0 and 1000 s/mm$^2$, yellow), f, D*, D based on free water using the IVIM model (Eq. 3), and K based on the non-Gaussian model (Eq. 2). ADC, apparent diffusion coefficient; DWI, diffusion-weighted imaging.

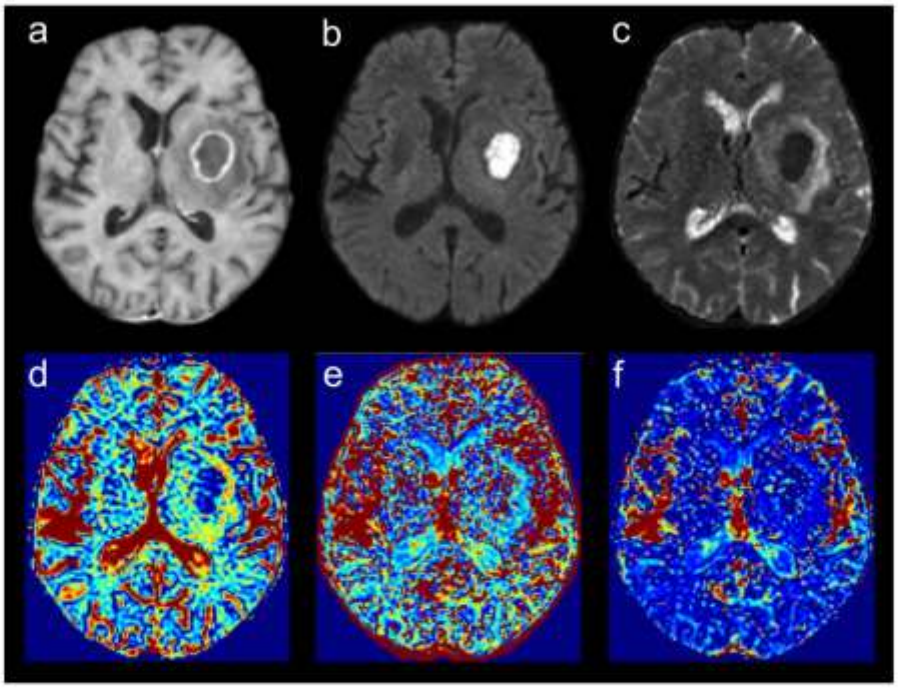



**Figure 2.** Brain abscess within left basal ganglia with enhancing rim visible on post-contrast T1-weighted image in axial plane (**a**). Various diffusion images and maps are shown: b1000 (**b**), D (**c**), f (**d**), D* (**e**), and f*D* (**f**). Mean D* values measured within the segmentation mask covering the entire enhancing rim of the lesion were lower (13253 *10$^{-6}$ mm$^2$/s) compared to those measured in the region of interest situated in contralateral white matter (16561 *10$^{-6}$ mm$^2$/s). No hypervascularization is visible on the f*D* map (**f**). The centre of the abscess demonstrates markedly restricted diffusion with low D values (**c**) and high signal intensity on the b1000 image (**b**).

distinguish brain abscesses from malignant lesions with high sensitivity and specificity based on diffusion restriction and the resulting low ADC values. This restriction of diffusion is attributed to the presence of cells, necrotic detritus, and macromolecules in the pus and its resulting high viscosity.(20) For better comparability with data from the literature, multi-b-value diffusion imaging data can be used to calculate ADC maps according to a simple monoexponential model that is analogous to conventional DWI imaging. The relevance of analysing the non-Gaussian diffusivity characteristics in this matter is presently unclear due to the lack of literature data. Using conventional perfusion imaging techniques like dynamic susceptibility contrast (DSC), however, reduced blood supply to the contrast-enhancing rim of abscesses has been previously demonstrated in comparison to malignant brain tumours.(21,22) Considering the relationship of some IVIM imaging parameters to brain tissue perfusion,(23) it can be assumed that this method could be helpful in addressing this problem while also offering the general advantage that dynamic contrast agent application is not necessary as it is for conventional perfusion imaging by DSC technique (Fig 2).

The differentiation of TDLs poses a greater diagnostic challenge wherein diffusion MRI may play a role. Acute demyelinating plaques are known to show signs of diffusivity restriction that gradually disappear over time.(24) In cases of TDL, diffusivity restriction has been demonstrated with a decrease in ADC values especially in the periphery of the lesions,(25) and some ability to differentiate TDLs from lymphomas and high-grade gliomas using conventional DWI through quantification of ADC values has been reported.(26) Differentiation of TDLs from tumours while using DWI imaging alone is not entirely reliable, however, and a combination of different characteristics in multiparametric imaging, including morphological features, may be helpful. The literature also includes papers using perfusion MRI (DSC) imaging to differentiate TDLs from tumours, and, in one earlier paper, reduced relative cerebral blood volume (rCVB) values were measured in TDLs compared to brain tumours.(27) In view of the previous results, we also see some potential for the use of additional IVIM imaging parameters that reflect to some extent, too, the degree of blood supply to the tissues, but relevant data on this topic are not yet available. The use of multi-b-value diffusion imaging in this application is therefore one of the desirable directions for future research.

Cerebral ischaemia, a condition affecting both diffusion and perfusion of brain tissue, is a logical research direction in the application of IVIM. Differentiation of acute or subacute cerebral ischaemia from other brain lesions is usually a simple task due to the combination of clinical picture and imaging, namely findings on conventional DWI, the role of which has been well documented.(28) Thus, the IVIM technique is being investigated as a possible alternative or complement to existing established imaging protocols,(29,30) as it has been demonstrated that perfusion-related parameters derived from IVIM technique in patients with stroke show excellent agreement with perfusion parameters derived from conventional perfusion imaging while simultaneously providing information on tissue diffusion.(31,32).

### Differential Diagnosis of Tumours

One of the major problems in the differential diagnosis of malignant brain lesions is the differentiation of primary central nervous system (CNS) lymphoma (PCNSL) from high-grade glioma (HGG), as the subsequent therapeutic approach to these lesions differs fundamentally.

Lymphoma has been shown to have significantly lower ADC values on DWI imaging in contrast to HGG, which is attributed to its high cellularity. A recent meta-analysis reported overall sensitivity and specificity of conventional diffusion MRI for differentiating PCNSL from HGG of 82% and 87%, respectively.(33) Studies investigating the use of perfusion imaging have shown significantly lower perfusion parameters in PCNSL, which probably results from the fact that PCNSL does not cause such significant neovascularization as does HGG.(34) Considering the relatively specific characteristics of lymphoma on conventional DWI and perfusion imaging, it can be assumed that the IVIM method might offer opportunities to further refine the differential diagnosis. In correlation with these assumptions, the parameter $f$ (perfusion fraction) has been found in studies published to date to be significantly lower in PCNSL than in HGG.(35) The number of studies investigating IVIM in this application is still limited, however, and further research in this area could be well justified.

Similarly, few studies using the DKI technique to differentiate lymphoma from HGG can be found. For example, the use of DKI imaging with multidirectional diffusion gradients has been reported to quantify axial and radial kurtosis parameters based on anisotropy analysis of this parameter, as well as of MK, a parameter somewhat analogous to the $K$ parameter that can be calculated from multi-b-value measurements of isotropic diffusivity as described previously. The MK and axial kurtosis values thus obtained were significantly higher for PCNSL compared to those for HGG.(5) Similar conclusions were reached in another recent work, which also demonstrated a significantly higher MK value in PCNSL.(36).

The results of the multi-b-value diffusion imaging analysis in a patient with HGG and lymphoma are shown in Figure 3.

Another problem of differential diagnostics consists in the differentiation of solitary brain metastases from HGG. In this case, the situation is complicated by the fact that metastases represent a highly heterogeneous group depending also on the characteristics of the primary tumour. At the same time, however, it is clear that distinguishing metastasis from HGG is important for the choice of further treatment strategy. Morphological imaging alone has relatively low sensitivity and specificity in differentiating between these lesions; one published paper reports accuracy of 68%, sensitivity of 84%,

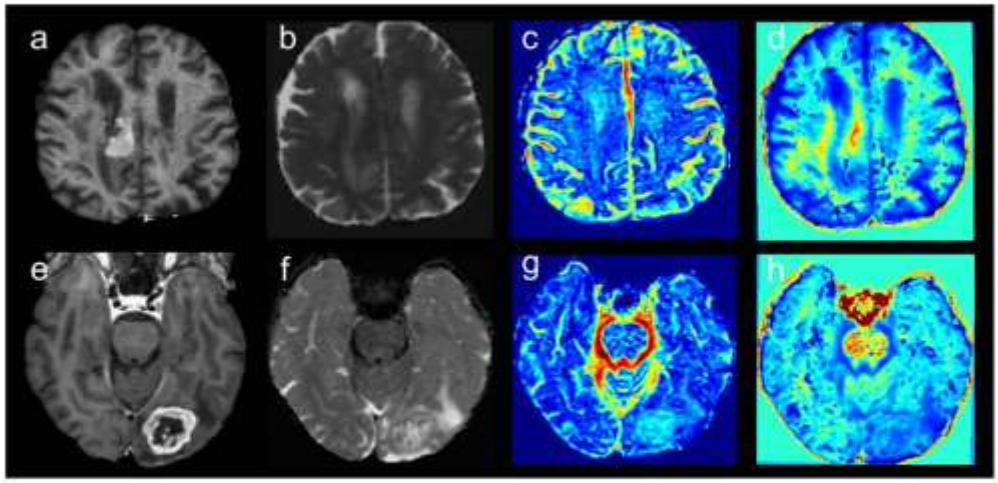



**Figure 3.** Comparison of multi-b diffusion findings in patient with primary CNS lymphoma (diffuse large B-cell lymphoma) (**a–d**) and glioblastoma (IDH wild type) (**e–h**). The two tumours demonstrate different patterns of contrast enhancement on T1-w images (**a–e**). Corresponding slices of several diffusion maps are shown: D (**b, f**), f (**c, g**), and K (**d, h**). Lymphoma lesion with typical homogeneous contrast enhancement demonstrates low D ($545 * 10^{-6}$ $mm^2/s$) and f (0.175) values (**b** and **c**, respectively) and comparatively high K (1.152) values (**d**). In contrast, glioblastoma in the left occipital lobe has comparatively higher D ($825 * 10^{-6}$ $mm^2/s$) and higher f (0.179) values (**f** and **g**, respectively) and lower K (0.844) values compared to lymphoma. The given values were calculated as means measured within the segmented enhancing parts of both tumours. IDH, isocitrate dehydrogenase.

and specificity of 45%.(37) For this reason, a number of studies have recently investigated the use of various advanced diffusion imaging techniques to differentiate between these two groups of brain tumours. This recent research has focused on the peritumoral region, which appears as a T2 hyperintense rim around the contrast-enhancing lesion and may be underlain by a relatively wide spectrum of such histopathological tissue abnormalities as vasogenic oedema, peritumoral infiltration, and the contribution of inflammation, necrosis, or gliosis.(38) Studies focusing on this area are based on the assumption that in HGG the peritumoral zone contains also tumour cell infiltration, whereas metastases are surrounded predominantly by vasogenic oedema, and it is believed that advanced imaging methods can discover and quantify this difference.(39) This hypothesis is supported by a number of studies investigating diffusion imaging (conventional DWI and DTI), which have found significantly lower ADC or MD measurements within peritumoral oedema in HGG compared to metastases. According to a meta-analysis, however, the results of these studies are not entirely uniform and they report rather widely ranging sensitivity and specificity values (46–96% and 40–100%, respectively).(40).

Exploring the benefits of IVIM in this application is still in its early stages, but the first studies show the method's potential. For example, the pseudo-diffusion coefficient $D^*$ values in the peritumoral region, which in one recent study were significantly higher in HGGs than in metastases, and the $f$ values in the peritumoral region, which were significantly lower in HGGs, seem promising for differentiating between those aetiologies.(41) A similar pattern of the diffusivity abnormalities within the peritumoral region is shown in Figure 4.

Tan et al. also demonstrated increased MK values within the peritumoral region in HGGs compared to metastases and they documented overall greater accuracy of DKI imaging in this situation compared to that of DTI.(42).

Studies using multi-b-value diffusion imaging techniques for the differential diagnosis of HGGs and metastases are still few in number, however, and the data published to date will need to be validated in other patient cohorts.

## Grading of Gliomas

Determining the correct grade of a tumour, or at least distinguishing whether it is high-grade glioma (HGG) or low-grade glioma (LGG), is essential for choosing adequate therapy and also for estimating a patient's prognosis. However, the possibilities of conventional MRI methods in this application are somewhat limited. There are already several papers dealing with the use of IVIM. Several authors have demonstrated increase in $f$ and/or $D^*$ parameters in HGGs compared to LGGs, probably as a consequence of increased blood flow and neovascularization in malignant tumours.(43–46).

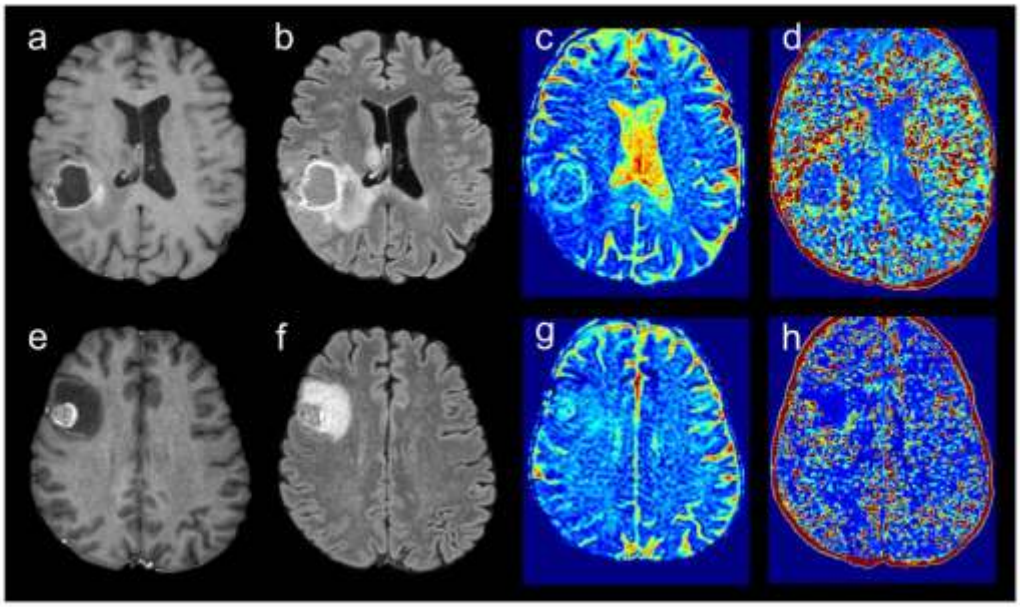



**Figure 4.** Different patterns of diffusion parameters in peritumoral oedema are shown in patients with glioblastoma (IDH wild type) (a–d) and metastasis of malignant melanoma (e–h). a,e - T1-w post-contrast images, b,f – FLAIR images, c,g – f maps, d,h – D* maps. Peritumoral FLAIR hyperintense area with small enhancing infiltration in patient with glioblastoma has relatively low f (0.162) values (c) and higher D* (15255 *$10^{-6}$ $mm^2/s$) values with marked area medial to the tumour core (d). Conversely, the oedema surrounding metastatic lesion has comparatively higher f (0.212) (g) and low D* (9787 *$10^{-6}$ $mm^2/s$) values (h). All the values were calculated as means measured within the segmentation mask covering the whole peritumoral FLAIR hyperintense area. FLAIR, fluid-attenuated inversion recovery; IDH, isocitrate dehydrogenase.

In any case, compared to the other applications discussed in this review, most papers dealing with IVIM technique focus on the topic of glioma grading. Upon closer examination of the data published to date, however, we can see that the results are not uniform, although some trends can be identified. A meta-analysis by Luo et al. summarising the results of 6 papers examining 252 tumours identified *f* as the parameter with the highest pooled specificity and sensitivity (sensitivity 89%, specificity 88%). This parameter also had the largest area under the curve (AUC = 0.94).(47) From another meta-analysis including nine papers examining a total of 318 gliomas, however, Li et al. reported that the *f* parameter did not differ significantly between LGGs and HGGs in the tumour parenchyma ($p = 0.056$) and, on the contrary, the *D**parameter ($p = 0.002$) was significantly higher in HGGs and the *D* parameter ($p = 0.001$) significantly lower in HGGs.(48) Both these meta-analyses also identify possible reasons for the equivocal and often divergent results. In every case, differences in methodology can be mentioned, as the different numbers of b-values used, as well as their different distributions, have been shown significantly to affect the values of the IVIM parameters obtained.(49). The varying methods and results of published studies are also illustrated in Table A (included in supplementary material), wherein we compare all relevant articles that we were able to find up to the time of submitting this paper.

Given the known impact of the degree of tissue vascularity on IVIM parameters, some papers have compared IVIM for glioma grading with conventional methods of perfusion imaging, either methods using contrast agent administration (DSC and dynamic contrast enhancement, DCE) or the arterial spin labelling (ASL) method, which does not require contrast agent administration. For example, a paper comparing IVIM with DSC found that the *f* parameter was significantly higher ($p < 0.0001$) in HGGs compared to LGGs, had the highest AUC (0.95) of all those parameters examined, and also correlated with relative cerebral blood volume (rCBV).(43) In another rather large study on 120 glioma patients comparing IVIM with perfusion techniques DCE and ASL, Yan et al. found that many parameters obtained by all these methods showed significant differences between the HGG and LGG groups. Specifically, the parameter *f* was again the most important in IVIM and was significantly higher in HGG ($p = 0.0029$). When comparing the methods with each other, however, the DCE method provided more accurate results.(50) Other work has found

that IVIM discriminates between HGG and LGG more accurately than does ASL, although IVIM performs more poorly compared to conventional diffusion parameters.(46).

A number of authors have also mentioned the possibility of using DKI imaging for the purpose of glioma grading, and there has been a general consensus across studies that increasing glioma grade correlates with increasing MK values. According to previous studies, this parameter reflects the microstructural complexity of tissues, which can be correlated with known attributes of gliomas of different malignancy grades. High-grade gliomas are characterised by a tendency towards greater structural complexity and heterogeneity of tumour tissue, richer vascularity, and more frequent occurrence of diffusivity barriers.(51–53).

The results of multi-b-value diffusion imaging in a patient with low-grade glioma are shown in Figure 5, where, among changes in other parameters, the decrease in *K* values within the tumour compared to normal brain tissues is particularly marked.

Currently, gliomas are classified according to not only histological but also molecular features, such as the mutational status of isocitrate dehydrogenase 1 (IDH1). In this context, the first papers documenting correlations of IDH status with IVIM imaging parameters are emerging, but the results are not yet uniform. One study reported significantly higher ADC values in LGGs with IDH1 mutation compared to LGGs with wild-type (i.e. non-mutated) IDH1, and, similarly, significantly higher ADC values as well as lower *D** and *f* values were found in HGGs with IDH1 mutation compared to HGGs with wild-type IDH1.(45) In another work, however, HGG with IDH1 mutation was reported to be associated with an increase in ADC and *D* parameters in addition to an increase in the *f* parameter compared to IDH1 wild-type tumours while the *D**parameter did not differ significantly between these subgroups.(54) Thus, further studies and research will be needed in this area to address also the methodology and reproducibility of this method.

### Distinguishing Progression From Pseudoprogression

A specific problem in the postoperative follow-up of patients with malignant intra-axial lesions is early and correct detection of tumour progression or recurrence. That is very difficult, however, as post-therapeutic changes may mimic true tumour progression (pseudoprogression).(55) The ability to differentiate true

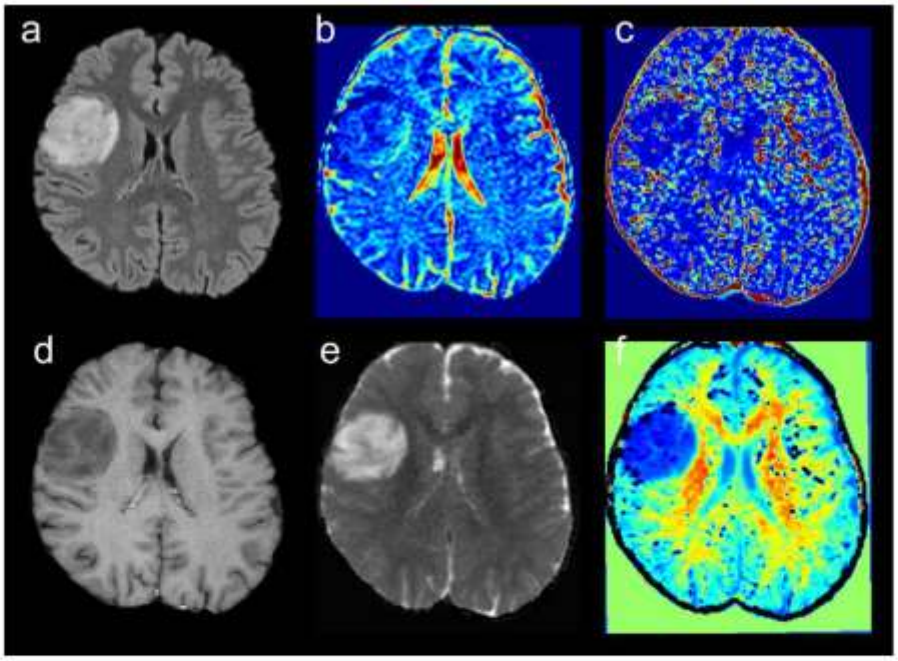



**Figure 5.** Low-grade glioma (diffuse astrocytoma grade II, IDH1 mutation) depicted on axial FLAIR (a) and post-contrast T1 (d) images, maps of f, D*, D, and K parameters (b, c, e, and f, respectively). The values of perfusion related parameters f and D* as well as K values measured as means within the whole tumour core (0.143, 13800 $*10^{-6}$ $mm^2/s$, and 0.460, respectively) were lower compared to those measured within contralateral normally appearing white matter (0.150, 20100$*10^{-6}$ $mm^2/s$, and 1.164 respectively). FLAIR, fluid-attenuated inversion recovery; IDH, isocitrate dehydrogenase.

progression from pseudoprogression is limited with conventional MR imaging. Therefore, multi-b-value diffusion imaging methods may provide options to refine the detection of recurrence. Again, the starting point here is that evidence of increased vascularity and blood flow is crucial for distinguishing true progression, as has been demonstrated by a number of studies using conventional perfusion techniques.(56–59).

The results of studies using the IVIM technique published to date support this assumption. It has been shown that the perfusion parameter *f* is significantly higher in true progression and has sensitivity and specificity comparable to those of perfusion parameters obtained by DSC.(60).

A study investigating use of the DKI technique in the same application showed increased MK and decreased MD values in patients with tumour recurrence compared to a subgroup of patients with post-therapeutic pseudoprogression.(61) A similar conclusion (i.e. increased MK values in patients with HGG recurrence) was reached by Shi et al. from their recent study wherein the DKI technique was investigated simultaneously with DSC perfusion imaging. They also mentioned the possibility of combining MK and CBV parameters to increase diagnostic accuracy.(62) The results of multi-b-value diffusion imaging in a patient with pseudoprogression are shown in Figure 6.

## CONTROVERSIES AND FUTURE DIRECTIONS

Although multi-b-value imaging appears to be an interesting technique, it does not yet seem ready for routine application. Therefore, increased effort in researching this topic is desirable as continued improvement of both data acquisition and post-processing methods may help address main limitations of the technique. These limitations consist particularly in low or not entirely clear reproducibility of the calculated parameters, which may be related to differences in the setting of acquisition parameters, and especially the number and distribution of b-values.(63) Furthermore, differences in data postprocessing, namely application of different curve-fitting models for diffusion data, have significant impact on the method's capability to differentiate between different tissues and lesions.(64).

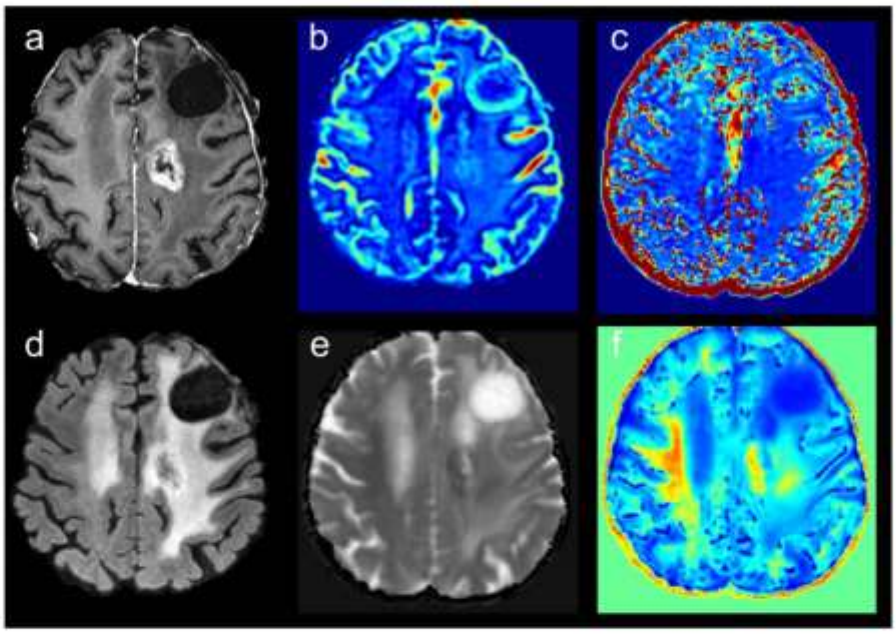



**Figure 6.** Contrast-enhancing lesion within the white matter of the left cerebral hemisphere corresponding with pseudoprogression in a patient treated for glioblastoma (grade IV, IDH wildtype) by surgical resection and radiotherapy. **a** – T1-w post-contrast axial image, **d** – FLAIR axial image, **b–f** map, **c** – D* map, **e** – D map, **f** – K map. The enhancing lesion (a) is located dorsally from the non-enhancing post-surgery area in the left frontal lobe; extensive leukoencephalopathy is visible in FLAIR image (d). Mean D* and K values measured within the enhancing parts of the lesion ($4338 * 10^{-6}$ $mm^2/s$ and 0.865 respectively) were lower compared to those measured within contralateral NAWM ($5112 * 10^{-6}$ $mm^2/s$ and 1.187), while D and f values of the lesion ($906 * 10^{-6}$ $mm^2/s$ and 0.206, respectively) were higher compared to those for NAWM ($538 * 10^{-6}$ $mm^2/s$ and 0.151, respectively). Especially low D* values are notable on D* map (c). FLAIR, fluid-attenuated inversion recovery; IDH, isocitrate dehydrogenase; NAWM, normally appearing white matter.

Moreover, different studies use different methodologies for selecting the regions of interest (ROIs) from which the IVIM parameters are extracted for statistical analysis. The ROIs are usually designated manually based on visual assessment of parametric maps. Therefore, it is suggested that more objective and reproducible approaches should be adopted for assessing diffusion parameters, such as segmentation of 3D masks of individual tumour components and peritumoral region that is followed by analysis of histograms or radiomics methods.(65) Inasmuch as implementing of these approaches has been reported in only a limited number of papers to date (see Table A in supplemental material), however, their utility cannot yet be fairly assessed.

Thus, future studies should carefully evaluate potential confounding factors, identify sources of data heterogeneity, and evaluate the reproducibility of this technique across different MRI devices.

Another known disadvantage of multi-b-value imaging is a longer scanning time. Therefore, development and evaluation of simplified methods of multi-b-value imaging using a limited number of b-values might be another possible research direction. Several initial studies on this topic have already been published(63,66–69) and, although it is suggested that a simplified method can lead to decreased accuracy,(63) the results are rather promising. Another opportunity for accelerating the acquisition can be to use a compressed sensing method or simultaneous multi-slice approach, which can lead to results similar to those from conventional acquisition methods but with shorter scan time.(70–72).

Another issue that is addressed variously by different studies is determination of the extent of tumour to be analysed. Most studies use the area of postcontrast enhancement, but the use of PET radiopharmaceuticals has also been suggested and could represent an interesting direction for research.(69) The available data concerning differentiation of the brain lesions of various aetiologies are still relatively scarce, so further research on larger cohorts of patients is desirable.

## CONCLUSION

Multi-b-value diffusion imaging is a promising technique having potential to provide valuable new biomarkers for the classification of brain lesions, whether for differentiating tumour from non-tumour aetiologies or in differential diagnosis of individual tumours. This technique is essentially an extension of the well-established conventional DWI imaging, but it requires new methodological approaches to data analysis using nonconventional diffusivity characteristics (pseudo-diffusion and non-Gaussian diffusion) in structurally complex tissues. Due to some effect of tissue perfusion in the case of IVIM analysis, this method can be seen as a possible alternative or complementary method to conventional perfusion techniques. It has the advantage over these techniques of not requiring contrast agent application while still allowing the assessment of conventional diffusivity parameters such as ADC. A limitation of this relatively new method is the inconsistency to date in image data analysis methods and acquisition protocols. Because results from the limited number of studies published so far are not yet uniform in some areas, further research as well as more methodological work on the reproducibility will be needed to validate the clinical relevance of these methods.

## FUNDING

This work was supported by the Ministry of Health of the Czech Republic, Grant No. NU21–08–00359.

## APPENDIX A. SUPPORTING INFORMATION

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.acra.2023.10.002.

## REFERENCES

1. Bihan DLe. What can we see with IVIM MRI? NeuroImage 2019; 187:56-67. https://doi.org/10.1016/j.neuroimage.2017.12.062
2. Bihan DL, Iima M. Diffusion magnetic resonance imaging: what water tells us about biological tissues. PLoS Biol 2015; 13(7):e1002203. https://doi.org/10.1371/journal.pbio.1002203
3. Attanwala R, Picker W. Whole body MRI: improved lesion detection and characterization with diffusion weighted techniques. J Magn Reson Imaging 2013; 38(2):253-268. https://doi.org/10.1002/jmri.24283
4. Zhang L, Min Z, Tang M, et al. The utility of diffusion MRI with quantitative ADC measurements for differentiating high-grade from low-grade cerebral gliomas: Evidence from a meta-analysis. J Neurol Sci 2017; 373:9–15. https://doi.org/10.1016/j.jns.2016.12.008
5. Gerstner ER, Sorensen AG. Diffusion and diffusion tensor imaging in brain cancer. Semin Radiat Oncol 2011; 21(2):141–146. https://doi.org/10.1016/j.semradonc.2010.10.005
6. Grant KB, et al. Comparison of calculated and acquired high b-value diffusion weighted imaging in prostate cancer. Abdom Imaging 2015; 40(3):578–586. https://doi.org/10.1007/s00261-014-0246-2
7. Eichner C, et al. Real diffusion-weighted MRI enabling true signal averaging and increased diffusion contrast. NeuroImage 2015; 122:373–384. https://doi.org/10.1016/j.neuroimage.2015.07.074
8. Paganelli C, et al. Optimizing b-values schemes for diffusion MRI of the brain with segmented Intravoxel Incoherent Motion (IVIM) model. J Appl Clin Med Phys 2023; 24(6):e13986. https://doi.org/10.1002/acm2.13986
9. Nai Y-H, et al. Effects of fitting methods, high b-values and image quality on diffusion and perfusion quantification and reproducibility in the calf. Comput Biol Med 2023; 157:106746. https://doi.org/10.1016/j.compbiomed.2023.106746
10. Guo D, Jiang B. Noninvasively evaluating the grade and IDH mutation status of gliomas by using mono-exponential, bi-exponential diffusion-weighted imaging and three-dimensional pseudo-continuous arterial spin labeling. Eur J Radiol 2023; 160:110721. https://doi.org/10.1016/j.ejrad.2023.110721
11. Liu Y-F, et al. Characterizing sensorimotor-related area abnormalities in amyotrophic lateral sclerosis: an intravoxel incoherent motion magnetic resonance imaging study. Acad Radiol 2022; 29(Suppl 3):S141–S146. https://doi.org/10.1016/j.acra.2021.07.006
12. Razek AAKAbdel. Editorial for "Preliminary Assessment of Intravoxel Incoherent Motion Diffusion-Weighted MRI (IVIM-DWI) Metrics in Alzheimer's Disease". J Magn Reson Imaging 2020; 52(6):1827–1828. https://doi.org/10.1002/jmri.27309
13. Paschoal AM, et al. Contrast-agent-free state-of-the-art MRI on cerebral small vessel disease—part 1. ASL, IVIM, and CVR. NMR Biomed 2022; 35(8):e4742. https://doi.org/10.1002/nbm.4742

14. Bihan DL, Turner R. The capillary network: a link between IVIM and classical perfusion. Magn Reson Med 1992; 27(1):171–178. https://doi.org/10.1002/mrm.1910270116
15. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. Biophys J 1994; 66(1):259–267.
16. Jensen JH, Helpern JA, Ramani A, et al. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magn Reson Med 2005; 53(6):1432–1440. https://doi.org/10.1002/mrm.20508
17. Pang H, et al. Diffusional kurtosis imaging for differentiating between high-grade glioma and primary central nervous system lymphoma. J Magn Reson Imaging 2016; 44(1):30–40. https://doi.org/10.1002/jmri.25090
18. Xu W, et al. Diffusion kurtosis imaging in evaluating the mild cognitive impairment of occupational aluminum workers. Acad Radiol 2023; S1076-6332(22). https://doi.org/10.1016/j.acra.2022.12.003
19. Faehndrich J, Weidauer S, Pilatus U, et al. Neuroradiological viewpoint on the diagnostics of space-occupying brain lesions. Clin Neuroradiol 2011; 21(3):123–139. https://doi.org/10.1007/s00062-011-0073-6
20. Xu X-X, et al. Can diffusion-weighted imaging be used to differentiate brain abscess from other ring-enhancing brain lesions? A meta-analysis. Clin Radiol 2014; 69(9):909–915. https://doi.org/10.1016/j.crad.2014.04.012
21. Erdogan C, Hakyemez B, Yildirim N, et al. Brain abscess and cystic brain tumor: discrimination with dynamic susceptibility contrast perfusion-weighted MRI. J Comput Assist Tomogr 2005; 29(5):663–667. https://doi.org/10.1097/01.rct.0000168868.50256.53
22. Muccio CF, Esposito G, Bartolini A, et al. Cerebral abscesses and necrotic cerebral tumours: differential diagnosis by perfusion-weighted magnetic resonance imaging. Radiol Med (Torino) 2008; 113(5):747–757. https://doi.org/10.1007/s11547-008-0254-9
23. Bisdas S, et al. Correlative assessment of tumor microcirculation using contrast-enhanced perfusion MRI and intravoxel incoherent motion diffusion-weighted MRI: is there a link between them? NMR Biomed 2014; 27(10):1184–1191. https://doi.org/10.1002/nbm.3172
24. Eisele P, et al. Reduced diffusion in a subset of acute MS lesions: a serial multiparametric MRI study. Am J Neuroradiol 2012; 33(7):1369–1373. https://doi.org/10.3174/ajnr.A2975
25. Suh CH, Kim HS, Jung SC, et al. MRI findings in tumefactive demyelinating lesions: a systematic review and meta-analysis. Am J Neuroradiol 2018; 39(9):1643–1649. https://doi.org/10.3174/ajnr.A5775
26. Mabray MC, et al. Performance of apparent diffusion coefficient values and conventional MRI features in differentiating tumefactive demyelinating lesions from primary brain neoplasms. Am J Roentgenol 2015; 205(5):1075–1085. https://doi.org/10.2214/AJR.14.13970
27. Chu S, et al. Dynamic contrast-enhanced T2*-weighted MR imaging of tumefactive demyelinating lesions. Am J Neuroradiol 2001; 22(6):1109–1116.
28. Osa Garcia A, Brambati SM, Desautels A, et al. Timing stroke: a review on stroke pathophysiology and its influence over time on diffusion measures. J Neurol Sci 2022; 441:120377. https://doi.org/10.1016/j.jns.2022.120377
29. Pavilla A, Gambarota G, Arrigo A, et al. Toward an intravoxel incoherent motion 2-in-1 magnetic resonance imaging sequence for ischemic stroke diagnosis? an initial clinical experience with 1.5T magnetic resonance. J Comput Assist Tomogr 2022; 46(1):110–115. https://doi.org/10.1097/RCT.0000000000001243
30. Federau C, et al. Intravoxel incoherent motion perfusion imaging in acute stroke: initial clinical experience. Neuroradiology 2014; 56(8):629–635. https://doi.org/10.1007/s00234-014-1370-y
31. Zhu G, et al. Comparison of MRI IVIM and MR perfusion imaging in acute ischemic stroke due to large vessel occlusion. Int J Stroke Off J Int Stroke Soc &;2020; 15(3):332–342. https://doi.org/10.1177/1747493019873519
32. Yao Y, et al. Intravoxel incoherent motion diffusion-weighted imaging in stroke patients: initial clinical experience. Clin Radiol 2016; 71(8). https://doi.org/10.1016/j.crad.2016.04.019
33. Du X, He Y, Lin W. Diagnostic accuracy of the diffusion-weighted imaging method used in association with the apparent diffusion coefficient for differentiating between primary central nervous system lymphoma and high-grade glioma: systematic review and meta-analysis. Front Neurol 2022; 13:882334. https://doi.org/10.3389/fneur.2022.882334
34. Choi YS, et al. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using the initial area under the curve derived from dynamic contrast-enhanced MR and the apparent diffusion coefficient. Eur Radiol 2017; 27(4):1344–1351. https://doi.org/10.1007/s00330-016-4464-2
35. Suh CH, Kim HS, Jung SC, et al. MRI 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(2):560–572. https://doi.org/10.1002/jmri.26662
36. Haopeng P, et al. Diffusion kurtosis imaging differs between primary central nervous system lymphoma and high-grade glioma and is correlated with the diverse nuclear-to-cytoplasmic ratio: a histopathologic, biopsy-based study. Eur Radiol 2020; 30(4):2125–2137. https://doi.org/10.1007/s00330-019-06544-7
37. Maurer MH, et al. Glioblastoma multiforme versus solitary supratentorial brain metastasis: differentiation based on morphology and magnetic resonance signal characteristics. RöFo - Fortschritte Auf Dem Geb Röntgenstrahlen Bildgeb Verfahr 2013; 185(03):235–240. https://doi.org/10.1055/s-0032-1330318
38. Villanueva-Meyer JE, Mabray MC, Cha S. Current clinical brain tumor imaging. Neurosurgery &;2017; 81(3):397–415. https://doi.org/10.1093/neuros/nyx103
39. Martin-Noguerol T, Mohan S, Santos-Armentia E, et al. Advanced MRI assessment of non-enhancing peritumoral signal abnormality in brain lesions. Eur J Radiol 2021; 143:109900. https://doi.org/10.1016/j.ejrad.2021.109900
40. Suh CH, Kim HS, Jung SC, et al. Diffusion-weighted imaging and diffusion tensor imaging for differentiating high-grade glioma from solitary brain metastasis: a systematic review and meta-analysis. Am J Neuroradiol 2018; 39(7):1208–1214. https://doi.org/10.3174/ajnr.A5650
41. Song S, et al. Intravoxel incoherent motion diffusion weighted imaging of high-grade gliomas and brain metastases: efficacy in preoperative differentiation. Int J Clin Exp Med 2015; 11:7964–7971.
42. Tan Y, et al. Differentiation of high-grade-astrocytomas from solitary-brain-metastases: comparing diffusion kurtosis imaging and diffusion tensor imaging. Eur J Radiol &;2015; 84(12):2618–2624. https://doi.org/10.1016/j.ejrad.2015.10.007
43. Togao O, et al. Differentiation of high-grade and low-grade diffuse gliomas by intravoxel incoherent motion MR imaging. Neuro-Oncol 2016; 18(1):132–141. https://doi.org/10.1093/neuonc/nov147
44. Shen N, et al. Intravoxel incoherent motion diffusion-weighted imaging analysis of diffusion and microperfusion in grading gliomas and comparison with arterial spin labeling for evaluation of tumor perfusion. J Magn Reson Imaging 2016; 44(3):620–632. https://doi.org/10.1002/jmri.25191
45. Wang X, Chen X-Z, Shi L, et al. Glioma grading and IDH1 mutational status: assessment by intravoxel incoherent motion MRI. Clin Radiol 2019; 74(8):651.e7–651.e14. https://doi.org/10.1016/j.crad.2019.03.020
46. Lin Y, et al. Comparison of Intravoxel Incoherent motion diffusion-weighted MR imaging and arterial spin labeling MR imaging in gliomas. BioMed Res Int 2015 :234245. https://doi.org/10.1155/2015/234245
47. Luo H, He L, Cheng W, et al. The diagnostic value of intravoxel incoherent motion imaging in differentiating high-grade from low-grade gliomas: a systematic review and meta-analysis. Br J Radiol 2021; 94(1121):20201321. https://doi.org/10.1259/bjr.20201321
48. Li W-F, Niu C, Shakir TM, et al. An evidence-based approach to assess the accuracy of intravoxel incoherent motion imaging for the grading of brain tumors. Medicine (Baltimore) 2018; 97(45):e13217. https://doi.org/10.1097/MD.0000000000013217
49. Chabert S, et al. 'Impact of b-Value Sampling Scheme on Brain IVIM Parameter Estimation in Healthy Subjects', *Magn. Reson. Med. Sci. MRMS Off J Jpn Soc Magn Reson Med* 2020; 19(3):216–226. https://doi.org/10.2463/mrms.mp.2019-0061
50. L.-F. Yan *et al.*, Perfusion, Diffusion, Or Brain Tumor Barrier Integrity: Which Represents The Glioma Features Best? , *Cancer Management and Research*, Nov. 27, 2019. https://www.dovepress.com/perfusion-diffusion-or-brain-tumor-barrier-integrity-which-represents-peer-reviewed-article-CMAR1 (accessed Jan. 06, 2021).
51. Zhang J, Chen X, Chen D, et al. Grading and proliferation assessment of diffuse astrocytic tumors with monoexponential, biexponential, and stretched-exponential diffusion-weighted imaging and diffusion kurtosis imaging. Eur J Radiol 2018; 109:188–195. https://doi.org/10.1016/j.ejrad.2018.11.003

52. Zhang L, et al. Noninvasively evaluating the grading of glioma by multiparametric magnetic resonance imaging. Acad Radiol 2021; 28(5):e137–e146. https://doi.org/10.1016/j.acra.2020.03.035
53. Van Cauter S, et al. Gliomas: diffusion kurtosis MR imaging in grading. Radiology 2012; 263(2):492–501. https://doi.org/10.1148/radiol.12110827
54. Wang C, Dong H. Intravoxel incoherent motion magnetic resonance imaging in predicting IDH1 gene mutations in high-grade gliomas. Acta Radiol Stockh Swed 1987 2021; 62(10):1412–1417. https://doi.org/10.1177/0284185121011914
55. Abbasi AW, Westerlaan HE, Holtman GA, et al. Incidence of tumour progression and pseudoprogression in high-grade gliomas: a systematic review and meta-analysis. Clin Neuroradiol 2018; 28(3):401–411. https://doi.org/10.1007/s00062-017-0584-x
56. Umemura Y, et al. DCE-MRI perfusion predicts pseudoprogression in metastatic melanoma treated with immunotherapy. J Neurooncol 2020; 146(2):339–346. https://doi.org/10.1007/s11060-019-03379-6
57. Thomas AA, et al. Dynamic contrast enhanced T1 MRI perfusion differentiates pseudoprogression from recurrent glioblastoma. J Neurooncol 2015; 125(1):183–190. https://doi.org/10.1007/s11060-015-1893-z
58. Wang S, et al. Differentiating tumor progression from pseudoprogression in patients with glioblastomas using diffusion tensor imaging and dynamic susceptibility contrast MRI. Am J Neuroradiol 2016; 37(1):28–36. https://doi.org/10.3174/ajnr.A4474
59. Manning P, et al. Differentiation of progressive disease from pseudoprogression using 3D PCASL and DSC perfusion MRI in patients with glioblastoma. J Neurooncol 2020; 147(3):681–690. https://doi.org/10.1007/s11060-020-03475-y
60. Kim HS, Suh CH, Kim N, et al. Histogram analysis of intravoxel incoherent motion for differentiating recurrent tumor from treatment effect in patients with glioblastoma: initial clinical experience. Am J Neuroradiol 2014; 35(3):490–497. https://doi.org/10.3174/ajnr.A3719
61. Wu X, et al. Differentiating high-grade glioma recurrence from pseudoprogression: comparing diffusion kurtosis imaging and diffusion tensor imaging. Eur J Radiol 2021; 135:109445. https://doi.org/10.1016/j.ejrad.2020.109445
62. Shi W, Qu C, Wang X, et al. Diffusion kurtosis imaging combined with dynamic susceptibility contrast-enhanced MRI in differentiating high-grade glioma recurrence from pseudoprogression. Eur J Radiol 2021; 144:109941. https://doi.org/10.1016/j.ejrad.2021.109941
63. Conklin J, Heyn C, Roux M, et al. A simplified model for intravoxel incoherent motion perfusion imaging of the brain. Am J Neuroradiol 2016; 37(12):2251–2257. https://doi.org/10.3174/ajnr.A4929
64. Keil VC, et al. Intravoxel incoherent motion MRI in the brain: impact of the fitting model on perfusion fraction and lesion differentiability. J Magn Reson Imaging 2017; 46(4):1187–1199. https://doi.org/10.1002/jmri.25615
65. Mesropyan N, et al. Comparison of different ROI analysis methods for liver lesion characterization with simplified intravoxel incoherent motion (IVIM). Sci Rep 2021; 11:22752. https://doi.org/10.1038/s41598-021-01168-6
66. Cao M, et al. Application of a simplified method for estimating perfusion derived from diffusion-weighted MR imaging in glioma grading. Front Aging Neurosci 2018; 9:432. https://doi.org/10.3389/fnagi.2017.00432
67. Hino T, et al. Clinical efficacy of simplified intravoxel incoherent motion imaging using three b-values for differentiating high- and low-grade gliomas. PLoS ONE 2018; 13(12):e0209796. https://doi.org/10.1371/journal.pone.0209796
68. Wang X, Cao M, Chen H, et al. Simplified perfusion fraction from diffusion-weighted imaging in preoperative prediction of IDH1 mutation in WHO grade II-III gliomas: comparison with dynamic contrast-enhanced and intravoxel incoherent motion MRI. Radiol Oncol 2020; 54(3):301–310. https://doi.org/10.2478/raon-2020-0027
69. Loução R, Oros-Peusquens A-M, Langen K-J, et al. A fast protocol for multiparametric characterisation of diffusion in the brain and brain tumours. Front Oncol 2021; 11:554205. https://doi.org/10.3389/fonc.2021.554205
70. Xu H, et al. Scan time reduction in intravoxel incoherent motion diffusion-weighted imaging and diffusion kurtosis imaging of the abdominal organs: using a simultaneous multislice technique with different acceleration factors. J Comput Assist Tomogr 2021; 45(4):507–515. https://doi.org/10.1097/RCT.0000000000001188
71. Landman BA, Bogovic JA, Wan H, et al. Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. NeuroImage 2012; 59(3):2175–2186. https://doi.org/10.1016/j.neuroimage.2011.10.011
72. Kuhnt D, et al. Fiber tractography based on diffusion tensor imaging compared with high-angular-resolution diffusion imaging with compressed sensing: initial experience. Neurosurgery 2013; 72(0 1):165–175. https://doi.org/10.1227/NEU.0b013e318270d9fb

nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-021-00588-0

Check for updates

# Generic acquisition protocol for quantitative MRI of the spinal cord

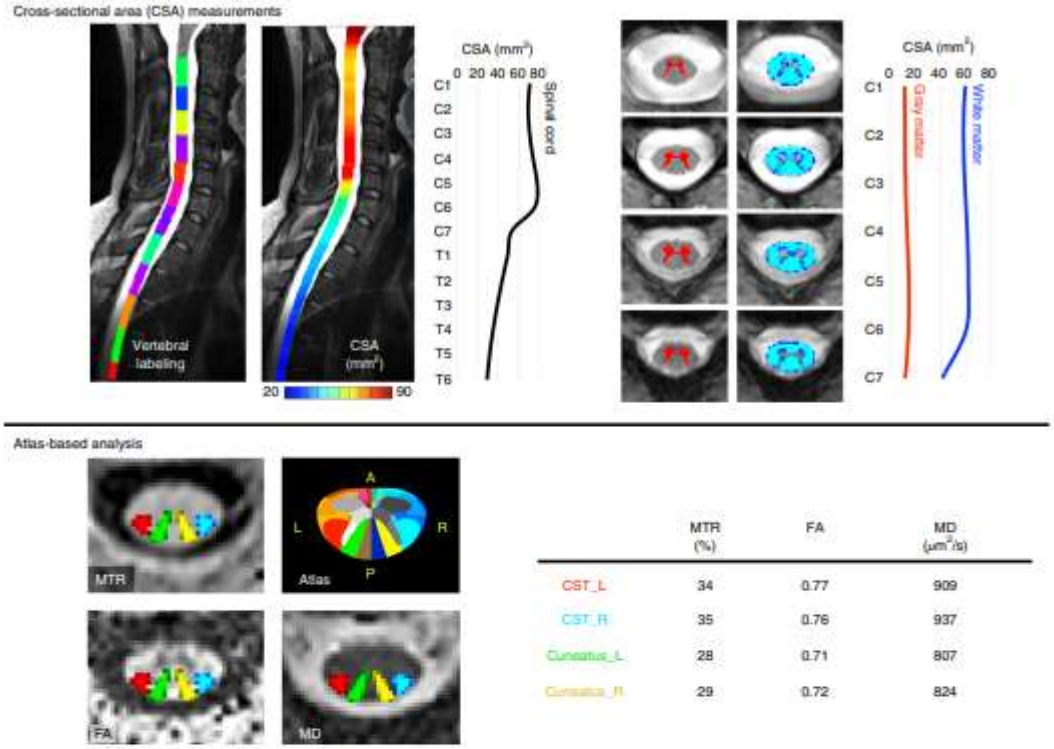
Julien Cohen-Adad[1,2,3], Eva Alonso-Ortiz[1], Mihael Abramovic[4], Carina Arneitz[4], Nicole Atcheson[5], Laura Barlow[6], Robert L. Barry[7,8,9], Markus Barth[10], Marco Battiston[11], Christian Büchel[12], Matthew Budde[13], Virginie Callot[14,15], Anna J. E. Combes[16], Benjamin De Leener[17,18], Maxime Descoteaux[19,20], Paulo Loureiro de Sousa[21], Marek Dostál[22], Julien Doyon[23], Adam Dvorak[24], Falk Eippert[25], Karla R. Epperson[26], Kevin S. Epperson[26], Patrick Freund[27], Jürgen Finsterbusch[12], Alexandru Foias[1], Michela Fratini[28,29], Issei Fukunaga[30], Claudia A. M. Gandini Wheeler-Kingshott[11,31,32], Giancarlo Germani[32], Guillaume Gilbert[33], Federico Giove[29,34], Charley Gros[1,5], Francesco Grussu[11,35], Akifumi Hagiwara[30], Pierre-Gilles Henry[36], Tomáš Horák[37], Masaaki Hori[38], James Joers[36], Kouhei Kamiya[39], Haleh Karbasforoushan[40,41], Miloš Keřkovský[22], Ali Khatibi[23,42], Joo-Won Kim[43], Nawal Kinany[44,45], Hagen Kitzler[46], Shannon Kolind[6,24,47], Yazhuo Kong[48,49,50], Petr Kudlička[37], Paul Kuntke[46], Nyoman D. Kurniawan[5], Slawomir Kusmia[51,52,53], René Labounek[54,55], Maria Marcella Laganà[56], Cornelia Laule[57], Christine S. Law[58], Christophe Lenglet[36], Tobias Leutritz[59], Yaou Liu[60,61], Sara Llufriu[62], Sean Mackey[58], Eloy Martinez-Heras[62], Loan Mattera[63], Igor Nestrasil[36,54], Kristin P. O'Grady[16,64], Nico Papinutto[65], Daniel Papp[1,50], Deborah Pareto[66], Todd B. Parrish[40], Anna Pichiecchio[31,32], Ferran Prados[11,52,67], Àlex Rovira[66], Marc J. Ruitenberg[68], Rebecca S. Samson[11], Giovanni Savini[32], Maryam Seif[27,59], Alan C. Seifert[43], Alex K. Smith[50], Seth A. Smith[16,64], Zachary A. Smith[69], Elisabeth Solana[62], Yuichi Suzuki[39], George Tackley[51], Alexandra Tinnermann[12], Jan Valošek[70], Dimitri Van De Ville[44,45], Marios C. Yiannakas[11], Kenneth A. Weber II[58], Nikolaus Weiskopf[59,71], Richard G. Wise[51,72], Patrik O. Wyss[4] and Junqian Xu[43]

**Quantitative spinal cord (SC) magnetic resonance imaging (MRI) presents many challenges, including a lack of standardized imaging protocols. Here we present a prospectively harmonized quantitative MRI protocol, which we refer to as the *spine generic* protocol, for users of 3T MRI systems from the three main manufacturers: GE, Philips and Siemens. The protocol provides guidance for assessing SC macrostructural and microstructural integrity: T1-weighted and T2-weighted imaging for SC cross-sectional area computation, multi-echo gradient echo for gray matter cross-sectional area, and magnetization transfer and diffusion weighted imaging for assessing white matter microstructure. In a companion paper from the same authors, the *spine generic* protocol was used to acquire data across 42 centers in 260 healthy subjects. The key details of the *spine generic* protocol are also available in an open-access document that can be found at https://github.com/spine-generic/protocols. The protocol will serve as a starting point for researchers and clinicians implementing new SC imaging initiatives so that, in the future, inclusion of the SC in neuroimaging protocols will be more common. The protocol could be implemented by any trained MR technician or by a researcher/clinician familiar with MRI acquisition.**

## Introduction

Quantitative MRI (qMRI) aims to provide objective continuous metrics that specifically reflect the morphology, microstructure and/or chemical composition of tissues[1,2], thereby enabling deeper insight and understanding of disease pathophysiology. While qMRI techniques have been successfully implemented in the brain for several decades, they remain largely underutilized for spinal cord (SC) imaging in both clinical and research settings, mostly as a direct consequence of the many challenges that need to be overcome in order to acquire good-quality data[3,4].

A full list of affiliations appears at the end of the paper.




| | MTR (%) | FA | MD ($\mu m^2/s$) |
|---|---|---|---|
| CST_L | 34 | 0.77 | 909 |
| CST_R | 35 | 0.76 | 937 |
| Cuneatus_L | 28 | 0.71 | 807 |
| Cuneatus_R | 29 | 0.72 | 824 |

**Fig. 1 | Illustration of the MRI metrics that could be extracted from the spine generic protocol.** The top panel 'Cross-sectional area (CSA) measurements' shows morphometric measures of the spinal cord and its gray and white matter. The bottom panel 'Atlas-based analysis' on the left shows axial views of qMRI maps: MTR, fractional anisotropy (FA) and mean diffusivity (MD), with an overlay of four spinal tracts of general interest: the descending corticospinal tract (CST) and the ascending cuneatus, left (L) and right (R). The 'Atlas' image corresponds to the white matter atlas[82], which includes 30 WM tracts that could be used for computing metrics within specific tracts of interest. This atlas also includes six parcellations of the GM. The table presents average values of each metric in the corresponding tract.

For the past 20 years, researchers have been developing methods to overcome the challenges around SC imaging, including more sensitive coil arrays[5,6] and advanced pulse sequences for mitigating motion and susceptibility artifacts[4,7]. As a result, it is now possible to acquire SC qMRI data that have a strong potential for providing new insights into SC anatomy and function. However, a remaining issue is that there is no clear consensus within the imaging community for acquiring SC qMRI data, leading to (i) wasted time and money spent on pilot scans for every new SC research initiative, and (ii) large variability in imaging parameters for multisite, multimanufacturer studies, hampering statistics for assessing biomarkers.

## Development of the protocol

The present study gathered a consortium of international SC researchers to provide a prospectively harmonized consensus protocol for acquiring high-quality qMRI of the human cervical SC at 3 Tesla (T) across the three main MRI manufacturers (GE, Philips and Siemens). We call this the *spine generic* protocol. qMRI techniques covered in the *spine generic* protocol (illustrated in Fig. 1) include:

### SC cross-sectional area (CSA)

The CSA of the whole SC has been shown to be a sensitive biomarker in multiple sclerosis (MS)[8–11], amyotrophic lateral sclerosis (ALS)[12–16], X-linked adrenoleukodystrophy with myelopathy[17], as well as both traumatic and nontraumatic SC injury[18,19]. Additionally, SC segmentation is useful for atlas-based analysis[20].

**CSA of the SC gray matter (GM)**

GM CSA is relevant for diagnosis[21] and prognosis in ALS[16]. Additionally, delineating the GM is relevant for quantifying pathologies juxtaposed with the GM (e.g., MS lesions), for functional MRI (fMRI) applications, and for atlas-based analysis.

**Diffusion tensor imaging (DTI)**

DTI is a technique that is based on multidirectionally encoded diffusion-weighted images (DWI). DTI can quantify microstructural integrity and has been deemed sensitive to degeneration and demyelination of SC white matter (WM) tracts in a variety of diseases[3,22–25] and after SC injury[19,23]. In nontraumatic SC injury, DWI appears to be a promising approach that is sensitive to presymptomatic microstructural changes[26,27].

**Magnetization transfer (MT)**

The MT technique has been shown to be sensitive to demyelination[28] and has been applied in various SC diseases, such as adrenomyeloneuropathy[29] and MS[30], as well as in SC injury[22,31].

To demonstrate the practical implementation and reproducibility of the proposed protocol, single-subject and multi-subject datasets were acquired across multiple centers. Relevant qMRI metrics were calculated using a fully automatic analysis pipeline, and those metrics were compared within site, across sites (for the same manufacturer) and across different manufacturers. Details of the datasets, processing pipelines and generated normative values are available in a companion Data Descriptor paper published in *Scientific Data*[32].

When optimizing protocols across manufacturers, a key question is: *should we minimize the differences in acquisition parameters across manufacturers, or should we optimize image quality on each platform?* The *spine generic* protocol was designed to reach a compromise between these two key aims: minimizing protocol differences in order to facilitate the interpretation of multimanufacturer studies, but at the same time we optimized parameters for each manufacturer separately when the hardware or software enabled it. For example, on the DWI protocol, the echo time (TE) was always minimized in order to maximize signal-to-noise ratio (SNR), which minimally affects the diffusion-specific signal (the b-value was kept the same). Given that platforms are equipped with different gradient nominal strength capabilities (ranging from 40 mT/m to 80 mT/m for current clinical systems), this yielded very different TEs depending on the platform. These aspects were taken into consideration when designing the spine generic protocol, resulting in a protocol with a high SNR regime that is hence less sensitive to changes in the TE. As illustrated in the companion data paper[32], fractional anisotropy values across Siemens sites equipped with gradient systems varying from 40 to 80 mT/m (TEs ranging from 55 to 99 ms), produced an intersite coefficient of variation of 3.5%, which was smaller than the intrasite coefficient of variation of 4.24%. The intersubject variability was thus higher than the intersite variability, despite the large changes in TEs. Another important consideration is that different TEs across manufacturers/models will likely result in different diffusion times. This may be an additional source of intermanufacturer variability, as it has been shown that common DWI metrics such as DTI radial diffusivity can exhibit diffusion time dependence, especially in anatomical regions containing large axons[33]. Similarly, some software versions were limited with respect to the minimum achievable repetition time (TR) on MT sequences; again, here the TR was optimized for each system separately, yielding full MT protocols (GRE-MT1/MT0/T1w) that varied from 5.4 min to 8.9 min, depending on the platform. However, in this case, magnetization transfer ratio (MTR) and magnetization transfer saturation (MTsat) were impacted by TR. This partly explains the discrepancies observed between GE and the two other manufacturers (see ref. [32]).

Because hardware and pulse sequence environments vary across manufacturers, it will never be possible to obtain the exact same acquisition configuration across manufacturers. Even for the same manufacturer, some variability could exist owing to the different specifications for different models and the adjustment and maintenance status of individual scanners (acoustic resonances, helium levels, eddy currents, software patches, etc.). From a practical standpoint, as in the case for the T1w versus T2w SC CSA (see Fig. 11 in ref. [32]), the relationship between qMRI metrics obtained from different manufacturers/models/sites can be modeled as fixed or random effects[34].

The *spine generic* protocol has been used (fully, in part or with modifications) in the following applications: imaging methods[35], methods development in healthy subjects[20,36–49], fMRI[50,51], MS[52–54], mucopolysaccharidoses[55], adrenoleukodystrophy[17], ALS[16,56], spinal muscular atrophy[57,58], degenerative cervical myelopathy[26,27,59–61] and stroke[62].

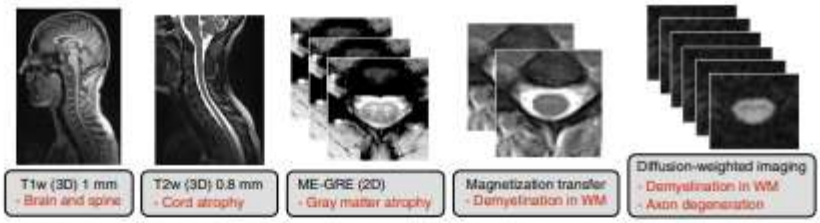



**Fig. 2 | Sequences included in the spine generic protocol (in black) with possible applications (in red).** The total acquisition time is 20–30 min, depending on the manufacturer/model.

The *spine generic* protocol has also been recommended in recent guidelines[24,63–66] and was adopted by multicenter initiatives such as the INSPIRED[67] and the CanProCo[54] studies, respectively dealing with cervical myelopathy and MS populations.

## Applications

The proposed protocol is not geared towards a specific disease, and it is suitable for imaging WM pathology (demyelination and Wallerian degeneration via axon/myelin-sensitive techniques), GM pathology (ALS, via GM CSA quantification), and traumatic and nontraumatic SC injury (structural scans to assess compression and/or to quantify atrophy above/below lesions or injury). Additional clinical scans (e.g., 2D FLAIR, STIR) that are specific to particular diseases and/or are part of the clinical routine can be added at the discretion of the researcher/clinician. Potential clinical uses of this protocol include improved diagnosis of pathology, monitoring of disease progression or recovery, and/or prediction of outcomes.

## Experimental design

### Sequences

The required sequences are illustrated in Fig. 2. Justifications for the sequence type and their pros and cons are summarized in Table 1. The manufacturer-specific sequence names are listed in Table 2.

### Shimming

Shimming refers to homogenizing the static magnetic field ($B_0$) and is a necessary step for recording reliable images, especially in regions that are prone to large $B_0$ inhomogeneities, such as the SC. Without proper shimming, fat saturation does not work effectively, slice excitation profiles are not accurate and echo planar imaging (EPI) data are prone to distortions and signal dropout, with the latter being particularly prevalent in gradient echo (GRE) imaging often used in fMRI studies.

The very first 'active' attempt to mitigate susceptibility artifacts is usually performed just before starting an MRI scan via a procedure called active shimming. This procedure consists of estimating a field map and then computing a set of 'shim coefficients', i.e., the amount of current that needs to go into each gradient and shim coil in order to minimize the static magnetic field inhomogeneity in a specified 'shim adjust volume'.

### Slice orientation

For 3D acquisitions with isotropic resolution (T1w and T2w), we recommend sagittal acquisition for an efficient superior–inferior (S-I) coverage with the minimum number of slices required to cover the cord (on the T2w). Note that typical clinical 2D scans use thick sagittal slices, which is popular for diagnosis with T1/T2/STIR/PDw contrasts, but these should not be used for measuring CSA or for template-based analysis due to the poor right–left (R-L) resolution.

For qMRI methods that produce microstructural metrics (MT, DWI), we recommend axial orientation (orthogonal to the SC) with high in-plane resolution and thick slices. This approach takes advantage of the (quasi-) coherently oriented fibers along the S-I direction to increase slice thickness and thus gain SNR. The high in-plane resolution, ideally submillimetric, is important for minimizing the partial volume effect between adjacent internal structures (WM tracts, GM), thereby ensuring accurate quantification of metrics. For 2D multislice sequences, if the sequence allows, each individual slice should ideally be orthogonal to the cord[68]. If not possible, slices should be oriented such that the

**Table 1 | Purpose, pros and cons of sequences of the spine generic protocol**

| | Purpose | Pros | Cons |
|---|---|---|---|
| T1w (3D sagittal) | - Measuring SC CSA and/or volume<br>- Registering to a template (preferred for disc labeling)<br>- Assessing lesions<br>- Measuring brain atrophy | - Efficient SNR per unit time<br>- High SC/CSF contrast (good for SC segmentation)<br>- 320 $mm^2$ FOV in ~5 min at 1 mm iso with full brain and cervical-spine coverage<br>- Low SAR<br>- High WM/GM contrast in the brain (good for cortical surface segmentation)<br>- Vertebral discs are well contrasted | - Sensitive to motion (pulsatile, swallowing)<br>- Poor WM/GM contrast in the SC |
| T2w (3D sagittal) | - Measuring SC CSA and/or volume (preferred over the 3D T1w owing to higher spatial resolution)<br>- Registering to a template registration (preferred for cord segmentation)<br>- Assessing lesions and compression | - Very high SC/CSF contrast<br>- Less sensitive to motion than the 3D T1w<br>- Better spatial resolution than the 3D T1w (0.8 mm versus 1 mm) | - High SAR<br>- Poor WM/GM contrast in the SC<br>- Cannot cover full brain in <10 min at 0.8 mm iso<br>- Poor visibility of vertebral discs<br>- More prone to Gibbs ringing artifact at high-contrast SC/CSF interface |
| DWI (2D axial) | - Computing DTI metrics (fractional anisotropy, mean diffusivity, radial diffusivity, axial diffusivity) that are sensitive to axonal damage, demyelination and degeneration[79] | - Quantify SC neural tissue microstructural properties<br>- Sensitive to WM pathologies (e.g., degenerative demyelination, injury, edema, tumor)<br>- Longitudinal monitoring of patient-specific SC microstructure (i.e., disease progression)<br>- Detect origin of microstructural damage before nonreversible changes (e.g., T2w hyperintensities, appearance of clinical symptoms)<br>- Short acquisition time (<5 min) | - Sensitive to $B_0$ inhomogeneities (EPI readout)<br>- DTI metrics are biased by SNR[80] |
| GRE-MT1/ MT0/T1w (3D axial) | - Computing MTR, MT-CSF and MTsat (requires T1w to partially compensate for B1+ homogeneity and T1 effects on the MTR[81])<br>- Detecting WM[60] and GM pathology (myelopathy) | - Quantify SC neural tissue microstructural properties<br>- Sensitive to WM pathologies (e.g., degenerative demyelination, injury, edema, tumor)<br>- Longitudinal monitoring of patient-specific SC microstructure (i.e., disease progression) | - Sensitive to motion<br>- Sensitive to $B_0$ inhomogeneities (signal dropout due to intravoxel dephasing, can be mitigated using thinner slices)<br>- Quantitative metrics sensitive to B1 (except for the ME-GRE sequence) |
| ME-GRE (2D axial) | - Segmenting the SC and GM for measuring cord/WM/GM CSA<br>- Registering to a template and accounting for GM shape<br>- Measuring SC and GM CSA | - Detect origin of microstructural damage before nonreversible changes (e.g., T2w hyperintensities, appearance of clinical symptoms)<br>- High in-plane axial resolution (good for atlas-based analysis of various WM tracts)<br>- The combined echoes provide high WM/GM contrast (depending on parameters)<br>- Fast<br>- Low SAR (except for the MT sequence) | |

region of most interest is orthogonal to the cord (leaving other regions with larger partial volume effects). Alternatively, if time allows, slices may be separated into several pseudocontiguous slabs, each orthogonal to the cord and containing three to five slices. Note that using thinner slices mitigates the partial volume effect, although this comes at the cost of lower SNR. Thinner slices also mitigate intravoxel dephasing due to inhomogeneities in the static magnetic field, which lead to signal dropout on GRE imaging[69]. Axial acquisitions with thick slices are also recommended for measuring GM CSA.

**Phase-encoding direction**

There are a few considerations to be made when choosing the phase-encoding direction. For transverse (perpendicular-to-the cord) image orientation, one advantage of R-L phase encoding is

**Table 2 | Sequences included in the spine generic protocol**

| | GE | Philips | Siemens |
|---|---|---|---|
| T1w (3D) | BRAVO/IR-FSPGR | TITFE | MPRAGE |
| T2w (3D) | CUBE | VISTA | SPACE |
| DWI (2D) | Spin Echo EPI with or without FOCUS* | Zoom Diffusion* | ep2d_diff with or without ZOOMit* |
| ME-GRE (2D) | MERGE | mFFE | GRE 'medic' |
| GRE-MT1/MT0/T1w (3D) | SPGR | FFE | GRE |

All sequences come by default with the MRI system, except those marked with an asterisk (*), which require a special license. Note that in the Philips system the MT1 and MT0 scans are acquired within the same sequence and the MTR is automatically calculated. ME, multi-echo.

that the SC is less curved along this axis, allowing for a smaller field of view (FOV) (only if using outer-volume suppression technique) and thus fewer *k*-space lines, yielding faster acquisition times in single-line readout schemes and fewer distortions on EPI. R-L phase encoding also allows for greater robustness in the presence of poor fat suppression (due to the fat in the posterior neck region) and less ghosting due to swallowing and pulsatile vessels. Alternatively, when using EPI, anterior–posterior (A-P) phase encoding will not create an R-L asymmetry, which could be problematic in some study designs where the R-L symmetry of the cord is part of the underlying study hypotheses (e.g., comparing diffusion metrics between the left and right corticospinal tract). A-P phase encoding is also less prone to peripheral nerve stimulation (although this also depends on the manufacturer, and how oblique the slices are).

**Thoracolumbar cord**

While the present protocol is optimized and validated for the cervical cord, most of the sequences proposed here could be ported to the thoracolumbar region with minimal or no adjustments. The amount of modification required mostly depends on the radiofrequency (RF) receive coil that is available. One notable advantage of the cervical region is the possibility of having coil elements around the neck, which provides better performance for accelerated acquisitions (GRAPPA, SENSE) and higher SNR. When imaging the lower cord, coil elements are typically arranged in a flat fashion, reducing acceleration and SNR. Hence, sequences already suffering from low SNR might need modifications, e.g., a larger voxel size.

In general, the T1w, T2w and MT sequences could likely be applied to the lower cord without modifications. The DWI protocol may require additional averaging and/or larger in-plane voxels to increase the SNR. Furthermore, using saturation bands for inner FOV DWI acquisitions may be much more challenging or even impossible owing to specific absorption rate (SAR) and saturation band thickness limits. The multi-echo (ME)-GRE sequence is feasible[70] but may require additional averaging[71], and/or the use of navigator echoes to compensate for respiration-related ghosting. Protocol optimization could be aided by the use of advanced SC phantoms made of 'tissue-like' materials that mimic respiration-related dynamic changes in the $B_0$ field, such as the one proposed by De Tillieux et al.[72].

**Other field strengths**

While the *spine generic* protocol was optimized and validated at 3T, only slight modifications would be required to adapt the protocol to 1.5T systems. Depending on what researchers would like to do (CSA measurements, lesion quantification, etc.), the SNR and contrast-to-noise ratio would need to be adjusted by finding the right tradeoff between spatial resolution and acquisition time. Relaxation parameters also change at lower and higher fields. For example, tissue T1 is shorter at 1.5T, which could help reduce TR in T1w sequences. Fortunately, SAR is also lower at 1.5T, which allows one to reduce the TR in SAR-intensive sequences, such as the MT protocol or the T2w sequence (including the DWI sequence). Another advantage of 1.5T is that susceptibility distortions on DWI EPI data are reduced.

At 7T, parameters would likely require greater changes than those needed to adapt to 1.5T. While SNR is higher at 7T, allowing one to reduce the voxel size, susceptibility effects are also increased. This is particularly problematic for the EPI-based DWI protocol (increased image distortions)[73] and

the long TE GRE sequences used for the T2* protocols[74]. Additionally, SAR is higher at 7T, which leads to challenges when using SAR-intensive sequences such as the MT protocol, the DWI sequence or the T2w sequence.

More challenges exist when moving between field strengths, including B1+ effects, dynamic $B_0$ changes, changes in T1, T2 and T2*, local versus body RF transmit coils, and different safety profiles. Further investigations are therefore needed to properly adapt the spine-generic protocol to other field strengths.

### Future directions

The *spine generic* acquisition protocol is a major milestone for the SC qMRI community. It provides a starting point for researchers and clinicians implementing new SC imaging initiatives. We would like to stress that the protocol will evolve with new MR hardware and software releases, as well as with research advances such as protocol optimizations and novel pulse sequence developments. Moreover, in future releases, the protocol will also be available for other manufacturers (e.g., Canon). For this reason, we suggest that researchers using and publishing with this protocol always refer to its release number (https://github.com/spine-generic/protocols/releases). The SC MRI community has initiated a forum (https://forum.spinalcordmri.org/) to encourage discussions about the generic protocol, how to use it, and how we could further improve it.

In the Supplementary Information, we discuss alternative techniques to those included in the main procedure (such as advanced shimming, navigator echoes, B1+ mapping, phase-sensitive inversion recovery, reconstruction, interpolation and filters), some of which are still at the research stage but could eventually be added to the protocol. In addition, we discuss additional equipment that can be used to immobilize the subject, including cervical collars and custom tight-fitting helmets.

We would like to reiterate that the *spine generic* protocol is not geared towards a specific disease. Researchers are at liberty to tune the proposed protocol by modifying parameters and/or adding/removing sequences as needed. A recent example is the development of a standardized brain and SC MRI protocol for patients with MS[75].

The present study also comes with two publicly available datasets (single- and multisubject)[32]. To the best of our knowledge, these are the first 'large-scale' multicenter qMRI SC datasets ever acquired and made public. The multisubject dataset could be used to create normative qMRI values, serving as age-matched healthy control references. More generally, these datasets could be used for developing new image processing tools dedicated to the SC, and the fact that they are publically available makes it possible for researchers to compare tools with the same data.

At a time when reproducibility of scientific results is a major concern[76], the proposed consensus acquisition protocol, along with publicly shared datasets and transparent analysis pipeline, aims to provide a basis for research reproducibility and study harmonization.

## Materials

### Equipment

- MRI scanner: a whole-body GE, Philips or Siemens 3T MRI scanner.
- Coils: image quality is largely affected by the receive coil. While most 1.5T and 3T systems use the integrated body coil for RF transmission to ensure adequate homogeneity, also referred to as the B1+ profile, reception can be done with various other coils, each having specific performance characteristics in terms of their sensitivity profile, which defines SNR, and g-factor, which describes the parallel imaging capability; i.e., how much one can accelerate (in the phase-encode and slice-select directions)[9]. The receive coils recommended for specific parts of the spine are listed in Table 3.
- Sequences: the required sequences are illustrated in Fig. 2, and manufacturer-specific sequence names are listed in Table 2. All the recommended sequences are available as a product; however, old software versions might not have all up-to-date product sequences, and there may be research sequences that are equivalent. When applicable, this information is mentioned within this manuscript. The protocols (pdf + import files) are freely available at https://github.com/spine-generic/protocols

**Table 3 | Recommended receive coils for SC imaging for the GE, Philips and Siemens systems**

| | Cervical | Thoracic/lumbar |
|---|---|---|
| **GE** | | |
| HD/HDxt | 8-channel cervical thoracic lumbar array | 8-channel cervical thoracic lumbar array |
| PETMR | 19-channel head neck unit array | 14-channel central molecular imaging array |
| MR750w | 16-channel head neck spine array | 48-channel geometry embracing method phased array |
| **Philips** | | |
| Achieva | 16-channel head/neck/neurovascular or 32-channel head coil | 15-channel posterior spine[b] |
| Ingenia[a] | 16-channel head/neck/neurovascular or 32-channel head coil | 12-channel posterior array[b] |
| **Siemens** | | |
| Trio | 12-channel brain + 4-channel neck array + spine array[b] | Spine array[b] |
| Verio | 12-channel brain + 4-channel neck array + spine array[b] | Spine array[b] |
| Skyra | 64-channel head/neck or 20-channel head/neck + spine array[b] | Spine array[b] |
| Prisma | 64-channel head/neck or 20-channel head/neck + spine array[b] | Spine array[b] |
| Vida | 64-channel head/neck or 20-channel head/neck + spine array[b] | Spine array[b] |

[a]A posterior spine coil could also be used, depending on coverage. For thoracic/lumbar SC imaging, an anterior coil could be used to improve image quality in sequences with anterior–posterior phase encoding. [b]The relevant elements of the spine array are selected depending on the region to cover. When using 'auto select' (Siemens) or 'SmartSelect' (Philips), elements will be automatically selected based on the slice positioning. It is advised to use it. The information in this table is subject to change with the evolution of the market.

## Procedure

### Equipment setup ● Timing 2 min

**Install coil**

1 Select the coil depending on your manufacturer and application (Table 3).
? TROUBLESHOOTING

### Subject and equipment preparation ● Timing 5–10 min

**Positioning and immobilization strategies**

2 Carefully position the subject to optimize image quality. Try to have the cervical SC as straight as possible, so that axial slices are orthogonal to the SC centerline. This minimizes partial volume effects with the surrounding cerebrospinal fluid (CSF). Reducing neck curvature also helps to improve field homogeneity because the shim volume (i.e., the 3D box centered over the region of interest where the MR system computes the optimal shim coefficients) is less likely to contain air–tissue interfaces. To minimize cervical lordosis, ask the subject to tilt their head slightly towards their chest. Placing some cushions below the head can help, as illustrated in Fig. 3. However, subjects should not be too uncomfortable and still be able to swallow in a way that minimizes motion. For thoracolumbar acquisitions, leg support helps minimize lumbar lordosis and provides more comfort for the subject.

3 Verify that the subject is aligned in the left–right direction, and ensure alignment of the spine with the sagittal plane whenever possible.
? TROUBLESHOOTING

4 Pad/clamp the subject's head tightly with cushions to avoid head motion. Note that, while doing this has the merit of not requiring additional purchases (e.g., cervical collar[77] or specialized immobilization apparatus), this setup is not easily reproducible and depends on the MR technician. It also does not ensure that subjects are always positioned in the same way for longitudinal experiments. Therefore, it is important that researchers specify the type of cushions used and, ideally, take a photograph showing how to position those cushions while the subject is in the coil.

5 Tell the subject that their neck/spine will be imaged and that if they move, image quality may be severely compromised. Mimic how *not* to swallow by exaggerating head and swallowing motions. Asking subjects not to swallow at all can sometimes lead to more motion due to the swallowing reflex that is triggered once a large volume of saliva is accumulated. This can also pose a choking risk, given that subjects are in a supine position. As a compromise, notify the subject when they can swallow between scans. Ask the subject to breathe normally and to avoid taking deep breaths. Breathing pattern affects image quality owing to the dynamic $B_0$ variations[78] that result from respiration. The latter can cause ghosting on GRE data and pixel displacement on EPI sequences.
? TROUBLESHOOTING

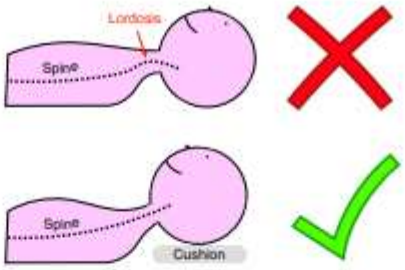



**Fig. 3 | Patient positioning.** Suggested subject positioning: use a cushion to minimize cervical lordosis (bottom panel).

**Pulse oximeter**

6 Install the pulse oximeter on one of the participant's fingers. The pulse oximeter will monitor the cardiac pulse, which will be used for cardiac gating on the DWI scan.

**Positioning the isocenter (laser marking)**

7 For thoracic/lumbar applications, set isocenter (laser) around the region of interest. If you are doing brain and cervical cord imaging, mark the isocenter right below the nose. This will ensure that the localizer will cover the desired region. Note that, for all other sequences, the table will move so that the center of the FOV is acquired at the scanner's isocenter (to ensure maximal gradient linearity).

## Image acquisition ● Timing 20–30 min

▲ CRITICAL STEP Before starting the acquisition, make sure the coil elements are properly selected. If you are using a coil that corresponds to the saved protocol (Table 3), the correct elements should be automatically selected. If you are not using a default coil, or if you are acquiring in the thoracolumbar region, then you will need to select the elements corresponding to the FOV. For some manufacturers and platforms, the elements will be automatically selected depending on the location and size of the FOV (mode 'auto select' or 'SmartSelect'), but regardless, it is always important to double check.

▲ CRITICAL STEP It is extremely important that you check each image right after its acquisition, not wait until the end of the imaging session. For example, if you notice that the wrong coil was used, fix the problem for the rest of the images (and reacquire the image if there is still time). Or if you spot excessive subject motion, talk to the subject before acquiring the next image.

**T1w scan**

8 Adjust the FOV so that it includes the whole head, as shown in Fig. 4.

9 (Optional) For GE users only: to have the images reconstructed at the proper matrix size, click on 'Save Rx' → 'Scan', then click on 'Research' → 'Download'. Then click on 'Research' → 'Display CVs'. Then, modify the following control variables (CVs) accordingly: rhimsize = 320, rhrcxres = 320, rhrcyres = 256. You can check on the console if the field was modified appropriately, by looking at the 'image header', after reconstruction. You should get: (0 × 0028, 0 × 0010) = 192; (0 × 0028, 0 × 0030) = 1\1.

10 Acquire the T1w scan. Further details of interest about the parameters used in the T1W scan can be found in Box 1.
? TROUBLESHOOTING

**T2w scan**

11 Center the FOV at C3–C4 as shown in Fig. 5. Align along the spine (see coronal view).

12 (Optional) For GE users only: to have the images reconstructed at the proper matrix size, click on 'Save Rx' → 'Scan', then click on 'Research' → 'Download'. Then click on 'Research' → 'Display CVs'. Then, modify the following CVs accordingly: rhimsize = 320, rhrcxres = 256, rhrcyres = 256.

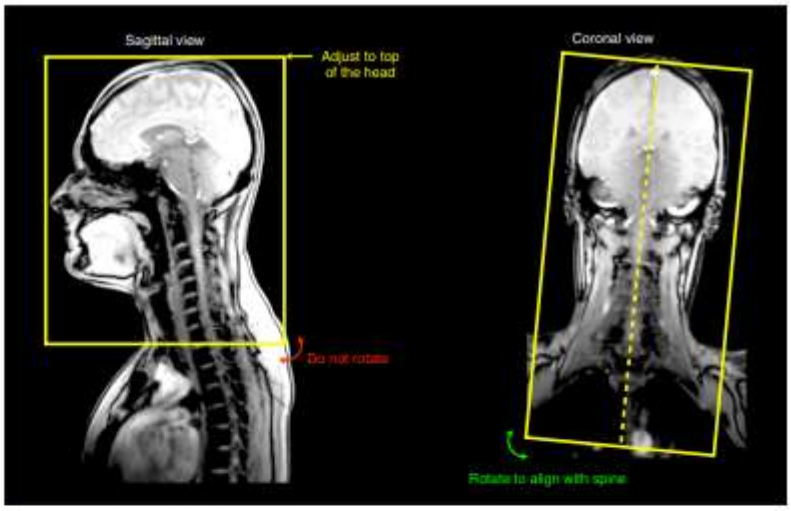



**Fig. 4 |** Positioning of FOV for T1w scans.

**Box 1 | Additional details about the parameters for T1w scan**

- **Slab-selective excitation**: to avoid R-L aliasing of the body (phase-encoding directions on this 3D sequence are R-L and S-I).
- **TR, TE, TI, flip angle**: inspired by the Human Connectome Project (HCP) protocol[11]. The TR was slightly reduced to find a compromise between satisfactory white/gray matter contrast in the brain and reducing the total acquisition time.

You can check on the console if the field was modified appropriately, by looking at the 'image header', after reconstruction. You should get: (0 × 0028, 0 × 0010) = 192; (0 × 0028, 0 × 0030) = 1\1.

13 Acquire the T2w scan. Further details of interest about the parameters used in the T2W scan can be found in Box 2.
**? TROUBLESHOOTING**

**DWI scan**

14 Use ZOOMit (Siemens), Zoom Diffusion (Philips) or FOCUS (GE), if available. Otherwise, use saturation bands for aliasing suppression (Fig. 6).

15 Center the FOV in the cord at the level of C3/C4 disc (Fig. 6). Rotate the FOV such that slices are orthogonal to the SC, in both the sagittal and coronal planes.
**▲ CRITICAL STEP** Phase-encode should be A-P.

16 Adjust the shim volume such that it covers the FOV, in both the sagittal and coronal planes (green box).

17 (Optional) For GE users only: click on 'shim volume', and then center on the SC. If you cannot modify the size of the shim box, do not worry.

18 (Optional) For GE users only: when tilting the slice, the TE might increase by a few ms. If you wish to use the same TE throughout an entire study, try tilting the FOV in the coronal and sagittal plane, and report what the minimum TE is. The more you tilt, the longer the TE will be (hence, lower SNR) but the more conservative you will be in keeping a fixed TE throughout the entire study.

19 (Optional) For GE users only: to have the images reconstructed at the proper matrix size, click on 'Save Rx' → 'Scan', then click on 'Research' → 'Download'. Then click on 'Research' → 'Display CVs'.

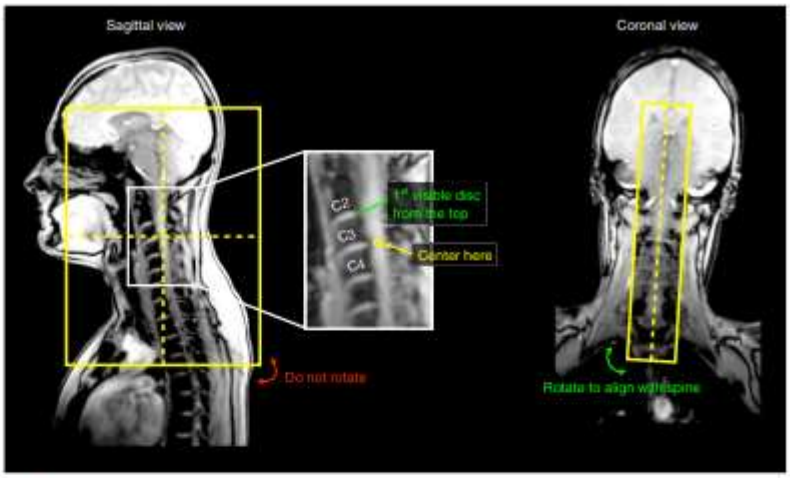



**Fig. 5 |** Positioning of FOV for T2w scans.

**Box 2 | Additional details about the parameters for T2w scan**

- **Slab-selective excitation**: to avoid R-L aliasing of the body (phase-encoding directions on this 3D sequence are R-L and S-I).
- **TR**: sufficiently high to prevent T1 recovery effects causing signal dropout in the CSF (see Fig. 3b in ref. [32]), while keeping it low to reduce total acquisition time.
- **TE, flip angle**: inspired by previous studies[31,63], optimized for satisfactory SC/CSF contrast and SAR.

Then, modify the following CVs accordingly: rhimsize = 96, rhrcxres = 86, rhrcyres = 43. You can check on the console if the field was modified appropriately, by looking at the 'image header', after reconstruction. You should get: (0 × 0028, 0 × 0010) = 192; (0 × 0028, 0 × 0030) = 1\1.

▲ CRITICAL Before starting the acquisition, make sure the PulseOx trigger is working (see Fig. 7 for an example).

20 Acquire the DWI scan. Further details of interest about the parameters used in the DWI scan can be found in Box 3.

? TROUBLESHOOTING

**GRE-MT1/MT0/T1w scans**

21 Make sure that the FOV center and orientation are the same as for the DWI scan. Normally, if you imported the full protocol, the FOV should be copied automatically from the DWI scan. If not, use 'copy parameters' (center of FOV and orientation). Use 'auto' mode for shimming.

22 (Optional) For GE users only: to avoid confusion with regard to the slice orientation, the protocol is saved as 'axial'. Please click on 'oblique' to be able to rotate the slice in the sagittal and coronal planes.

23 (Optional) For GE users only: to match the RF frequency of other manufacturers, modify the CV `off_rfmt`.

24 (Optional) For GE users only: to have the images reconstructed at the proper matrix size, click on 'Save Rx' → 'Scan', then click on 'Research' → 'Download'. Then click on 'Research' → 'Display CVs'. Then, modify the following CVs accordingly: rhimsize = 192, rhrcxres = 172, rhrcyres = 172. You can check on the console if the field was modified appropriately, by looking at the 'image header', after reconstruction. You should get: (0 × 0028, 0 × 0010) = 192; (0 × 0028, 0 × 0030) = 1\1.

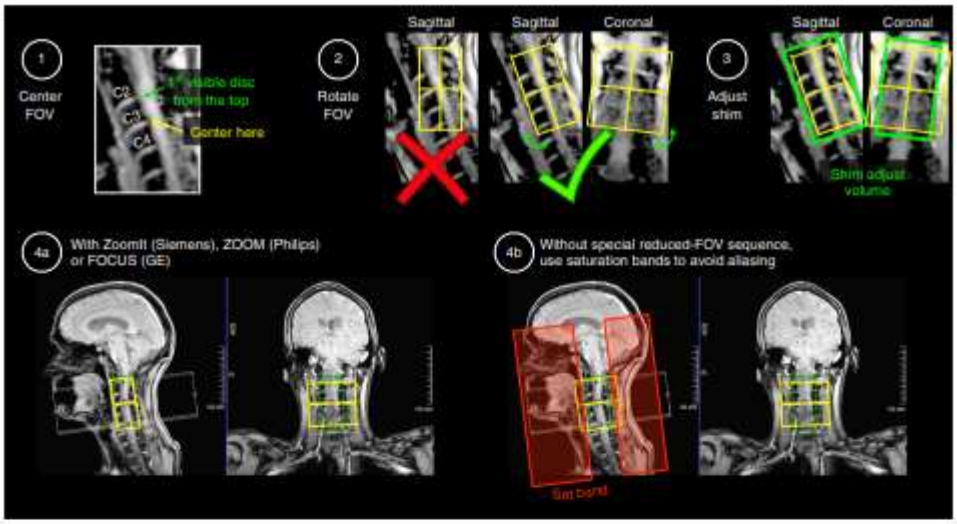



**Fig. 6 |** Positioning of FOV, shim box and saturation bands for the DWI scan.

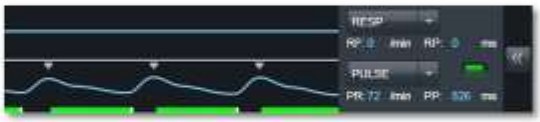



**Fig. 7 | Checking a pulse oximeter trace.** Example of a pulse oximeter trace on a Siemens scanner for triggered acquisition (small triangles).

▲ CRITICAL STEP If you get a SAR limitation on the MT scan, increase the TR to the minimum suggested (e.g., going from 35 ms to 36 ms). If the TR is increased, it is very important that you also change the TR on the GRE-MT0 sequence (TR should be the same on the MT1 and MT0 scans).

25 Acquire GRE-MT1/MT0/T1w scan. Further details of interest about the parameters used in the GRE-MT1/MT0/T1w scan can be found in Box 4.

? TROUBLESHOOTING

**GRE-ME scan**

26 Make sure that the FOV center and orientation are the same as for the DWI scan. Normally, if you imported the full protocol, the FOV should be copied automatically from the DWI scan. If not, please do 'copy parameters' (center of FOV and orientation).

27 Adjust the shim box so that it follows the spine as closely as possible (Fig. 8).

28 (Optional) For GE users only: to avoid confusion with regard to the slice orientation, save the protocol as 'axial'. Click on 'oblique' to rotate the slice in the sagittal and coronal planes.

29 (Optional) For GE users only: to have the images reconstructed at the proper matrix size, click on 'Save Rx' → 'Scan', then click on 'Research' → 'Download'. Then click on 'Research' → 'Display CVs'. Then, modify the following CVs accordingly: rhimsize = 448, rhrcxres = 224, rhrcyres = 224.

30 Acquire the GRE-ME scan. Further details of interest about the parameters used in the GRE-ME scan can be found in Box 5.

? TROUBLESHOOTING

**Box 3 | Additional details about the parameters for DWI scan**

- **2D axial**: for spinal tract-based metric quantifications (see section *Slice orientation*).
- **Spatial resolution**: 0.9 × 0.9 × 5 mm (no interpolation).
- **FOV**: reducing the FOV along the phase-encoding direction for EPI reduces susceptibility-related distortions. SC geometry is favorable to such acquisition techniques. There are three main techniques for reducing the FOV:
  - 2DRF excitation: FOCUS[84] (GE, version DV25 and onwards), ZOOMit[85] (Siemens). Paid licence. Not available on all versions.
  - Cross-sectional RF excitations 'Zoom Diffusion'[86,87] (Philips): available for software versions 5.3 and higher. For older software, it is possible to use a research sequence that is equivalent to the product sequence.
  - Outer volume suppression with sat bands (available for all manufacturers). Note: because the saturation band approach is not perfect (fat can be missed and aliased on top of the SC, as exemplified in Fig. 3f in ref. [32]), it is strongly recommended that it be used only if other options are not available.

  For more details about the pros/cons of each reduced FOV technique for the SC, the reader is referred to ref. [88].
- **Contiguous slices**
  - Pros: can do tractography, greater statistical power.
  - Cons: reduced coverage (more slices can be added at the discretion of the researcher).
- **b-Value**: single-shell at 800 s/mm$^2$
  - A single-shell protocol was chosen owing to time constraints. For more advanced diffusion models (e.g., NODDI, DBSI), additional shells can be added, potentially increasing TE/TR as the b-value is increased.
  - If the SNR is too low (the SC is barely visible on DWI), the b-value could be lowered (e.g., 600–700 s/mm$^2$), thereby reducing TE and increasing the SNR. Note that increasing the number of repetitions is not helpful because, in this low-SNR regime, the noise distribution is more Rician-like; hence, averaging of the magnitude signal results in an upward bias. On the other hand, if the SNR is sufficient, on scanners with strong gradients (≥80 mT/m), it is recommended to set an additional b-value shell (2,000–3,000 s/mm$^2$ or higher) to provide better sensitivity to diffusion-based contrast (to see demyelination/degeneration) and enable the use of more advanced diffusion models. For any multishell acquisition, using a fixed TE across shells is recommended.
  - Users should note that a low b-value (800 s/mm$^2$) may not detect complex fiber geometry for tractography applications (e.g., crossing, fanning). The b-value and number of diffusion directions can be modified to fit researchers' needs.
- **Number of directions**: 30–32 uniformly distributed. The exact diffusion gradient scheme is manufacturer-specific. The DWI protocol also includes five b = 0 images acquired at the beginning or interspersed (this is possible by editing 'DiffusionVectors.txt' on Siemens, 'dti_vectors_input.txt' on Philips, or 'tensor.dat' on GE, but probably only sensible for research sites). All b = 0 should have the same TE as the DWI data.
- **TR and cardiac gating**: it is recommended to acquire EPI data during the quiescent phase of the cardiac-related SC motion[89]. The quiescent phase of the SC lasts for ~500 ms within a cardiac cycle[89]. On the Siemens platform, we added a concatenation to break down the volume acquisition and only acquire approximately three slices during the quiescent phase of the cord. On the Philips platform, one or two slices are acquired per beat (depending on the heart rate). We suggest using a pulse oximeter instead of an ECG (it has adequate precision and is less cumbersome to use). The trigger delay is subject-dependent[89], and its definition depends on the platform and sequence parameters, making it difficult to provide an optimal number that fits all subjects and platforms. For example, the SPAIR fat saturation on the Siemens platform adds ~83 ms before each excitation, while on the Philips platform the minimum delay is about 110 ms (used by the SPIR fat sat pulse and the outer volume suppression pulses). For convenience, we set the delay to the minimum value, but researchers can optimize this value accordingly.
- **TE**: should be minimum (to maximize SNR).
- **Echo spacing** (controlled with the bandwidth): should be minimum (to minimize distortions).
- **Diffusion gradient mode**: monopolar (if available) to ensure lower TE. If needed, bipolar mode can be used to minimize eddy current-related distortions. On older Siemens platforms (e.g., VB17), the ep2d_diff product sequence is bipolar; however, research or work-in-progress sequences exist, such as WIP511 (monopolar option, with polarity alternation, freq stab and skewed fat sat).
- **Dynamic stabilization**: frequency stabilization (Siemens), dynamic stabilization (Philips), real-time field adjustment (GE): these options help reduce artifacts related to changes in the magnetic field over time.
- **Phase encoding**: P-A instead of A-P so that (i) susceptibility distortions have the effect of 'stretching' instead of 'compressing' the SC (no information loss), and (ii) in case of poor fat saturation, posterior neck fat will be aliased outside of the FOV.
- **Shimming**: box (+ advanced shimming) should be carefully positioned around the SC. See Fig. 7.
- **Acceleration** (GRAPPA/SENSE): We recommend no in-plane acceleration because a small matrix size (96 × 96) combined with a reduced FOV (~60% reduction) and partial Fourier (7/8) leaves too few phase-encoding lines for reliable image reconstruction. Besides, in-plane acceleration reduces SNR by factor sqrt(R), where R is the in-plane acceleration factor.
- **Simultaneous multislice**: while this technique is gaining popularity for reducing the overall acquisition time by exciting several slices at the same time, thereby reducing the volume TR, we do not suggest its use here because the number of slices (n = 15) and their gap is small, which hampers the acceleration performance; further reducing the TR would hamper longitudinal relaxation.
- **Fat suppression**: 2DRF protocols use water excitation, while the Philips Zoom sequence employs a SPIR fat saturation pulse to minimize fat contribution. If insufficient, skewed techniques[90] could be tried.
- **Partial Fourier**: 75% *k*-space to mitigate phase errors, while still being able to reduce TE. A larger *k*-space window also minimizes the risk of a total dropout that can happen when the peak of an echo moves entirely outside the readout window.
- **Fieldmap** (blip-up/down): given the difficulty of acquiring a robust fieldmap and correcting for susceptibility-related distortions using a blip-up/down sequence in the SC (partly because the manifestation of artifacts in the up/down directions might be slightly different, due to, e.g., CSF presence, $B_0$ field differences between up/down because of respiratory-related $B_0$ variation), we do not recommend that these be acquired. Any type of correction might introduce more artifacts if misused. Instead, we suggest acquiring data with minimal distortions in the first place (by minimizing the echo spacing, optimizing shimming, etc.) and correcting residual distortions by registering the DWI data on to a structural scan (e.g., b = 0 on the T2 space). For more details, see this forum post: http://forum.spinalcordmri.org/t/how-to-correct-for-distortions-in-spinal-cord-diffusion-mri-data/326

**Box 4 | Additional details about the parameters for GRE-MT1 / MT0 / T1w Scans**

- **2D versus 3D**: 3D is recommended because with some manufacturers (including Siemens) the MT pulse is not selective, so using a 2D sequence will result in a different MT effect across slices.
- **TR and flip angle**: optimized so as to minimize the standard deviation of the MTR, given the limitations on different systems (e.g., SAR, TR, etc.).
- **TE**: minimized to reduce T2* relaxation.
- **Saturation bands**: not recommended because the offset excitation creates an MT effect, which could vary across manufacturers (different implementation).
- **Fat saturation pulses**: should not be used with a quantitative MT protocol (MTR, MTsat, qMT, MTCSF, etc.) because the off-resonance pulse will create an additional MT effect.
- **Water excitation** (using composite binomial pulses): not recommended since the effect on the off-resonance fat signal will impact the MT effect.
- **MT frequency offset**: using product sequences, the frequency offset cannot be changed. Siemens and GE use 1.2 kHz, while Philips uses 1.1 kHz, which is the recommendation for the spine generic protocol. If, however, one has access to research sequences that allow changing that parameter, increasing the frequency offset (e.g., to 4 kHz) will result in higher WM/GM contrast that could be advantageous when segmenting the GM. The RF strength and pulse pattern for the MT pulse cannot be disclosed here because it is proprietary information of the manufacturers.
- **Multiecho combined**: although combining echoes provides higher SNR, we do not always recommend it because signal dropout at later echo times could bias MT metrics. Also, on GE systems, it is not possible to use multiecho with the MT pulse.

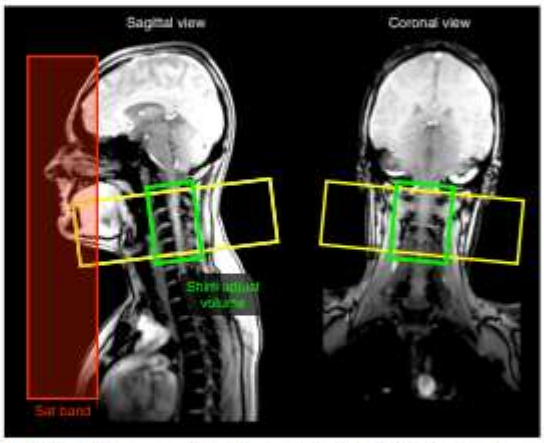



**Fig. 8 | Positioning of the FOV, shim box and saturation bands for the GRE-ME scan.** Siemens and GE users: the saturation band is already automatically positioned. Philips users: the saturation bands are 'invisible' on this sequence, but they are nevertheless applied.

**Box 5 | Additional details about the parameters for GRE-ME scan**

- **Optimization**: the chosen parameters for this sequence result from a consensus that arose from the 'Gray Matter Acquisition Challenge', which was organized during the 5th Spinal Cord MRI workshop (http://www.spinalcordmri.org/2018/06/22/workshop.html).
- **2D versus 3D**: while 3D acquisitions are more SNR efficient, we recommend using 2D acquisitions as they produce 'cleaner' images: no aliasing along the second phase encoding direction when using 3D, more homogeneous B1+ profile than 3D acquisition, less sensitive to motion.
- **Spatial resolution**: 0.5 × 0.5 × 5 mm (no interpolation).
- **Saturation band**: adds a slight MT effect due to the off-resonance pulse, which has the effect of slightly increasing white/gray matter contrast. A corollary benefit of this saturation band, positioned coronal and anteriorly (Fig. 8), is that it also removes signal from a region prone to motion (swallowing and vessel pulsatility in the neck).
- **Monopolar versus bipolar**: this concerns the filling of *k*-space across the different echoes. It is more time-efficient to fill the *k*-space by alternating polarities across echoes; however, this leads to a slight inter-echo shift caused by field inhomogeneities. For this reason, we recommend using monopolar encoding, with the downside of slightly longer TE and TR.
- **Multiecho combined**: if individual echo images are available, they should be combined during postprocessing for more transparency on the aggregation method. MEDIC (Siemens) and MERGE (GE) automatically combine all echoes. The Philips mFFE sequence outputs all the echoes with the option to also output an 'accumulated' image, which corresponds to the sum of all echoes. Depending on the version, the MEDIC sequence does not feature the phase stabilization option (navigator-based phase correction that minimizes ghosting), whereas the FLASH does (this depends on the versio; e.g., VE11C does not).

## Troubleshooting

Troubleshooting advice can be found in Table 4.

**Table 4 | Troubleshooting table**

| Step | Problem | Possible reason | solution |
|---|---|---|---|
| 1 | Insufficient signal in some parts of the image (Supplementary Fig. 1) | Wrong coil selection: for each region in the FOV, the proper coil needs to be selected (Table 3). For example, if you are planning to cover the head and neck region, then the head/neck coil should be used. In addition, for each sequence, the proper coil elements need to be selected | If you notice such artifacts in the image, make sure to check those coil parameters |
| 2-5 | Blurry images (Supplementary Figs. 2 and 3) | Improper subject positioning: proper subject positioning is important both for the subject's comfort (which has an indirect positive impact on image quality) and for the reduction of some artifacts. For example, excessive lordosis can create more pronounced CSF flow and SC motion | Reposition the subject |
| | | Subject motion: subject motion can negatively affect all sequences, with some sequences being particularly sensitive: T1w (Step 10), GRE-MT0/MT1/T1w (Step 25) and GRE-ME (Step 30) | Talk to the subject; ask them to not move for the rest of the imaging session. If there is still time before the end of the imaging session, it is recommended to reacquire the problematic image(s) |
| 3 | Misaligned images (Supplementary Fig. 4) | Improper subject/FOV positioning: if the medial plane of the spine and head is not aligned with the MRI bore, it could lead to subject discomfort and image misalignment when the FOV is not properly rotated about the anteroposterior axis | Reposition the subject |
| 10 (T1w) and 13 (T2w) | Artifact: multiple fine parallel lines adjacent to high-contrast interfaces | Gibbs ringing | Can be reduced with an apodization filter, e.g., 'Raw filter' (Siemens), 'Image filter' (Philips) |
| 13 (T2w) | Signal drops in the CSF (Supplementary Fig. 5) | Incorrect acquisition parameter: changes to the acquisition parameters can result in undesired artifacts. In the example shown in Supplementary Fig. 5, the flip angle was increased, causing the CSF signal to not recover fully (hypointense signal). Other changes in sequence parameters (TR, TE, matrix size, etc.) could also lead to undesired artifacts or biases in the computed qMRI metrics | Keep the protocol parameters as close as possible to what is prescribed in the spine generic protocol |
| 8, 11, 15, 21 and 26 | Variable coverage across the studied population | Wrong FOV placement: it is important to follow the prescribed FOV placement, as failure to do so could result in variable coverage across the studied population, and be a source of inconsistencies and biases. Supplementary Fig. 6 shows an example of wrong FOV placement for a GRE-MT scan | Keep the FOV as close as possible to what is prescribed in the spine generic protocol |
| 20 (DWI) | Artifacts on DWI scans (Supplementary Fig. 7) | The DWI scan is based on an EPI sequence, which is prone to susceptibility artifacts manifesting as image distortions. Other effects can lead to artifacts when using this sequence, including poor fat saturation and excessive subject or pulsatile motion | Poor fat saturation: this can cause fat to overlay on the SC (see Fig. 3f in ref. [52]); the cause is likely related to poor shimming. In this case, try to move the table, re-shim and/or try other fat saturation methods (e.g., frequency-selective, inversion-recovery). If saturation bands are used, look for poorly saturated signals in the saturation band region by prescribing full FOV and looking at the area where saturation bands are located. If the signal outside the FOV is too high, causing it to alias over the SC, try to increase the number of phase encode lines (this will cause slightly more distortions) or to unselect coil elements if they are not necessary (e.g., switching off the anterior neck element) |
| 25 (MT) | Blurry slice edges | The 3D excitation does not have a sharp profile at the edges | Discard two to three slices at each edge |
| 30 (GRE-ME) | Signal dropout | Signal dropout can be caused by intravoxel dephasing | If you notice substantial signal dropout, try reshimming, using thinner slices, or reducing the TE (and/or number of echoes) |

## Anticipated results

In this section, we show images of the same subject acquired across the three manufacturers. Additional examples of good-quality data with interactive 3D visualization are shown in the spine generic website (https://spine-generic.rtfd.io/en/latest/data-acquisition.html#example-of-datasets). The interactive embedding in the website is powered by Brainsprite (https://brainsprite.github.io/).

### Good-quality T1w scans (Steps 8–10)

Figure 9 illustrates what good-quality T1w scans for all three manufacturers look like. All scans are devoid of any motion artifacts, and the signal is homogeneous throughout the SC. The SC is nicely visible in the medial sagittal plane.

### Good-quality T2w scans (Steps 11–13)

In Fig. 10, we show good-quality T2w scans for all three manufacturers. All scans are devoid of any motion artifacts, and the signal is homogeneous throughout the SC. Like for the T1w scans, the SC is nicely visible in the medial sagittal plane.

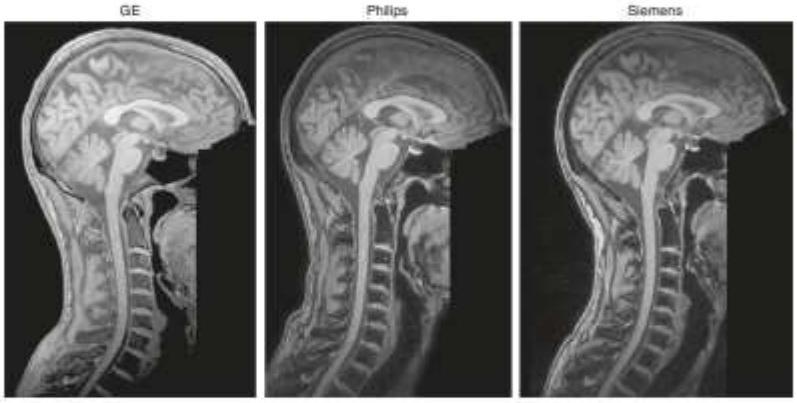



**Fig. 9 |** Sagittal views of good-quality T1w scans for each manufacturer.

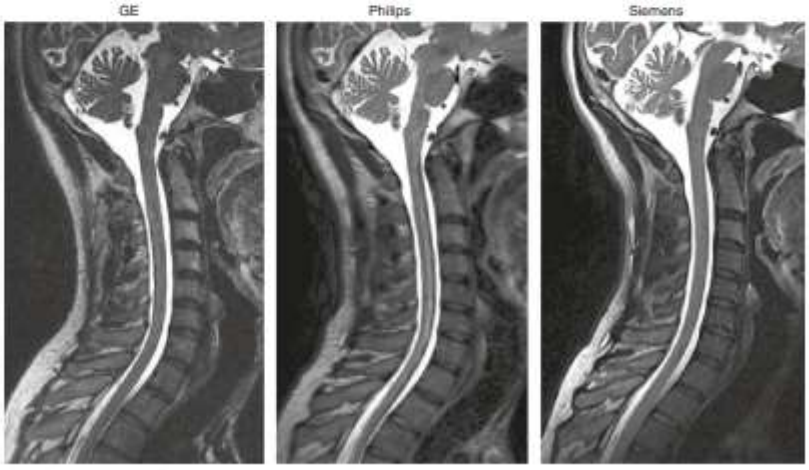



**Fig. 10 |** Sagittal views of good-quality T2w scans for each manufacturer.

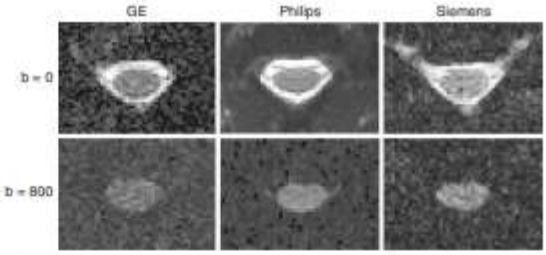



**Fig. 11 | Axial views of good quality data for DWI scans at b = 0 s/mm$^2$ (top row) and b = 800 s/mm$^2$ (bottom row).** The DW image corresponds to a diffusion gradient vector fairly orthogonal to the cord axis, hence the visible spinal cord. Notice the different noise patterns across the manufacturers, which is due to the different types of filters applied across manufacturers; these filters were present in an older version of the protocol, but have been removed in the latest version of the protocol in order to minimize differences across manufacturers.

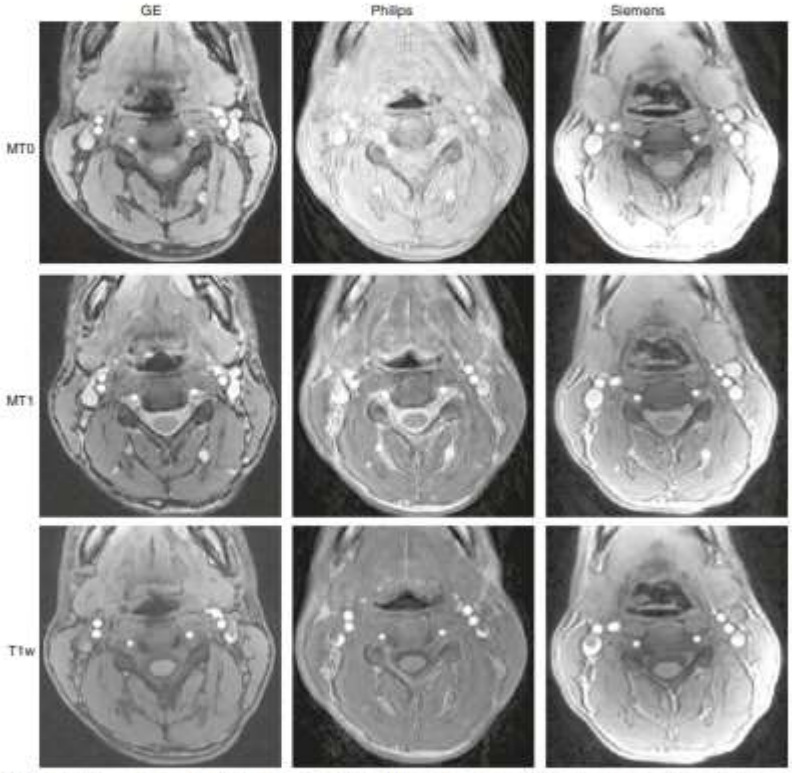



**Fig. 12 | Axial views of good-quality data for MT0, MT1 and T1w scans.** Notice the slight motion artifact on the Philips MT0 scan. Also notice the strong signal intensity at the periphery of the tissue on the Siemens scans, which is due to the inactivation of the intensity bias filter. This filter is not relevant when computing qMRI metrics such as MTR or MTsat.

### Good-quality DWI scans (Steps 14–20)

In Fig. 11, we show good-quality DW scans for all three manufacturers. These DW images correspond to a diffusion gradient vector fairly orthogonal to the cord axis, hence the visible SC. When the diffusion gradient is oriented quasi-parallel to the cord, the signal in the cord almost vanishes. Notice the different noise patterns across the manufacturers, which is due to the different types of filters applied. These filters were present in the old version of the protocol but removed in the latest version.

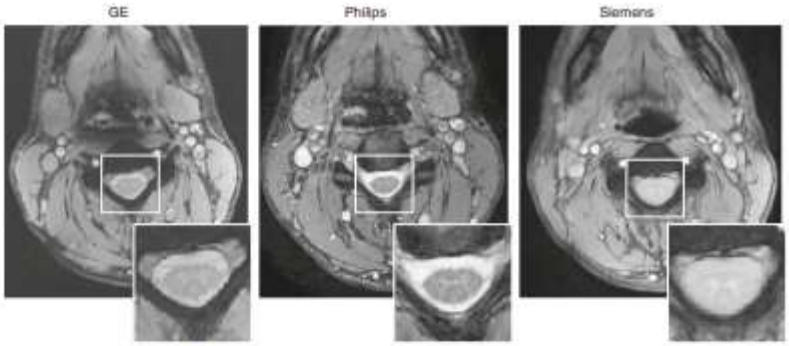



**Fig. 13 |** Axial views of good-quality ME-GRE scans for each manufacturer.

### Good-quality MT scans (Steps 21–25)

Figure 12 illustrates good-quality MT0, MT1 and T1w scans for all three manufacturers. Notice the slight motion artifact on the Philips MT0 scan. Also notice the strong signal intensity at the periphery of the tissue on the Siemens scans, which is due to the inactivation of the intensity bias filter. This filter is not relevant when computing qMRI metrics such as MTR or MTsat.

### Good-quality ME-GRE scans (Steps 26–30)

In Fig. 13, we show good-quality ME-GRE scans for the three manufacturers. The contrast between GM and WM is good, and there is no visible ghosting or signal dropout.

## References

1. Cercignani, M., Dowell, N. G. & Tofts, P. S. *Quantitative MRI of the Brain: Principles of Physical Measurement* 2nd edn (CRC Press, 2018).
2. Cohen-Adad, J. & Wheeler-Kingshott, C. (eds). *Quantitative MRI of the Spinal Cord* (Academic Press, 2014).
3. Wheeler-Kingshott, C. A. et al. The current state-of-the-art of spinal cord imaging: applications. *Neuroimage* **84**, 1082–1093 (2014).
4. Stroman, P. W. et al. The current state-of-the-art of spinal cord imaging: methods. *Neuroimage* **84**, 1070–1081 (2014).
5. Cohen-Adad, J. & Wald, L. L. Array coils. in *Quantitative MRI of the Spinal Cord* (Cohen-Adad, J. & Wheeler-Kingshott, C. A. M. eds) 59–67 (Academic Press, 2014).
6. Barry, R. L., Vannesjo, S. J., By, S., Gore, J. C. & Smith, S. A. Spinal cord MRI at 7T. *Neuroimage* **168**, 437–451 (2018).
7. Saritas, E. U., Holdsworth, S. J. & Bammer, R. Susceptibility artifacts. in *Quantitative MRI of the Spinal Cord* (Cohen-Adad, J. & Wheeler-Kingshott, C. A. M. eds) 91–104 (Academic Press, 2014).
8. Bonati, U. et al. Cervical cord and brain grey matter atrophy independently associate with long-term MS disability. *J. Neurol. Neurosurg. Psychiatry* **82**, 471–472 (2011).
9. Cohen, A. B. et al. The relationships among MRI-defined spinal cord involvement, brain involvement, and disability in multiple sclerosis. *J. Neuroimaging* **22**, 122–128 (2012).
10. Kearney, H. et al. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. *Mult. Scler.* **20**, 72–80 (2014).
11. Lukas, C. et al. Relevance of spinal cord abnormalities to clinical disability in multiple sclerosis: MR imaging findings in a large cohort of patients. *Radiology* **269**, 542–552 (2013).
12. Branco, L. M. T. et al. Spinal cord atrophy correlates with disease duration and severity in amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **15**, 93–97 (2014).
13. El Mendili, M.-M. et al. Multi-parametric spinal cord MRI as potential progression marker in amyotrophic lateral sclerosis. *PLoS One* **9**, e95516 (2014).
14. de Albuquerque, M. et al. Longitudinal evaluation of cerebral and spinal cord damage in amyotrophic lateral sclerosis. *Neuroimage Clin.* **14**, 269–276 (2017).
15. Querin, G. et al. Spinal cord multi-parametric magnetic resonance imaging for survival prediction in amyotrophic lateral sclerosis. *Eur. J. Neurol.* **24**, 1040–1046 (2017).
16. Paquin, M.-È. et al. Spinal cord gray matter atrophy in amyotrophic lateral sclerosis. *AJNR Am. J. Neuroradiol.* **39**, 184–192 (2018).
17. van de Stadt, S. I. W. et al. Spinal cord atrophy as a measure of severity of myelopathy in adrenoleukodystrophy. *J. Inherit. Metab. Dis.* **43**, 852–860 (2020).

18. Kadanka, Z. Jr et al. Predictors of symptomatic myelopathy in degenerative cervical spinal cord compression. *Brain Behav.* **7**, e00797 (2017).
19. Seif, M. et al. Cervical cord neurodegeneration in traumatic and non-traumatic spinal cord injury. *J. Neurotrauma* **37**, 860–867 (2020).
20. De Leener, B. et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage* **145**, 24–43 (2017).
21. Rasoanandrianina, H. et al. Region-specific impairment of the cervical spinal cord (SC) in amyotrophic lateral sclerosis: a preliminary study using SC templates and quantitative MRI (diffusion tensor imaging/inhomogeneous magnetization transfer). *NMR Biomed.* **30**, e3801 (2017).
22. Martin, A. R. et al. Translating state-of-the-art spinal cord MRI techniques to clinical use: a systematic review of clinical studies utilizing DTI, MT, MWF, MRS, and fMRI. *Neuroimage Clin.* **10**, 192–238 (2016).
23. David, G. et al. Traumatic and nontraumatic spinal cord injury: pathological insights from neuroimaging. *Nat. Rev. Neurol.* **15**, 718–731 (2019).
24. Cadotte, D. W., Akbar, M. A., Fehlings, M. G., Stroman, P. W. & Cohen-Adad, J. What has been learned from magnetic resonance imaging examination of the injured human spinal cord: a Canadian perspective. *J. Neurotrauma* **35**, 1942–1957 (2018).
25. Huffnagel, I. C. et al. Longitudinal diffusion MRI as surrogate outcome measure for myelopathy in adrenoleukodystrophy. *Neurology* **93**, e2133–e2143 (2019).
26. Martin, A. R. et al. Can microstructural MRI detect subclinical tissue injury in subjects with asymptomatic cervical spinal cord compression? A prospective cohort study. *BMJ Open* **8**, e019809 (2018).
27. Labounek, R. et al. HARDI-ZOOMit protocol improves specificity to microstructural changes in presymptomatic myelopathy. *Sci. Rep.* **10**, 17529 (2020).
28. Schmierer, K., Scaravilli, F., Altmann, D. R., Barker, G. J. & Miller, D. H. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Ann. Neurol.* **56**, 407–415 (2004).
29. Fatemi, A. et al. Magnetization transfer MRI demonstrates spinal cord abnormalities in adrenomyeloneuropathy. *Neurology* **64**, 1739–1745 (2005).
30. Lema, A. et al. A comparison of magnetization transfer methods to assess brain and cervical cord microstructure in multiple sclerosis. *J. Neuroimaging* **27**, 221–226 (2017).
31. Cohen-Adad, J. et al. Demyelination and degeneration in the injured human spinal cord detected with diffusion and magnetization transfer MRI. *Neuroimage* **55**, 1024–1033 (2011).
32. Cohen-Adad, J. et al. Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers. *Sci. Data* https://doi.org/10.1038/s41597-021-00941-8 (2021).
33. Grussu, F. et al. Relevance of time-dependence for clinically viable diffusion imaging of the spinal cord. *Magn. Reson. Med.* **81**, 1247–1264 (2019).
34. Feaster, D. J., Mikulich-Gilbertson, S. & Brincks, A. M. Modeling site effects in the design and analysis of multi-site trials. *Am. J. Drug Alcohol Abus.* **37**, 383–391 (2011).
35. Fratini, M. et al. Multiscale imaging approach for studying the central nervous system: methodology and perspective. *Front. Neurosci.* **14**, 72 (2020).
36. Grussu, F. et al. Multi-parametric quantitative in vivo spinal cord MRI with unified signal readout and image denoising. *Neuroimage* **217**, 116884 (2020).
37. Gros, C. et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage* **184**, 901–915 (2019).
38. Papinutto, N. & Henry, R. G. Evaluation of intra-and interscanner reliability of MRI protocols for spinal cord gray matter and total cross-sectional area measurements. *J. Magn. Reson. Imaging* **49**, 1078–1090 (2019).
39. Perone, C. S., Ballester, P., Barros, R. C. & Cohen-Adad, J. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. *Neuroimage* **194**, 1–11 (2019).
40. Perone, C. S., Calabrese, E. & Cohen-Adad, J. Spinal cord gray matter segmentation using deep dilated convolutions. *Sci. Rep.* **8**, 5966 (2018).
41. Lévy, S. et al. Test-retest reliability of myelin imaging in the human spinal cord: Measurement errors versus region- and aging-induced variations. *PLoS One* **13**, e0189944 (2018).
42. Gros, C. et al. Automatic spinal cord localization, robust to MRI contrasts using global curve optimization. *Med. Image Anal.* **44**, 215–227 (2018).
43. Duval, T., Smith, V., Stikov, N., Klawiter, E. C. & Cohen-Adad, J. Scan-rescan of axcaliber, macromolecular tissue volume, and g-ratio in the spinal cord. *Magn. Reson. Med.* **79**, 2759–2765 (2018).
44. De Leener, B. et al. PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. *Neuroimage* **165**, 170–179 (2018).
45. Prados, F. et al. Spinal cord grey matter segmentation challenge. *Neuroimage* **152**, 312–329 (2017).
46. De Leener, B. et al. Topologically preserving straightening of spinal cord MRI. *J. Magn. Reson. Imaging* **46**, 1209–1219 (2017).
47. Duval, T. et al. g-Ratio weighted imaging of the human spinal cord in vivo. *Neuroimage* **145**, 11–23 (2017).
48. Dupont, S. M. et al. Fully-integrated framework for the segmentation and registration of the spinal cord white and gray matter. *Neuroimage* https://doi.org/10.1016/j.neuroimage.2016.09.026 (2017).
49. Papp, D., Smith, A. K., Mariano, R. & Clare, S. High-resolution quantitative maps of magnetisation transfer, R1 and R2* of the cervical spinal cord in clinically feasible acquisition time using vendor-provided sequences. in *Proceedings of the 27th Annual Meeting of ISMRM, Montreal, Canada* 4992 (2019).

50. Vahdat, S. et al. Resting-state brain and spinal cord networks in humans are functionally integrated. *PLoS Biol.* **18**, e3000789 (2020).
51. Di Nuzzo, M. et al. Towards a standard pipeline for the analysis of human spinal cord fMRI data series. in *Proceedings of the 27th Annual Meeting of ISMRM, Montreal, Canada* (2019).
52. Moccia, M. et al. Longitudinal spinal cord atrophy in multiple sclerosis using the generalized boundary shift integral. *Ann. Neurol.* **86**, 704–713 (2019).
53. Prados, F. et al. Generalised boundary shift integral for longitudinal assessment of spinal cord atrophy. *NeuroImage* **209**, 116489 (2020).
54. Oh, J. et al. The Canadian prospective cohort (canproco) study to understand progression in multiple sclerosis: rationale, aims, and study design. in *35th Congress of the European Committee for Treatment and Research in Multiple Sclerosis, Stockholm, Sweden* P753 (2019).
55. Nestrasil, I. et al. Cervical spinal cord diffusion MRI and intraspinal space restriction at the occipito-cervical junction in mucopolysacharidoses patients. in *Proceedings of the 27th Annual Meeting of ISMRM, Montreal, Canada* (2019).
56. Querin, G. et al. Presymptomatic spinal cord pathology in c9orf72 mutation carriers: a longitudinal neuroimaging study. *Ann. Neurol.* **86**, 158–167 (2019).
57. Querin, G. et al. The spinal and cerebral profile of adult spinal-muscular atrophy: a multimodal imaging study. *Neuroimage Clin.* **21**, 101618 (2019).
58. Savini, G. et al. Pilot study on quantitative cervical cord and muscular MRI in spinal muscular atrophy: promising biomarkers of disease evolution and treatment? *Front. Neurol.* **12**, 613834 (2021).
59. Martin, A. R. et al. Monitoring for myelopathic progression with multiparametric quantitative MRI. *PLoS One* **13**, e0195733 (2018).
60. Martin, A. R. et al. A novel MRI biomarker of spinal cord white matter injury: T2*-weighted white matter to gray matter signal intensity ratio. *AJNR Am. J. Neuroradiol.* **38**, 1266–1273 (2017).
61. Martin, A. R. et al. Clinically feasible microstructural MRI to quantify cervical spinal cord tissue injury using DTI, MT, and T2*-weighted imaging: assessment of normative data and reliability. *AJNR Am. J. Neuroradiol.* **38**, 1257–1265 (2017).
62. Karbasforoushan, H., Cohen-Adad, J. & Dewald, J. P. A. Brainstem and spinal cord MRI identifies altered sensorimotor pathways post-stroke. *Nat. Commun.* **10**, 3524 (2019).
63. Seif, M., Gandini Wheeler-Kingshott, C. A., Cohen-Adad, J., Flanders, A. E. & Freund, P. Guidelines for the conduct of clinical trials in spinal cord injury: neuroimaging biomarkers. *Spinal Cord.* **57**, 717–728 (2019).
64. Bagnato, F. et al. Imaging mechanisms of disease progression in multiple sclerosis: beyond brain atrophy. *J. Neuroimaging* **30**, 251–266 (2020).
65. Tinnermann, A., Büchel, C. & Cohen-Adad, J. Cortico-spinal imaging to study pain. *Neuroimage* **224**, 117439 (2020).
66. Cohen-Adad, J. Microstructural imaging in the spinal cord and validation strategies. *Neuroimage* **182**, 169–183 (2018).
67. Wheeler-Kingshott, C. A. M. G. et al. Imaging spinal cord injury and assessing its predictive value—the INSPIRED study. in *Wings for Life Scientific Meeting, Salzburg, Austria* 29 (2017).
68. Xu, J. et al. Improved in vivo diffusion tensor imaging of human cervical spinal cord. *Neuroimage* **67**, 64–76 (2013).
69. Summers, P. E., Brooks, J. & Cohen-Adad, J. Spinal cord fMRI. in *Quantitative MRI of the Spinal Cord* (Cohen-Adad, J. & Wheeler-Kingshott, C. A. M. eds) 221–236 (Academic Press, 2014).
70. Fradet, L., Arnoux, P.-J., Ranjeva, J.-P., Petit, Y. & Callot, V. Morphometrics of the entire human spinal cord and spinal canal measured from in vivo high-resolution anatomical magnetic resonance imaging. *Spine* **39**, E262–E269 (2014).
71. Yiannakas, M. C., Kakar, P., Hoy, L. R., Miller, D. H. & Wheeler-Kingshott, C. A. M. The use of the lumbosacral enlargement as an intrinsic imaging biomarker: feasibility of grey matter and white matter cross-sectional area measurements using MRI at 3T. *PLoS One* **9**, e105544 (2014).
72. De Tillieux, P. D. et al. A pneumatic phantom for mimicking respiration-induced artifacts in spinal MRI. *Magn. Reson. Med.* **79**, 600–605 (2018).
73. Massire, A. et al. Feasibility of single-shot multi-level multi-angle diffusion tensor imaging of the human cervical spinal cord at 7T. *Magn. Reson. Med.* **80**, 947–957 (2018).
74. Massire, A. et al. High-resolution multi-parametric quantitative magnetic resonance imaging of the human cervical spinal cord at 7T. *Neuroimage* **143**, 58–69 (2016).
75. Li, D. K. B. et al. Developing a universally useful, useable and used standardized MRI protocol for patients with multiple sclerosis. in *Proceedings of the 28th Annual Meeting of ISMRM, Sydney, Australia* (2020).
76. Stikov, N., Trzasko, J. D. & Bernstein, M. A. Reproducibility and the future of MRI research. *Magn. Reson. Med.* **82**, 1981–1983 (2019).
77. Yiannakas, M. C. et al. Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: a pilot study with application to magnetisation transfer measurements. *Neuroimage* **63**, 1054–1059 (2012).
78. Verma, T. & Cohen-Adad, J. Effect of respiration on the B0 field in the human spinal cord at 3T. *Magn. Reson. Med.* **72**, 1629–1636 (2014).
79. Song, S. K. et al. Demyelination increases radial diffusivity in corpus callosum of mouse brain. *Neuroimage* **26**, 132–140 (2005).

80. Jones, D. K. & Basser, P. J. 'Squashing peanuts and smashing pumpkins': how noise distorts diffusion-weighted MR data. *Magn. Reson. Med.* **52**, 979–993 (2004).
81. Helms, G., Dathe, H., Kallenberg, K. & Dechent, P. High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. *Magn. Reson. Med.* **60**, 1396–1407 (2008).
82. Levy, S. et al. White matter atlas of the human spinal cord with estimation of partial volume effect. *Neuroimage* **119**, 262–271 (2015).
83. Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
84. Saritas, E. U., Cunningham, C. H., Lee, J. H., Han, E. T. & Nishimura, D. G. DWI of the spinal cord with reduced FOV single-shot EPI. *Magn. Reson. Med.* **60**, 468–473 (2008).
85. Finsterbusch, J. High-resolution diffusion tensor imaging with inner field-of-view EPI. *J. Magn. Reson. Imaging* **29**, 987–993 (2009).
86. Wilm, B. J. et al. Diffusion-weighted imaging of the entire spinal cord. *NMR Biomed.* **22**, 174–181 (2009).
87. Jeong, E.-K., Kim, S.-E., Guo, J., Kholmovski, E. G. & Parker, D. L. High-resolution DTI with 2D interleaved multislice reduced FOV single-shot diffusion-weighted EPI (2D ss-rFOV-DWEPI). *Magn. Reson. Med.* **54**, 1575–1579 (2005).
88. Samson, R. S. et al. ZOOM or Non-ZOOM? Assessing spinal cord diffusion tensor imaging protocols for multi-centre studies. *PLoS One* **11**, e0155557 (2016).
89. Summers, P. et al. A preliminary study of the effects of trigger timing on diffusion tensor imaging of the human. *AJNR Am. J. Neuroradiol.* **27**, 1952–1961 (2006).
90. Pfeuffer, J. et al. Zoomed functional imaging in the human brain at 7 Tesla with simultaneous high spatial and high temporal resolution. *Neuroimage* **17**, 272–286 (2002).



[1]NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, Quebec, Canada. [2]Functional Neuroimaging Unit, CRIUGM, University of Montreal, Montreal, Quebec, Canada. [3]Mila-Quebec AI Institute, Montreal, Quebec, Canada. [4]Department of Radiology, Swiss Paraplegic Centre, Nottwil, Switzerland. [5]Centre for Advanced Imaging, The University of Queensland, Brisbane, Queensland, Australia. [6]Department of Radiology, University of British Columbia, Vancouver, British Columbia, Canada. [7]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA. [8]Department of Radiology, Harvard Medical School, Boston, MA, USA. [9]Harvard-Massachusetts Institute of Technology Health Sciences & Technology, Cambridge, MA, USA. [10]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Queensland, Australia. [11]NMR Research Unit, Queen Square MS Centre, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, UK. [12]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [13]Department of Neurosurgery, Medical College of Wisconsin, Milwaukee, WI, USA. [14]Aix-Marseille Univ, CNRS, CRMBM, Marseille, France. [15]APHM, Hopital Universitaire Timone, CEMEREM, Marseille, France. [16]Vanderbilt University Institute of Imaging Science, Vanderbilt University Medical Center, Nashville, TN, USA. [17]Department of Computer and Software Engineering, Polytechnique Montreal, Montreal, Quebec, Canada. [18]CHU Sainte-Justine Research Centre, Montreal, Quebec, Canada. [19]Centre de Recherche CHUS, CIMS, Sherbrooke, Quebec, Canada. [20]Sherbrooke Connectivity Imaging Lab (SCIL), Computer Science department, Université de Sherbrooke, Sherbrooke, Quebec, Canada. [21]Université de Strasbourg, CNRS, ICube, Strasbourg, France. [22]UHB - University Hospital Brno and Masaryk University, Department of Radiology and Nuclear Medicine, Brno, Czech Republic. [23]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada. [24]Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada. [25]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. [26]Richard M. Lucas Center, Stanford University School of Medicine, Stanford, CA, USA. [27]Spinal Cord Injury Center Balgrist, University of Zurich, Zurich, Switzerland. [28]Institute of Nanotechnology, CNR, Rome, Italy. [29]IRCCS Santa Lucia Foundation, Rome, Italy. [30]Department of Radiology, Juntendo University School of Medicine, Tokyo, Japan. [31]Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy. [32]Brain MRI 3T Research Centre, IRCCS Mondino Foundation, Pavia, Italy. [33]MR Clinical Science, Philips Healthcare, Markham, Ontario, Canada. [34]CREF - Museo storico della fisica e Centro studi e ricerche Enrico Fermi, Rome, Italy. [35]Radiomics Group, Vall d'Hebron Institute of Oncology, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain. [36]Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, MN, USA. [37]Multimodal and functional imaging laboratory, Central European Institute of Technology (CEITEC), Brno, Czech Republic. [38]Department of Radiology, Toho University Omori Medical Center, Tokyo, Japan. [39]Department of Radiology, the University of Tokyo, Tokyo, Japan. [40]Interdepartmental Neuroscience Program, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. [41]Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA, USA. [42]Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, UK. [43]BioMedical Engineering and Imaging Institute (BMEII), Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [44]Institute of Bioengineering/Center for Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne, Geneva, Switzerland. [45]Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland. [46]Institute of Diagnostic and Interventional Neuroradiology, Carl Gustav Carus University Hospital, Technische Universität Dresden, Dresden, Germany. [47]Department of Medicine (Neurology), University of British Columbia, Vancouver, British Columbia, Canada. [48]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China. [49]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China. [50]Wellcome Centre For Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [51]CUBRIC, Cardiff University, Wales, UK. [52]Centre for Medical Image Computing (CMIC), Medical Physics and Biomedical Engineering Department, University College London, London, UK. [53]Epilepsy Society MRI Unit, Chalfont St Peter, UK. [54]Division of Clinical Behavioral Neuroscience, Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA. [55]Departments of Neurology and Biomedical Engineering, University Hospital Olomouc, Olomouc, Czech Republic. [56]IRCCS Fondazione Don Carlo Gnocchi ONLUS, Milan, Italy. [57]Departments of Radiology, Pathology & Laboratory Medicine, Physics & Astronomy; International Collaboration on Repair Discoveries (ICORD), University of British Columbia, Vancouver, British Columbia, Canada. [58]Division of Pain Medicine, Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, CA, USA. [59]Department of Neurophysics, Max Planck Institute for

Human Cognitive and Brain Sciences, Leipzig, Germany. [60]Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China. [61]Tiantan Image Research Center, China National Clinical Research Center for Neurological Diseases, Beijing, China. [62]Center of Neuroimmunology, Laboratory of Advanced Imaging in Neuroimmunological Diseases, Hospital Clinic Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) and Universitat de Barcelona, Barcelona, Spain. [63]Fondation Campus Biotech Genève, Geneva, Switzerland. [64]Department of Radiology, Vanderbilt University Medical Center, Nashville, TN, USA. [65]UCSF Weill Institute for Neurosciences, Department of Neurology, University of California San Francisco, San Francisco, CA, USA. [66]Neuroradiology Section, Vall d'Hebron University Hospital, Barcelona, Spain. [67]E-health Centre, Universitat Oberta de Catalunya, Barcelona, Spain. [68]School of Biomedical Sciences, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia. [69]University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA. [70]Department of Neurology, Faculty of Medicine and Dentistry, Palacký University and University Hospital Olomouc, Olomouc, Czech Republic. [71]Felix Bloch Institute for Solid State Physics, Faculty of Physics and Earth Sciences, Leipzig University, Leipzig, Germany. [72]Institute for Advanced Biomedical Technologies, Department of Neuroscience, Imaging and Clinical Sciences, "G. D'Annunzio University" of Chieti-Pescara, Chieti, Italy. [58]e-mail: jcohen@polymtl.ca

## Acknowledgements

We thank G. Moran and B. Schrau (Siemens Healthcare), S. Banerjee and N. Takei (GE Healthcare) for sharing proprietary information and for their help with setting up manufacturer-specific protocols, C. Hurst, A. Cyr, A. Boré and P. Bellec (Functional Neuroimaging Unit), C. Tremblay (Polytechnique Montreal), A. Melek and H. Benali (PERFORM center, Concordia University), I. Levesque (McGill University), C. Nguyen (University of Minnesota), Prof. S. Aoki (Juntendo University Hospital) for helping with data acquisitions, Compute Ontario (https://computeontario.ca/) and Compute Canada (www.computecanada.ca) for providing the supercomputer infrastructure and all the volunteers who participated in the Spinal Cord MRI Public Database. This work was funded by the Canada Research Chair in Quantitative Magnetic Resonance Imaging (950-230815), the Canadian Institute of Health Research (CIHR FDN-143263), the Canada Foundation for Innovation (32454, 34824), the Fonds de Recherche du Québec–Santé (28826), the Fonds de Recherche du Québec–Nature et Technologies (2015-PR-182754), the Natural Sciences and Engineering Research Council of Canada (435897-2013), the Canada First Research Excellence Fund (IVADO and TransMedTech), the Quebec BioImaging Network (5886), Spinal Research (UK), Wings for Life (Austria, #169111) and Craig H. Neilsen Foundation (USA) for the INSPIRED project, the National Institutes of Health (NIH) through grants R00EB016689 (R.L.B.), R01EB027779 (R.L.B.), P41 EB027061 (CMRR) and P30 NS076408 (CMRR), the Instituto Investigación Carlos III (Spain, PI18/00823), the Czech Health Research Council grant no. NV18-04-00159, the Ministry of Health, Czech Republic–conceptual development of research organization (FNBr, 65269705), the National Imaging Facility and Queensland NMR Network (UQ), and SpinalCure Australia (M.J.R.), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 616905; European Union's Horizon 2020 research and innovation programme under the grant agreement No 681094, and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0137; BMBF (01EW1711A & B) in the framework of ERA-NET NEURON, the European Union's Horizon 2020 research and innovation programme under grant agreement no. 634541, the Engineering and Physical Sciences Research Council (R006032/1, M020533/1) and Rosetrees Trust (UK), UK Multiple Sclerosis Society (892/08, 77/2017), NIHR Biomedical Research Centres, UCLH, the Italian Ministry of Health Young Researcher Grant 2013 (GR-2013-02358177), the FISR Project 'Tecnopolo di nanotecnologia e fotonica per la medicina di precisione' (funded by MIUR/CNR, CUP B83B17000010001), TECNOMED project (funded by Regione Puglia, CUP B84I18000540002), Million Dollar Bike Ride from the University of Pennsylvania (MDBR-17-123-MPS), investigator-initiated PREdICT study at the Vall d'Hebron Institute of Oncology (Barcelona), funded by AstraZeneca and CRIS Cancer Foundation, the Wellcome Trust (UK) (203139/Z/16/Z), Systems, Technologies and Applications for Radiofrequency and Communications (STARaCOM), Swiss National Science Foundation (PCEFP3_181362/1) and the Max Planck Society and European Research Council (ERC StG 758974). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## Competing interests

G. Gilbert is an employee of Philips Healthcare.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41596-021-00588-0.

**Correspondence and requests for materials** should be addressed to J.C.-A.

**Peer review information** *Nature Protocols* thanks Felix W. Wehrli and the other, anonymous reviewer(s) for their contribution to the peer review of this work.



Received: 23 October 2020; Accepted: 10 June 2021;
Published online: 16 August 2021

**Related links**

**Key references using this protocol**

Cohen-Adad, J. Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers. *Sci. Data* https://doi.org/10.1038/s41597-021-00941-8 (2021).

Karbasforoushan, H., Cohen-Adad, J. & Dewald, J. P. A. *Nat. Commun.* **10**, 3524 (2019): https://doi.org/10.1038/s41467-019-11244-3

Martin, A. R. et al. *AJNR Am. J. Neuroradiol.* **38**, 1257–1265 (2017): https://doi.org/10.3174/ajnr.A5163

scientific **data**

OPEN

DATA DESCRIPTOR

# Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers

Julien Cohen-Adad *et al.*[#]

**In a companion paper by Cohen-Adad *et al.* we introduce the *spine generic* quantitative MRI protocol that provides valuable metrics for assessing spinal cord macrostructural and microstructural integrity. This protocol was used to acquire a single subject dataset across 19 centers and a multi-subject dataset across 42 centers (for a total of 260 participants), spanning the three main MRI manufacturers: GE, Philips and Siemens. Both datasets are publicly available via git-annex. Data were analysed using the Spinal Cord Toolbox to produce normative values as well as inter/intra-site and inter/intra-manufacturer statistics. Reproducibility for the *spine generic* protocol was high across sites and manufacturers, with an average inter-site coefficient of variation of less than 5% for all the metrics. Full documentation and results can be found at https://spine-generic.rtfd.io/. The datasets and analysis pipeline will help pave the way towards accessible and reproducible quantitative MRI in the spinal cord.**

## Background & Summary

Quantitative MRI (qMRI) aims at providing objective continuous metrics that specifically reflect the morphology, microstructure and/or chemical composition of tissues[1,2], thereby enabling deeper insight and understanding of disease pathophysiology. While qMRI techniques have been successfully implemented in the brain for several decades, they remain largely underutilized for spinal cord (SC) imaging in both clinical and research settings, mostly as a direct consequence of the many challenges that need to be overcome in order to acquire good quality data[3,4]. In a companion paper[5], we introduce the *spine generic* protocol for acquiring high-quality qMRI of the human SC at 3 Tesla (T). The *spine generic* protocol includes relevant sequences and contrasts for calculating metrics sensitive to macrostructural and microstructural integrity: T1w and T2w imaging for SC cross-sectional area (CSA) computation, multi-echo gradient echo for gray matter CSA, as well as magnetization transfer and diffusion weighted imaging for assessing white matter microstructure.

To demonstrate the practical implementation and reproducibility of the *spine generic* protocol, single subject and multi-subject datasets were acquired across multiple centers. Relevant qMRI metrics were calculated using a fully-automatic analysis pipeline, and those metrics were compared within site, across sites (within manufacturer), and across different manufacturers. The generated normative values will be useful as reference for future clinical studies.

## Methods

**Data acquisition.** Single-participant and multi-participant datasets were acquired across multiple centers (see details below). The scan operator (researcher or MR technician) was instructed to follow the *spine-generic* protocol[5]. Briefly, each participant was positioned in the head-first supine position. The following sequences were run: localizer, 3D sagittal T1w, 3D sagittal T2w, 2D axial diffusion weighted echo planar imaging (30 directions at a b-value of 800 s/mm$^2$), 3D axial gradient echo with/without an MT pulse and an additional T1w scan, 2D axial

[#]A full list of authors and their affiliations appears at the end of the paper.

multi-echo gradient echo. Data were collected, organized and analyzed according to a fully-documented procedure available at https://spine-generic.rtfd.io.

**Ethical compliance.** We have complied with all relevant ethical regulations. Local ethics committees of the participating institutions listed in Online-only Table 1 approved the study protocol. Signed informed consent was obtained from all participants under the compliance of the corresponding local ethics committee and stored in the corresponding local research center under responsibility of the local principal investigator/s (listed as "*Contact*" in the Online-only Table 1).

**Single subject.** The same participant (JCA, 38 y.o., male) was scanned at 19 different centers within 77 days. This represents a "best case scenario" in terms of reproducibility because the participant was very familiar with being scanned, knew how to position himself in the scanner and how to breathe, directly interacted with the MR technician at each site to ensure the standard operating procedure (SOP) was understood, and was able to adapt the sequence parameters on the scanner if the protocol was not imported properly (e.g., hardware or software version incompatibility).

**Multiple subjects.** In order to evaluate a more realistic (routine) scenario, 42 different groups worldwide with varying levels of expertise each scanned six different healthy participants (3 males, 3 females), aged between 19 and 56 y.o., median age 28 y.o., resulting in 260 data sets. Participant-specific list of age and sex distribution is available at https://github.com/spine-generic/data-multi-subject/blob/r20201130/participants.tsv. Each group used the *spine generic* protocol and SOP, and obtained consent to scan and upload participants' anonymized data onto a publicly-available repository. Anatomical scans (T1w, T2w) where facial features are visible were "defaced" before being released to the public domain using *pydeface*[6]. Some centers equipped with more than one system were asked to scan 6 different participants on each scanner, although they were also given the possibility to scan the same 6 participants across systems for convenience. In the latter case, this could slightly bias the inter-site coefficient of variation (COV), however it would not affect the inter-manufacturer COV. The list of centers and scanner models is shown in Online-only Table 1.

Data from each participant was then entered into the processing pipeline described in the next section. COVs were computed within site and manufacturers, and compared across manufacturers.

**Data processing.** Data were processed using Spinal Cord Toolbox (SCT) v5.0.1[7] and spine-generic v2.6 (https://github.com/spine-generic/spine-generic/releases/tag/v2.6). The processing pipeline is illustrated in Fig. 1 and is fully documented at https://spine-generic.rtfd.io. Briefly, for the T1w scan, the SC was segmented using deep-learning[8], vertebral levels were identified[9] and the SC was registered to the PAM50 template[10] using the C3 and C5 vertebrae as labels. Then, the SC CSA was computed slice-wise (corrected for angulation between the SC and the slice) and averaged between the C2 and C3 vertebral levels. For the T2w scan, the SC was segmented[8], registered with the PAM50 template using the transformation from the T1w scan and the CSA was computed and averaged between C2 and C3. For the ME-GRE (T2*w) scan, the SC[8] and gray matter (GM)[11] were segmented and registered with the PAM50 template using the transformation from the T1w scan. Then, GM CSA was computed and averaged between C3 and C4. For the magnetization transfer (MT) protocol, the SC on the GRE-T1w scan was segmented[8] and registered with the PAM50 template using the initial transformation from the T1w scan. GRE-MT1 and GRE-MT0 scans were registered to the GRE-T1w scans using an automatically-generated mask tightly fitting the SC for more accurate registration. Magnetization transfer ratio (MTR) and MTsat[12] were computed and their values extracted for the white matter (WM) between C2 and C5 using the WM probabilistic atlas[13]. For the diffusion weighted imaging (DWI) scan, the time series of diffusion-weighted scans were averaged and the SC was segmented so that a mask could be created around it. The time series were motion-corrected using slice-wise 2D transformations regularized along z[7], with time-adjacent volumes grouped together for increased robustness[14]. The PAM50 template was then registered to the DWI dataset using the initial transformation from the T1w scan and diffusion tensor imaging (DTI) metrics were computed using the weighted least squares fitting algorithm implemented in Dipy[15] and wrapped in SCT's function `sct_dmri_compute_dti`. No post-processing was performed. Lastly, DTI metrics were extracted within the WM between C2 and C5.

Processing was run on a supercomputer cluster (https://docs.computecanada.ca/wiki/Graham), by distributing 32x OpenMP jobs across 9 nodes in parallel (each node equipped with 2 x Intel E5-2683 v4 Broadwell @ 2.1 GHz, with 128GB reserved RAM), enabling us to process all 260 participants in parallel (one participant per CPU core). Software parallelization was achieved using Python's *multiprocessing* package available in SCT's `sct_run_batch`. Total processing times were 37 min (single subject) and 40 min (multi-subject).

**Statistics.** Intra- and inter-manufacturer differences were tested using a one-way ANOVA. Post-hoc analyses testing pairwise differences across manufacturers were performed using the Tukey Honestly Significant Difference (HSD) using the family-wise error rate to account for multiple comparisons. Significance level was set to $p = 0.05$. Interactive plots available online were generated with Plotly (https://plotly.com/).

## Data Records

The two datasets associated with this publication are:

- The Spine Generic Public Database (Single Subject)[16]
- The Spine Generic Public Database (Multi-Subject)[17]

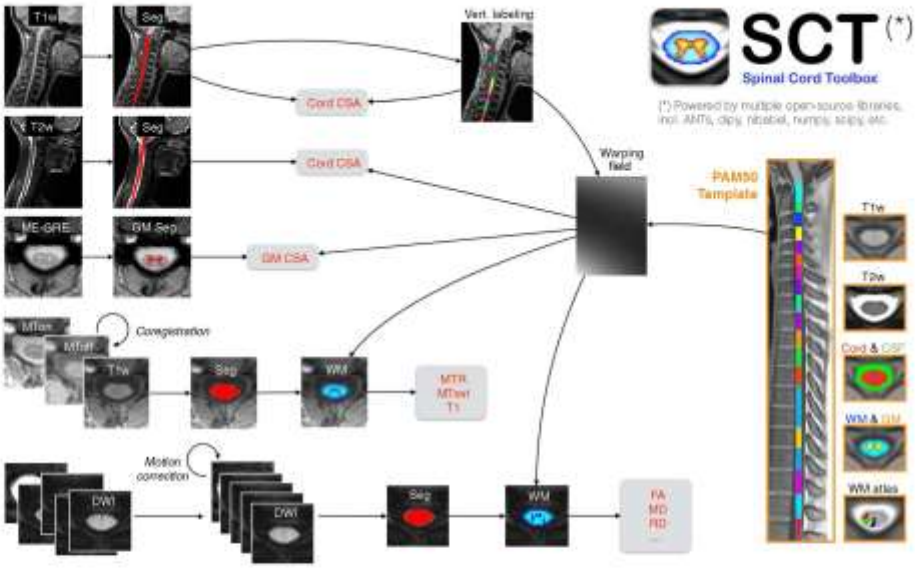



Fig. 1 Overview of the processing pipeline based on SCT. Briefly, for each participant, the SC is automatically segmented on the T1w, T2w, GRE-T1w, and mean DWI scans, while the gray matter is segmented on the ME-GRE scan (after averaging across echoes). Vertebral labeling is run on the T1w scan, followed by registration of the PAM50 template to each contrast. Estimated metrics are shown in red.

**Dataset management.** Figure 2 illustrates the data management workflow. Datasets are managed using git-annex (https://git-annex.branchable.com/); git-annex is built on git technology and enables the separation of large files (NIfTI images, hosted on Amazon Web Services, AWS) from small files (metadata and documentation, hosted on GitHub). This decision was based on the modularity of git-annex (multiple mirrors can be added) and its compatibility with Datalad[18]. The documentation for contributing to the repositories is hosted on a wiki (https://github.com/spine-generic/spine-generic/wiki).

To facilitate data aggregation across centers, we used the *Brain Imaging Data Structure* (BIDS) convention[19]. BIDS notably features JSON metadata files as a sidecar for each NIfTI file, which includes relevant acquisition parameters, making it easy to assess how well each site followed the generic protocol and which parameters were modified. Parameter verification (within a specified tolerance) as well as file and folder naming is automated (https://spine-generic.rtfd.io/en/latest/data-acquisition.html#checking-acquisition-parameters) such that, every time new participants are added to the database, a notification is sent to a continuous integration system (e.g., https://github.com/spine-generic/data-multi-subject/runs/2730553396?check_suite_focus=true) that downloads the dataset and runs custom scripts to verify the validity of the dataset. For example, if a flip angle for a particular volume exceeded the tolerance range, the BIDS validator would fail and the data would not be merged. In that case, the management team would reach out to the data contributors asking if they can reacquire the data. If not, the data would not be added to the dataset. Another (less problematic) example: if a file was incorrectly named (`sub_amu01_T2w.nii.gz` instead of `sub-amu01_T2w.nii.gz`), the BIDS validator would fail. In that case, the management team would manually correct the file name, commit and push the change to the working branch and wait for the BIDS validator to pass before being able to merge the new data on the main (master) branch. Below is an example of the "BIDS Validator" script output (only showing part of it):

```
WARNING: Incorrect FlipAngle: sub-amu01_T2w.nii.gz; FA=180 instead of 120
WARNING: Incorrect RepetitionTime: sub-amu02_T2w.nii.gz; TR=2 instead of 1.5
WARNING: Incorrect FlipAngle: sub-amu02_T2w.nii.gz; FA=180 instead of 120
WARNING: Incorrect RepetitionTime: sub-amu03_T2w.nii.gz; TR=2 instead of 1.5
WARNING: Incorrect FlipAngle: sub-amu03_T2w.nii.gz; FA=135 instead of 120
Missing jsonSidecar: ./derivatives/labels/sub-oxfordOhba05/anat/
sub-oxfordOhba05_acq-T1w_MTS_seg-manual.json
Missing jsonSidecar: ./derivatives/labels/sub-oxfordOhba05/anat/
sub-oxfordOhba05_T1w_labels-disc-manual.json
Missing jsonSidecar: ./derivatives/labels/sub-oxfordOhba05/anat/
sub-oxfordOhba05_T1w_RPI_r_labels-manual.json
```


Fig. 2 Illustration of the dataset management, from acquisition to end-user consumption.

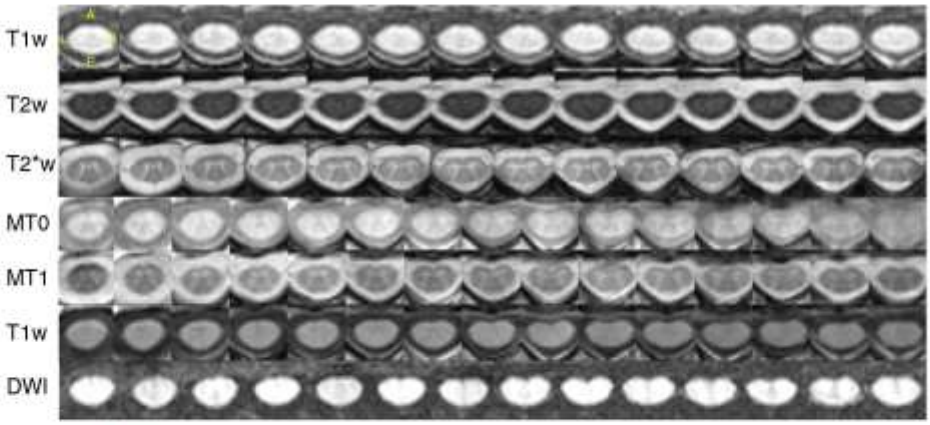



Fig. 3 Axial views of good quality data for all sequences in the spine generic protocol across various slices (the exact coverage along the SC varies because the slice thickness varies across sequences). DWI corresponds to the mean DWI data after motion correction. The images are from different participants. T1w: *vuisAchieva02*; T2w: *milan01*; T2*w (ME-GRE): *brnoCettec01*; MT0, MT1, T1w (for the MTS protocol) and DWI : *barcelona04*. Axial views were automatically generated by SCT's QC report.

## Technical Validation

**Data quality.** Overall, data quality was satisfactory based on qualitative visual inspection. Criteria included the correctness of field of view prescription, proper selection of receive coils, quality of shimming (assessed by looking at fat saturation performance and the presence of susceptibility distortions), and the presence and severity of motion artifacts. Figure 3 shows examples of good quality data for all sequences. A few operator errors occurred, including: mis-labeled MT0 for MT1 and MT1 for MT0, shim parameters changed between MT0 and MT1 scans (causing different signal intensities, and hence not suitable for MT-based metrics), change of FFT scaling factor between the MT1/MT0 scans and the T1w scan used to compute MTsat and T1 maps (causing different signal quantization and hence not suitable for MT-based metrics unless corrected for), and repositioning of the participants, causing mis-alignment between the images before/after repositioning and violation of the analysis pipeline assumptions (all images are supposed to be acquired with the patient in the same position). These errors were not caught by the BIDS validator, but by the managing team during visual inspection of the data and interpretation of the qMRI metrics results. In future work, the data validator could be made sensitive to these issues. For example, the FFT scaling factor and shim coefficient are sometimes retrievable from the DICOM data and could be checked. Also, the qform (affine matrix present in the NIfTI header) could be checked to ensure consistency across data from the same series, e.g. MT1, MT0, GRE-T1w. Regarding the mis-labeling of MT1/MT0, training a deep learning model to recognize image contrast could address this issue.

Figure 4 illustrates some of the image artifacts encountered during QC. A list of poor data quality scans is available on the Github's issues of the dataset under the label "data-quality" (https://github.com/spine-generic/data-multi-subject/labels/data-quality); most of these were caused by patient motion. Mosaics of images for every contrast and every participant are available in the supplementary materials (Figures S1–S5). Additional examples of good quality data are also available in the spine generic website (https://spine-generic.rtfd.io/en/latest/data-acquisition.html#example-of-datasets).

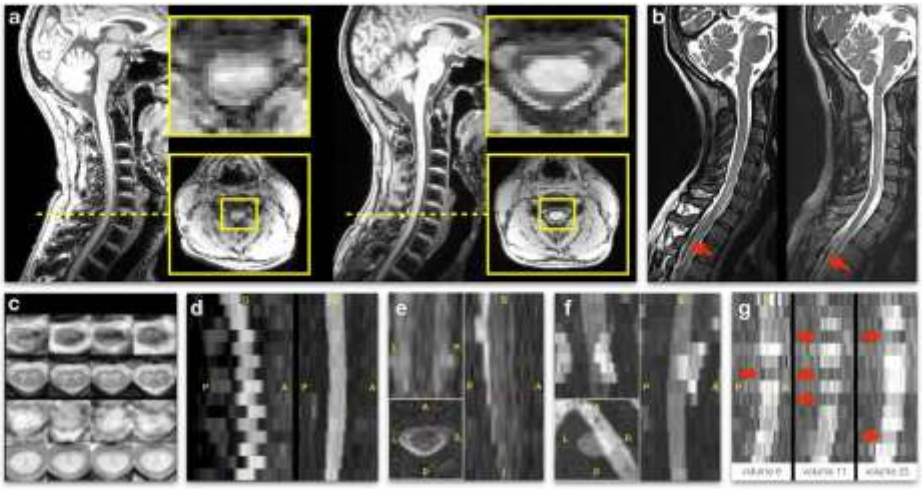


Fig. 4 Examples of image artifacts: (**a**) T1w MPRAGE taken in the same participant (from the single subject database) at two different sites on a Siemens Prisma: oxfordFmrib (left) and juntendoPrisma (right). The slightly larger cervical lordosis on the left likely induced more pronounced cerebrospinal fluid (CSF) flow and SC motion resulting in the artifact shown in the axial view. (**b**) T2w scans showing signal drops in the CSF likely due to a poorly-recovered CSF signal combined with flow effects. These two participants (beijingVerio01 and strasbourg03), were acquired with a flip angle of 180° instead of the recommended 120°, which likely explained the presence of those artifacts (longer TR was required for sufficient T1 recovery). (**c**) Axial view of ME-GRE scans with (fslAchieva04, 1st row) and without motion (brnoCeitec01, 2nd row), and axial view of GRE-MT0 with (fslAchieva04, 3rd row) and without motion (barcelona04, 4th row). (**d**) Mean DWI scan from a Philips site (ubc02, left panel) with a concatenated acquisition wherein odd slices are acquired during the first half of the entire acquisition (spanning all b-vectors) and the even slices are acquired during the second half. In the event of participant motion between those two acquisition sub-sets, apparent motion will be visible between the odd and even slices. When odd and even slices are acquired closer in time (in ascending/descending mode, or interleaved but sequentially within the same b-vector), this artifact is not visible (mountSinai03, right panel). Such an artifact could be problematic for image registration with regularization along the S-I axis, or for performing diffusion tractography. (**e**) b=0 image from a DWI scan (perform02) acquired with poor shimming and resulting signal dropout. (**f**) Another example of poor shimming resulting in sub-efficient fat saturation, with the fat being aliased on top of the SC. Here we show the mean DWI scan of a participant from the single subject database (perform). (**g**) Effect of pulsatile movement on a non-cardiac gated acquisition (single subject, juntendoAchieva). Diffusion-weighted scans (sagittal view) acquired at three b-vecs fairly orthogonal to the SC (i.e., diffusion-specific signal attenuation should be minimum in the SC), showing abrupt signal drop at a few slices (red arrows), likely due to cardiac-related pulsatile effects.

**Quantitative results: Single subject.** Overall, data quality was satisfactory. All images were visually inspected to ensure that there were no significant errors in the masks used to average the signals in the SC, WM or GM, and any errors were manually corrected. A list of poor quality scans is available on Github in the issues for the dataset, under the label "data-quality" (https://github.com/spine-generic/data-single-subject/labels/data-quality). Complete metrics and statistical tests are available in the r20201130 release assets (https://github.com/spine-generic/data-single-subject/releases/download/r20201130/results.zip).

Figure 5 shows the SC CSA data from the T1w scan, averaged between cervical levels 2 and 3 (C2 and C3), for the single participant across the 19 centers. Within each manufacturer, the inter-site standard deviation ranges from 0.65 $mm^2$ (Siemens) to 1.56 $mm^2$ (GE), which is remarkably small considering that the size of a pixel is 1 $mm^2$. The inter-site COVs were 2.3% for GE, 1.8% for Philips and 0.9% for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), with the Tukey test showing significant differences between GE and Philips (p-adjusted = 0.03) and between GE and Siemens (p-adjusted < 0.01).

Figure 6 shows the SC CSA for the T2w scan, again averaged between cervical levels 2 and 3 (C2 and C3). The inter-site COVs were 2.3% for GE, 2.1% for Philips and 1.5% for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), with the Tukey test showing significant differences between Philips and Siemens (p-adjusted < 0.01).

Figure 7 shows the gray matter CSA for the ME-GRE scan, averaged between cervical levels C3 and C4. The inter-site COVs were 2.5% for GE, 3.4% for Philips and 3.4% for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), with the Tukey test showing significant differences between GE and Philips (p-adjusted < 0.01) and between Philips and Siemens (p-adjusted < 0.01).

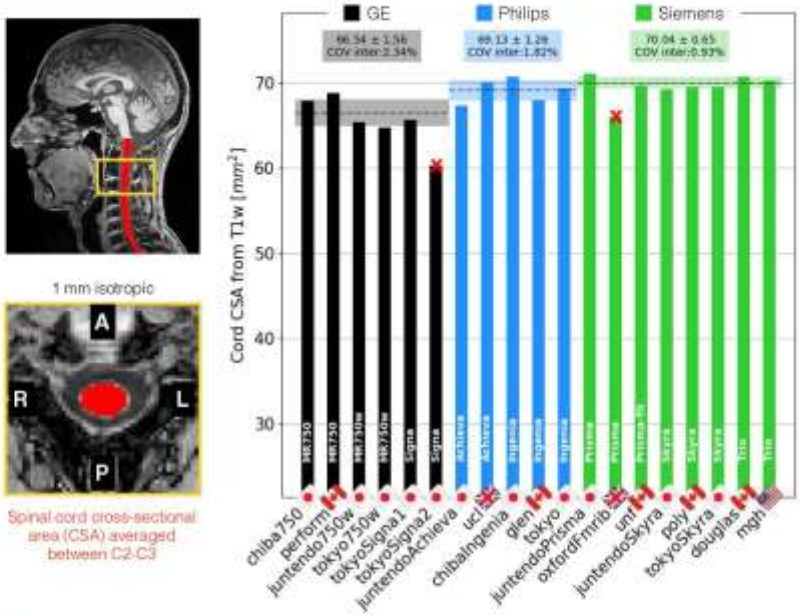



Fig. 5 Results of the single subject study for the T1w scan. The cross-sectional area (CSA) of the SC was averaged between the C2 and C3 vertebral levels. Sites *tokyoSigna2* and *oxfordFmrib* were excluded from the statistics due to excessive motion.

Figure 8a shows the MTR average for the WM between C2 and C5. The inter-site COVs were 8.0% for GE, 4.2% for Philips and 3.6% for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), but the Tukey test showed no significant difference across pairwised manufacturers.

Figure 8b shows the MTsat results. The inter-site COVs were 11.3% for GE, 2.9% for Philips and 5.2% for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), with the Tukey test showing significant differences between GE and Philips (p-adjusted = 0.03), between GE and Siemens (p-adjusted < 0.01), and between Philips and Siemens (p-adjusted < 0.01).

Sites *perform* and *juntendo750w* were excluded from the MTR statistics because the TR for the GRE-MT0 and GRE-MT1 was set to 62 ms (vs. 35 ms for the other GE sites), causing a drastic decrease in MTR values. These sites were not excluded from MTsat, because this metric is supposed to account for the T1 recovery effect[12] as was indeed observed, with those sites now falling inside the 1σ interval. The site *tokyoSigna1* fell outside the 1σ interval because of issues related to image registration.

Figure 9 shows the average fractional anisotropy (FA) in WM across C2 and C5. The inter-site COVs were 0.8% for GE, 4.5% for Philips and 2.8% for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), but the Tukey test showed no significant difference across pairwised manufacturers. One of the outliers (*tokyo750w*) was due to the absence of the FOCUS license, which led us to rely on saturation bands to prevent aliasing. However, those were not efficient (likely due to poor shimming in the region), with poor fat saturation efficiency that yielded spurious diffusion tensor fits (e.g. FA >1 or <0).

Average +/− standard deviation (SD) and COVs for mean diffusivity were, respectively, (0.62 +/− 0.03) $mm^2/s$ and 5.6% for GE, (1.00 +/− 0.06) $mm^2/s$ and 5.71% for Philips, and (1.01 +/− 0.05) $mm^2/s$ and 4.83% for Siemens. Average +/− SD and COVs for radial diffusivity were, respectively, (0.36 +/− 0.03) $mm^2/s$ and 7.21% for GE, (0.51 +/− 0.06) $mm^2/s$ and 11.34% for Philips, and (0.54 +/− 0.03) $mm^2/s$ and 6.37% for Siemens.

**Quantitative results: Multiple subjects.** As in the case of the single subject data, all images were visually inspected to ensure that there were no significant errors in the masks used to average the signals in the SC, WM or GM and any errors were manually corrected. Complete metrics and statistical tests are available in the r20201130 release assets (https://github.com/spine-generic/data-multi-subject/releases/download/r20201130/results.zip). Interactive plots are available on the *spine generic* website (https://spine-generic.readthedocs.io/en/latest/analysis-pipeline.html#results).

In Figure 10 we show the multi-subject, multi-center results for the SC CSA (averaged between C2 and C3) obtained from the T1w scan. The intra-site COVs were averaged for each manufacturer and found to be all just under 7.8%. The inter-site COVs (and inter-site ANOVA p-values) were 3.08% ($p = 0.52$) for GE, 3.22% ($p = 0.44$) for Philips and 4.41% ($p = 0.12$) for Siemens. The inter-manufacturer difference was significant ($p = 0.0007$), with

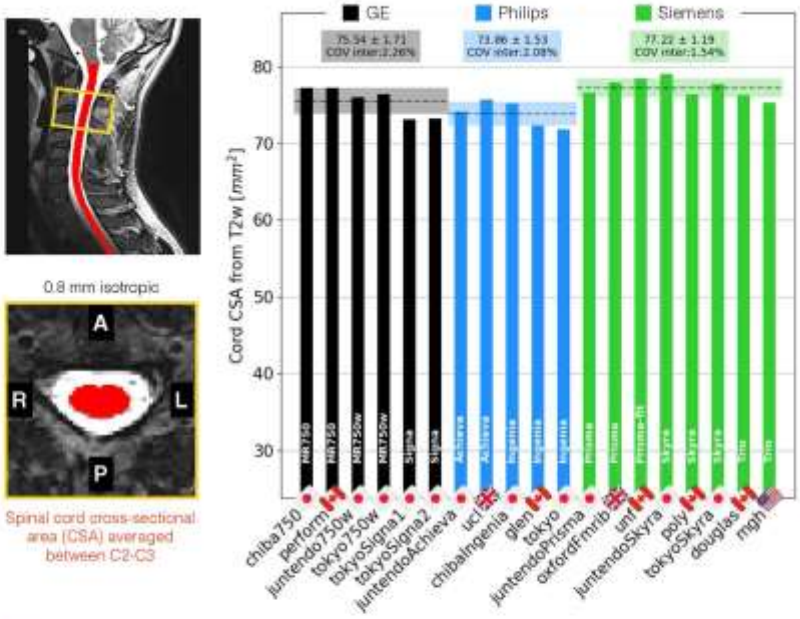



Fig. 6 Results of the single subject study for the T2w scan. The cross-sectional area (CSA) of the SC was averaged between the C2 and C3 vertebral levels.

the Tukey test showing significant differences between GE and Philips (p-adjusted $< 0.01$), and between GE and Siemens (p-adjusted $< 0.01$).

Figure 11 has the CSAs obtained from the T2w scans (also averaged between C2 and C3). Again, intra-site COVs were close to 8%. Inter-site COVs (and ANOVA results) were 4.24% ($p = 0.13$) for GE, 3.39% ($p = 0.35$) for Philips, and 5.07% ($p = 0.004$) for Siemens. The inter-manufacturer difference was not significant ($p = 0.17$).

Interestingly, T2w images were found to lead to larger cord CSAs than T1w images. In Figure 12 we show the relationship between T1w and T2w cord CSAs for all 3 manufacturers. Linear regressions led to $R^2$ values that ranged from 0.63 for GE scanners (note that the same sequence was not used for all GE scanners) to 0.90 for Philips scanners.

Figure 13 shows the GM CSA, averaged across C3 and C4. The intra-site COV ranges from 5.83% (Siemens) to 9.16% (Philips). The inter-site COVs (and inter-site ANOVA p-values) were 4.22% ($p = 0.14$) for GE, 5.62% ($p = 0.03$) for Philips, and 3.76% ($p = 0.005$) for Siemens. The inter-manufacturer difference was significant ($p = 2.3 \cdot 10^{-11}$), with the Tukey test showing significant differences between GE and Philips (p-adjusted $< 0.01$), and between Philips and Siemens (p-adjusted $< 0.01$). The larger intra-site COV on Philips and the significantly lower values are likely due to the fact that some Philips sites used older versions of the consensus protocol, which produced lower contrast between white and gray matter and, as a result, less reliable gray matter segmentations.

Figure 14 shows MTR results averaged between C2 and C5. The intra-site COVs were averaged for each manufacturer and found to be all under 3.6%. The inter-site COVs (and inter-site ANOVA p-values) were 2.0% ($p = 0.03$) for GE, 1.8% ($p = 0.17$) for Philips, and 2.3% ($p < 0.01$) for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), with the Tukey test showing significant differences between GE and Philips (p-adjusted $= 0.02$), and between GE and Siemens (p-adjusted $= 0.01$).

Figure 15 shows MTsat results, also averaged between C2 and C5. The intra-site COVs were all under 11%. The inter-site COVs (and inter-site ANOVA p-values) were 7.5% ($p < 0.01$) for GE, 4.9% ($p = 0.11$) for Philips, and 9.0% ($p = 0.09$) for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), with the Tukey test showing significant differences between GE and Philips (p-adjusted $= 0.04$), between GE and Siemens (p-adjusted $< 0.01$), and between Philips and Siemens (p-adjusted $< 0.01$). Some outliers have notable impacts on the standard deviations: *nottwil03*, *nottwil04*, *pavia05*. These outliers are likely caused by poor image quality due to participant motion on the MT0 scan (see the full reports on the Github issue https://github.com/spine-generic/data-multi-subject/issues/36). Interestingly, these participants did not produce such outliers on the MTR results (which is computed from the MT1 and MT0 scans), and the T1w scan looked visually normal. We therefore decided to keep these participants on the figure in order to highlight possible implications about the reliability of the MTsat measures as a myelin biomarker (see discussion). We also decided to keep the *stanford* site (removed for MTR computation), because the T1 recovery effect induced by the different TR compared to other sites is supposed to be taken into account by the additional GRE-T1w scan when computing the MTsat metric, as is indeed confirmed in the figure (average MTsat for this site falls within the $1\sigma$-$2\sigma$ interval).

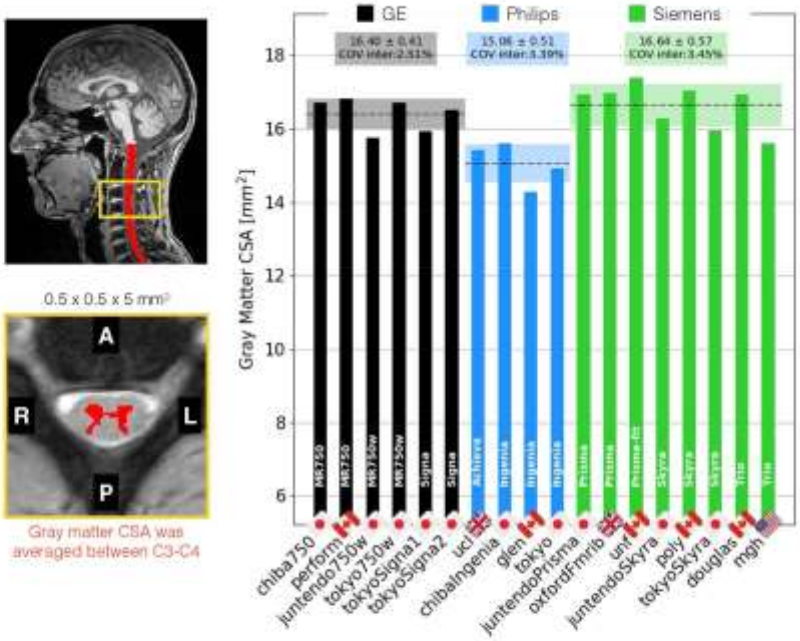



Fig. 7 Results of the single subject study for the ME-GRE scan. Gray matter CSA was computed after automatic gray matter segmentation and averaged between C3 and C4 vertebral levels.

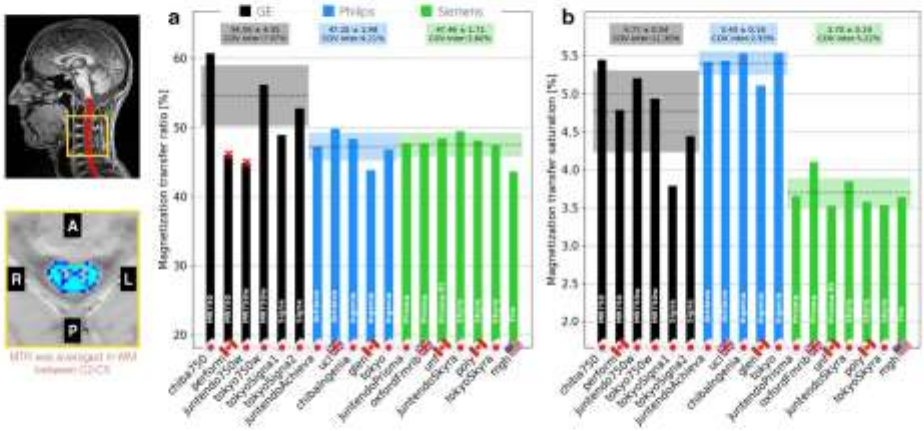



Fig. 8 Results of the single subject study for the MT protocol. The mean MTR (**a**) and MTsat (**b**) were computed in the white matter between C2 and C5. Sites *perform* and *juntendo750w* were excluded from the statistics because the TR for the GRE-MT0 and GRE-MT1 was set to 62 ms (vs. 35ms for the other GE sites), causing drastic decrease of MTR values. These sites were not excluded from MTsat.

Figure 16 shows FA results from the DWI scans, averaged between C2 and C5. The intra-site COVs were averaged for each manufacturer and found to be all under 5.2%. The inter-site COVs (and inter-site ANOVA p-values) were 3.0% ($p = 0.25$) for GE, 3.6% ($p < 0.01$) for Philips and 3.5% ($p < 0.01$) for Siemens. The inter-manufacturer difference was significant ($p < 0.01$), with the Tukey test showing significant differences between GE and

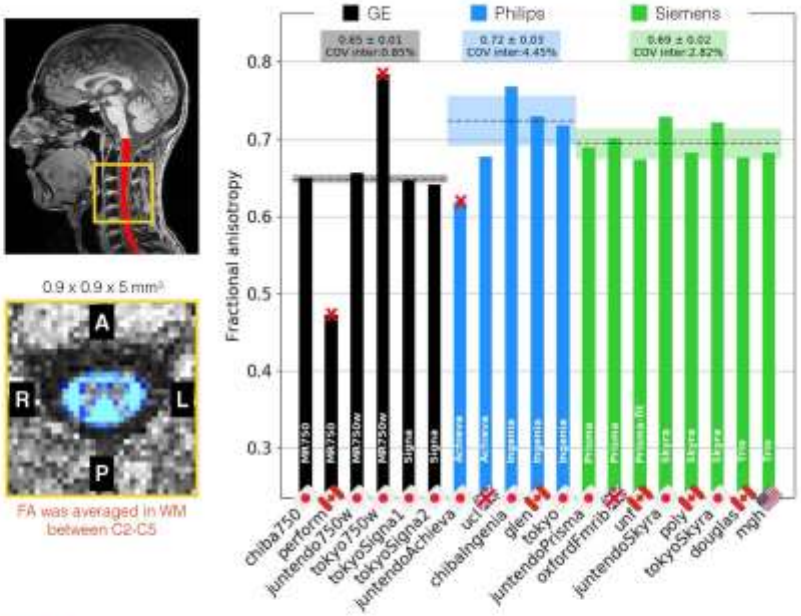



Fig. 9 Results of the single subject study for the DWI protocol. The FA in the SC WM was averaged between the C2 and C5 vertebral levels. The following sites were excluded: *perform* (strong fat aliasing artifact), *tokyo750w* (poor shimming) and *juntendoAchieva* (no cardiac gating).

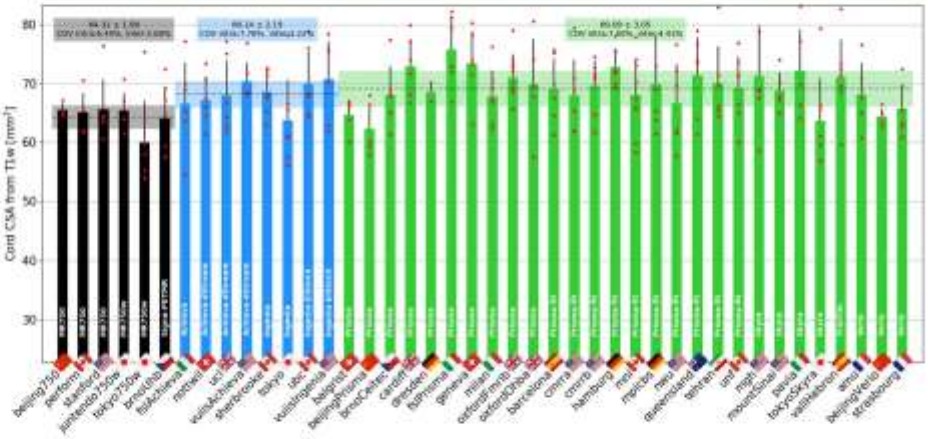



Fig. 10 Results of multi-subject study for the T1w scan. As in the single subject study, the cross-sectional area of the SC was averaged between the C2 and C3 vertebral levels. Black, blue and green bars respectively correspond to GE, Philips and Siemens, with the manufacturer's model indicated in white letters on each bar. The following participants were excluded from the statistics: *balgrist01* (motion), *beijingGE04* (motion), *mnt806* (motion), *mountSinai03* (participant repositioning), *oxfordFmrib04* (participant repositioning), *pavia04* (motion) and *perform06* (motion).

Philips (p-adjusted < 0.01), between GE and Siemens (p-adjusted < 0.01), and between Philips and Siemens (p-adjusted < 0.01). Average +/− SD and inter-site COVs for mean diffusivity were, respectively, (0.73 +/− 0.09) mm²/s and 12.52% for GE, (0.97 +/− 0.08) mm²/s and 7.82% for Philips, and (0.99 +/− 0.04) mm²/s and 4.40%

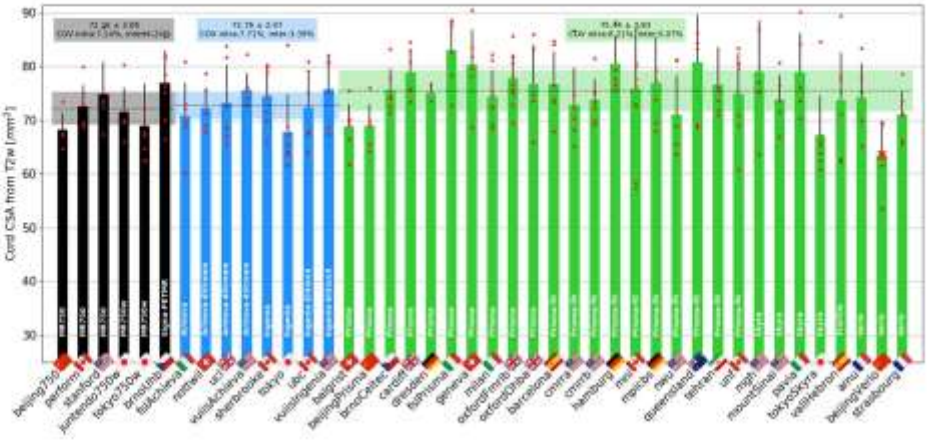



Fig. 11 Results of multi-subject study for the T2w scan. The cross-sectional area of the SC was averaged between the C2 and C3 vertebral levels. The Siemens site *beijingVerio* was excluded from statistics (red cross) due to different TR and FA causing biases in the segmentation volume. The following participants were excluded: *oxfordFmrib04* (T1w scan was not aligned with other contrasts due to participant repositioning), *pavia04* (motion) and *mountSinai03* (participant repositioning).

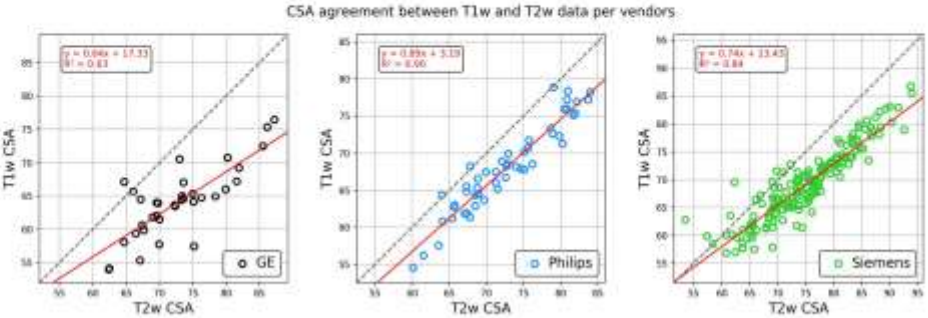



Fig. 12 Relationship between CSA calculated from the T1w vs. T2w scans. The same site and participants excluded in Figs 10 and 11 were also excluded here.

for Siemens. Average +/− SD and COVs for radial diffusivity were, respectively, (0.42 +/− 0.04) $mm^2/s$ and 10.31% for GE, (0.48 +/− 0.06) $mm^2/s$ and 12.25% for Philips, and (0.52 +/− 0.05) $mm^2/s$ and 8.91% for Siemens.

**Differences in qMRI results between manufacturers.** Before discussing differences across and within manufacturers, we would like to stress that results presented here will become further refined with time because, as for any neuroimaging analysis pipeline, the algorithms evolve. Moreover, visual QC and manual corrections are prone to human error. We therefore encourage users of this living database to provide feedback. As it is an open source project, contributions are welcome. Also, as future participants are added, the statistics will be updated.

*Spinal cord CSA.* Within manufacturers, SC CSAs showed a maximum inter-site COV of 2.4% for the single subject study and 5% for the multi-subject study, for both T1w and T2w contrasts, which is highly encouraging. Overall, intra-site COVs were higher than inter-site COVs, which is expected because CSAs are known to vary substantially across individuals[21]. Hence, taking the mean within each site and comparing it across sites somewhat smooths this inherent inter-individual variability, putting aside geographical differences. This could be the goal of follow-up investigations.

Regardless of the manufacturer, intra-site COVs were about two-fold higher for SC CSAs (8%) compared to MTR and DTI-FA (4–5%). This result is not surprising, considering that, as noted above, SC size is known to vary across healthy adults, while white matter microstructure (which MTR and DTI-FA measure) is not expected to

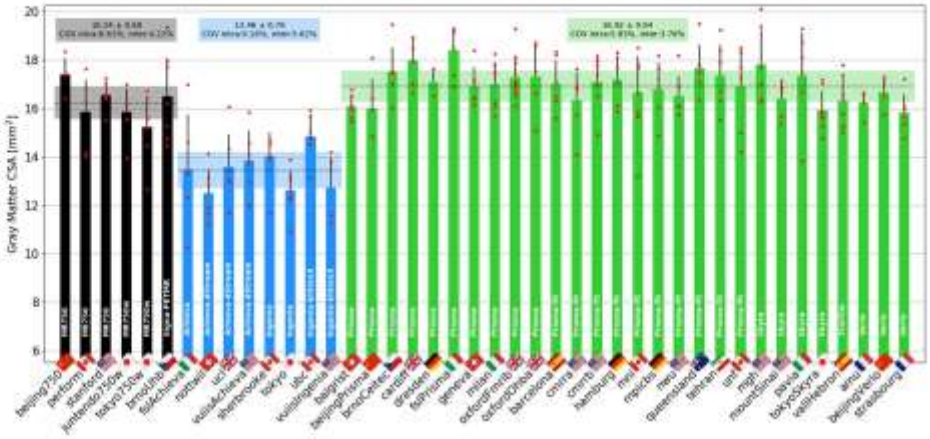



Fig. 13 Gray matter CSA computed after automatic gray matter segmentation on the ME-GRE scan and averaged between C3 and C4 vertebral levels. The following participants were excluded due to motion artifacts: *amu03*, *fslAchieva04*, *vuiisIngenia04* and *vuiisIngenia05*.

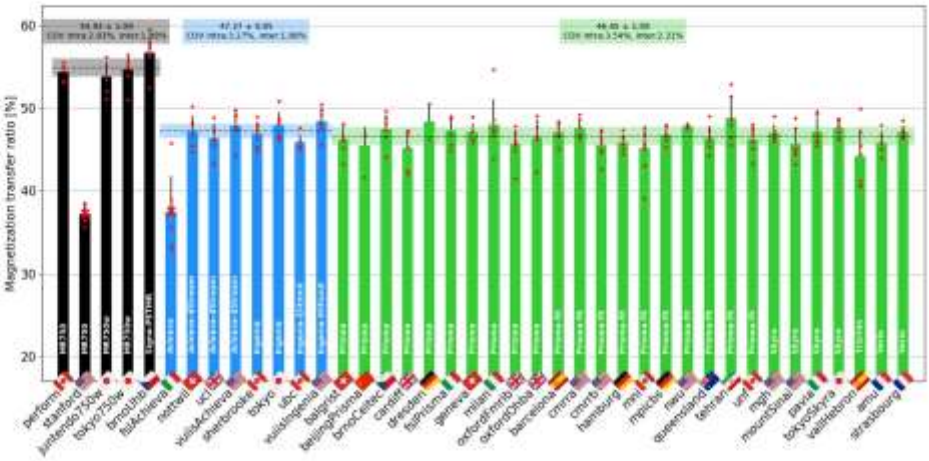



Fig. 14 MTR results computed from the GRE-MT0 and GRE-MT1 scans and averaged in the SC WM between the C2 and C5 vertebral levels. The following sites were removed from the statistics: *stanford* (large difference in the TR), *fslAchieva* (wrong field of view (FOV) placement). The following participants were also removed: *beijingPrisma04* (different coil selection, shim value and FOV placement between MT1 and MT0), *geneva02* (FOV positioning changed between MT1 and MT0), *oxfordFourth04* (T1w scan was not aligned with other contrasts due to participant repositioning).

vary much between healthy individuals[21]. Note that there is no conclusive evidence of a correlation of SC CSA with age[22], although some studies do report smaller cord area in older participants[23,24]. There is currently no accepted consensus on an effective and reliable normalization method for SC CSA[20]. Given that CSA is a widely used biomarker for neurodegenerative diseases such as MS, reducing that inter-subject variance is a much needed goal for the research community.

T1w scans showed slightly better intra- and inter-site COVs compared to T2w scans. This is rather surprising, given that T2w scans look visually cleaner, with a sharper SC/CSF border, and the fact that they are less prone to participant or SC motion artifacts. The SC CSAs obtained from the T1w scans were significantly lower for GE scanners, compared to both Philips and Siemens, whereas for T2w scans, the CSA was comparable across all three

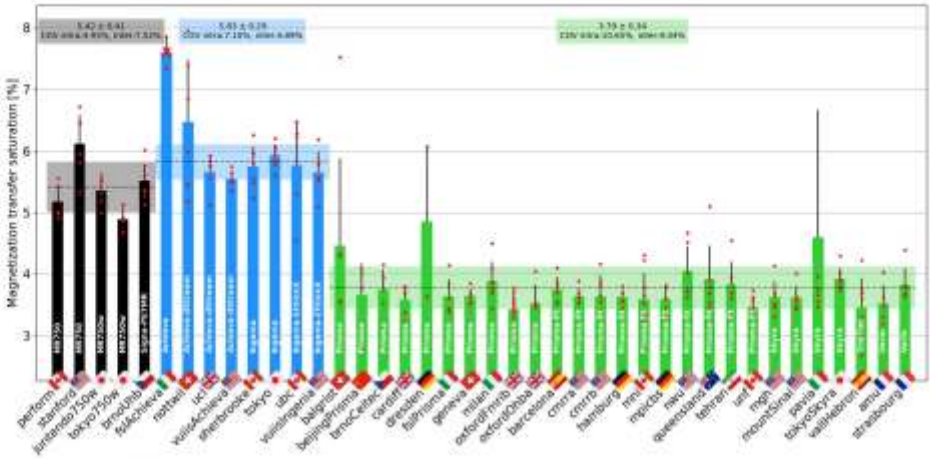



Fig. 15 MTsat results computed from the GRE-MT0, GRE-MT1 and GRE-T1w scans and averaged in the SC WM between the C2 and C5 vertebral levels. The following site was removed from the statistics: *fslAchieva* (wrong FOV placement). The following participants were also removed: *beijingPrisma04* (different coil selection, shim value and FOV placement between MT1 and MT0), *geneva02* (FOV positioning changed between MT1 and MT0), *oxfordFmrib04* (T1w scan was not aligned with other contrasts due to participant repositioning).

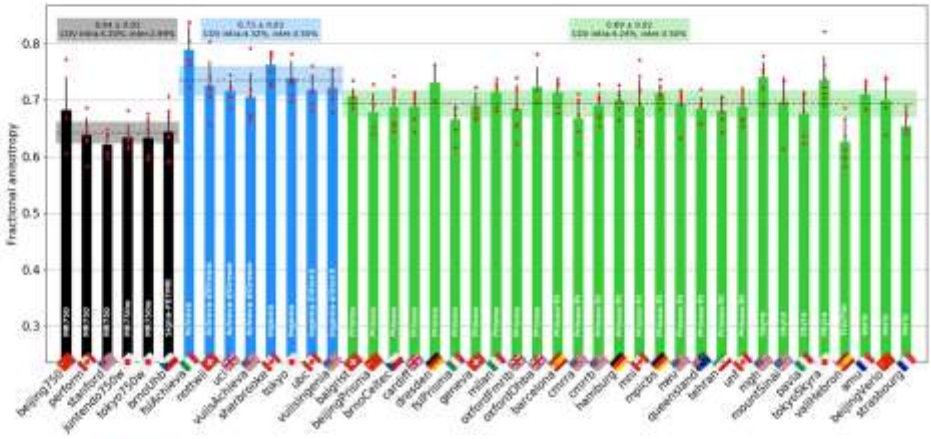



Fig. 16 Results of multi-subject study for the DWI scan. The FA of the SC WM was averaged between the C2 and C5 vertebral levels. The following participants were excluded: *beijingPrisma03* (wrong FOV placement), *mountSinai03* (T2w was re-acquired, causing wrong T2w to DWI registration), *oxfordFmrib04* (participant repositioning) and *oxfordFmrib01* (registration issue).

manufacturers. Variability of CSA across manufacturers could be due to (i) sequence parameters and/or reconstruction filters (e.g. smoothing) that alter the boundary definition, and/or (ii) differing field-strength between manufacturers (2.89 T for Siemens, 3.00 T for GE and Philips MRI) that change the apparent tissue contrasts[25].

Interestingly, the SC CSA was overall higher on T2w vs. T1w sequences (see Fig. 12). The sensitivity of image contrast to CSA measurements has already been reported in a study comparing T2w SPACE and T1w MPRAGE sequences[26], and in another study comparing T1w MPRAGE (3D-TFE) and 3D phase sensitive inversion recovery (PSIR) sequences[27]. As discussed elsewhere[28], discrepancies in measurements across MR sequences and parameters could be caused by the slightly darker contour of the T2w image, accentuating partial volume effects with the surrounding CSF, T2* blurring, Gibbs ringing, motion and flow artifacts. These differences would thus change the

identification of the SC boundaries by either the user (in case of manual segmentation) or an algorithm (in case of automated segmentation). It is worth noting that the type of MRI contrast can also impact the physical appearance of the boundaries. For example, the dura mater has a relatively short T2* value and hence its apparent location varies with the choice of TE in gradient echo sequences[29]. Age-related increases in iron deposition in the dura mater can also lead to CSA under-estimation, due to T2* reduction, which can be a confounding factor in longitudinal studies.

In order to measure CSA in retrospective or longitudinal studies, we therefore recommend sticking to exactly the same sequence and parameters. Users of our proposed protocols have the option of deriving the SC CSA from T1w or T2w images. While the two contrasts did lead to different SC CSA values, these have been modeled for each manufacturer. This means that when users compare SC CSA values that were obtained from different contrasts, they can account for differences between them by either acquiring sufficient data themselves, using our protocol and modeling the relationship between T1w and T2w SC CSA, or by using our estimated regression coefficients linking T1w CSA with T2w CSA (Fig. 12).

*Gray matter CSA.* In terms of the GM CSA, for the multi-subject dataset, GM CSAs showed a maximum inter-site COV of 5.6% (3.5% for the single subject dataset), which is highly encouraging, especially considering the small size of the GM, making CSA measures very sensitive to segmentation errors. Also worth mentioning is the inter-site standard deviation ranging from 0.64 to 0.76 $mm^2$ (0.41 to 0.57 $mm^2$ for single subject), which is remarkable considering that the effective in-plane spatial resolution of the image is $0.5 \times 0.5$ $mm^2$, i.e., the precision is roughly the size of the pixel.

Philips scanners led to significantly lower CSA values here and also larger intra-site COVs, which is likely due to the fact that some Philips sites used older versions of the consensus protocol that produced lower contrast between white and gray matter and, as a result, less reliable gray matter segmentations. The current Philips protocol has different echo times and an increased saturation band power. The latter has the effect of generating a greater MT effect and, consequently, improved WM/GM contrast. The only site benefiting from these changes was *ubc*, which explains the GM CSA values being slightly closer here to those of the Siemens and GE sites.

*Magnetization transfer.* The MT protocol includes MTR and MTsat metrics, both of which are sensitive to myelin loss[30,31]. Owing to the use of GRE-T1w images (in addition to the MT1 and MT0 scans), MTsat is less sensitive to T1 recovery effects[32,33], as has been confirmed by results from both the single- and multi-subject studies. However, this benefit is largely outweighed by it being noisier than MTR with maximum intra- and inter-site COVs of 11% and 9%, respectively, versus 4% and 2.3% for MTR. On the other hand, the higher COVs may be compensated by a higher sensitivity to myelin loss, given that myelin content appears to be more closely related to MTsat than to MTR[30]. This warrants further investigation in a patient population exhibiting abnormal myelination. Despite these somewhat discouraging results for the MTsat and T1 metrics, the GRE-T1w scan could still be kept in the *spine generic* protocol because it is short (~1 min) and could be useful for detecting hypointense lesions.

We otherwise noticed larger differences for the GE site compared to Philips and Siemens, which is likely attributed to the different MT pulse shape (Fermi for GE vs. sinc for Philips, and Gaussian for Siemens), and possibly different offset frequencies and energy. Another potential source of difference is that the acquisition matrix for the GE sites had to be reduced to 192 (instead of 256 for Philips/Siemens) because older software versions did not have ASSET (parallel imaging technique used by GE) on the GRE sequence that features the MT pulse.

*Diffusion weighted imaging.* As with MTR, DTI metrics showed very little intra- and inter-site variabilities. FA values were similar between Siemens and Philips, but significantly lower for GE. One possible explanation may lie in the different noise properties, which are known to impact DTI metrics[35]. Differences in noise properties could be related to receive coil properties, reconstruction of the images (GE data are reconstructed on a finer grid) or filters applied by the image reconstructor, among other factors. Another possible cause for the lower FA observed on the GE sites is the diffusion pulse sequence and the way diffusion gradients are played out (slew rate, mixing time, maximum gradient strength). For Siemens, the lower FA for the *vallHebron* (Tim Trio, within the $2\sigma$-$3\sigma$ interval) and *strasbourg* (Verio, within the $1\sigma$-$2\sigma$ interval) sites compared to other Siemens sites is likely caused by a much longer TE (99 ms for Trio and 95 ms for Verio, versus 55–60 ms for Skyra and Prisma), increasing noise amplitude with an impact in the DTI metrics. That said, *amu* and *beijingVerio* sites were also Verio, but their FA values were within the $1\sigma$ interval. Other DTI metrics followed the same trends in terms of intra- and inter-manufacturer variability, although COVs were higher, which could be explained by the less forgiving behaviour of these DTI metrics with respect to image quality (motion, ghosting, low signal-to-noise ratio).

Another factor which likely impacted the variability was the non-use/misuse of cardiac gating. As observed in the single subject study, DTI metrics were abnormal for sites that did not use cardiac gating, as this led to a sudden drop in signal not related to microscopic water diffusion (see Fig. 4g). The present study reiterates the benefits of cardiac gating in SC DWI experiments.

## Usage Notes

**BIDS convention.** We recommend that researchers planning to contribute to the *spine generic* database or creating other databases check the validity of the json sidecars associated with BIDS datasets. This will help assess how well protocols are followed by different centers. For json files to contain the relevant information, it is necessary that (i) DICOM fields include the relevant fields themselves, including the obvious (TR, TE, flip angle) as well as lesser known parameters that can have a strong impact on the computed metrics (water excitation, fat saturation, monopolar vs. bipolar readout, etc.), and (ii) that these fields are populated in the json files. Checking these parameters as well as the files and folder names can be automatized via continuous integration (e.g., GitHub Actions, as used in the present study).

Another advantage of the BIDS convention is that it enables the standardization of the inputs/outputs of complex analysis pipelines, or so-called "BIDS Apps" (https://bids-apps.neuroimaging.io/). For example, the proposed analysis pipeline for the *spine generic* project can be applied 'as is' to another dataset organized according to BIDS.

**Concluding remarks and future directions.** To the best of our knowledge, this study features the first "large-scale" multi-center SC qMRI datasets ever acquired and made public. These datasets are shared according to the 'Findable, Accessible, Interoperable and Reusable' (FAIR) principles[34]. The normative values from the multi-subject dataset could serve as age-matched healthy control references. More generally, these datasets will be useful for developing new image processing tools dedicated to the SC, and the fact that they are public and version-tracked with git-annex technology makes it possible for researchers to compare tools with the same data.

Lastly, important efforts were deployed to make the data analysis methods fully transparent and the results reproducible. The analysis is fully automated - aside from minor manual corrections when necessary-, minimizing user bias and facilitating large multi-center studies. We hope this analysis framework can serve as an example for future studies and we encourage researchers to use it. The SC MRI community has initiated a forum (https://forum.spinalcordmri.org/) to encourage discussions about these open-access datasets, and to pitch new ideas for subsequent analyses and acquisitions.

In a time where reproducibility of scientific results is a major concern[35], we believe a consensus acquisition protocol along with publicly-shared datasets and a transparent analysis pipeline provide a solid foundation for the field of SC qMRI so that, in the future, inclusion of the SC in neuroimaging protocols will become a "no-brainer".

## Code availability

Data were processed using Python and shell scripts contained in the spine-generic package (https://github.com/spine-generic/spine-generic/releases/tag/v2.6), which is distributed under the MIT license. A comprehensive procedure is described in the "Analysis pipeline" section of the spine generic website (https://spine-generic.rtfd.io/). This procedure includes the list of dependent software packages to install, a step-by-step analysis procedure with a list of commands to run, a procedure for quality control and for manual correction of intermediate outputs (e.g. cord segmentation and vertebral labeling). The procedure includes embedded video tutorials and has been tested by external users. The analysis documentation also includes a section on how to generate the static figures that are shown in this article (in PNG format) as well as the interactive figures embedded in the spine-generic website. Notable software used in this study include: the Spinal Cord Toolbox v5.0.1 (https://spinalcordtoolbox.com) to analyse the MRI data, pandas[36] to perform statistics, plotly v4.12.0 (https://plotly.com) to display the interactive plots, brainsprite v0.13.3 (https://brainsprite.github.io/) for embedding in the online documentation an interactive visualization of example datasets, pybids[37] for checking the acquisition parameters on the BIDS datasets, FSLeyes v0.34.0 (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLeyes) for manually-correcting the segmentations.

Received: 9 December 2020; Accepted: 26 March 2021;
Published: 16 August 2021

## References

1. Cercignani, M., Dowell, N. G. & Tofts, P. S. *Quantitative MRI of the Brain: Principles of Physical Measurement. Second edition.* (CRC Press, 2018).
2. Cohen-Adad, J. & Wheeler-Kingshott, C. Quantitative MRI of the Spinal Cord. (2014).
3. Wheeler-Kingshott, C. A. *et al.* The current state-of-the-art of spinal cord imaging: applications. *Neuroimage* **84**, 1082–1093 (2014).
4. Stroman, P. W. *et al.* The current state-of-the-art of spinal cord imaging: methods. *Neuroimage* **84**, 1070–1081 (2014).
5. Cohen-Adad, J. *et al.* Generic acquisition protocol for quantitative MRI of the spinal cord. *Nature Protocols* https://doi.org/10.1038/s41596-021-00588-0 (2021).
6. Gulban, O. F. *et al.* poldracklab/pydeface: v2.0.0. *Zenodo* https://doi.org/10.5281/zenodo.3524401 (2019).
7. De Leener, B. *et al.* SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage* **145**, 24–43 (2017).
8. Gros, C. *et al.* Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage* **184**, 901–915 (2019).
9. Ullmann, E., Paquette, P. J. F., Thong, W. E. & Cohen-Adad, J. Automatic labeling of vertebral levels using a robust template-based approach. *Int. J. Biomed. Imaging* **2014**, 719520 (2014).
10. De Leener, B. *et al.* PAM50: Unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. *Neuroimage* **165**, 170–179 (2018).
11. Perone, C. S., Calabrese, E. & Cohen-Adad, J. Spinal cord gray matter segmentation using deep dilated convolutions. *Sci. Rep.* **8**, 5966 (2018).
12. Helms, G., Dathe, H., Kallenberg, K. & Dechent, P. High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. *Magn. Reson. Med.* **60**, 1396–1407 (2008).
13. Levy, S. *et al.* White matter atlas of the human spinal cord with estimation of partial volume effect. *Neuroimage* **119**, 262–271 (2015).
14. Xu, J. *et al.* Improved in vivo diffusion tensor imaging of human cervical spinal cord. *Neuroimage* **67**, 64–76 (2013).
15. Garyfallidis, E. *et al.* Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* **8**, 8 (2014).
16. Cohen-Adad, J. Spine Generic Public Database (Single Subject). *Zenodo* https://doi.org/10.5281/zenodo.4299148 (2020).
17. Cohen-Adad, J. *et al.* Spine Generic Public Database (Multi-Subject). *Zenodo* https://doi.org/10.5281/zenodo.4299149 (2020).
18. Halchenko, Y. O. *et al.* datalad/datalad 0.12.0rc6. *Zenodo* https://doi.org/10.5281/zenodo.3512712 (2019).
19. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* **3**, 160044 (2016).
20. Papinutto, N. *et al.* Intersubject Variability and Normalization Strategies for Spinal Cord Total Cross-Sectional and Gray Matter Areas. *J. Neuroimaging* **30**, 110–118 (2020).
21. Smith, S. A. *et al.* Reproducibility of tract-specific magnetization transfer and diffusion tensor imaging in the cervical spinal cord at 3 tesla. *NMR Biomed.* **23**, 207–217 (2010).
22. Agosta, F. *et al.* Evidence for cervical cord tissue disorganisation with aging by diffusion tensor MRI. *Neuroimage* **36**, 728–735 (2007).
23. Kameyama, T., Hashizume, Y. & Sobue, G. Morphologic features of the normal human cadaveric spinal cord. *Spine* **21**, 1285–1290 (1996).

24. Papinutto, N. *et al.* Age, Gender and Normalization Covariates for Spinal Cord Gray Matter and Total Cross-Sectional Areas at Cervical and Thoracic Levels: A 2D Phase Sensitive Inversion Recovery Imaging Study. *PLoS One* **10**, e0118576 (2015).
25. Rooney, W. D. *et al.* Magnetic field and tissue dependencies of human brain longitudinal 1H2O relaxation in vivo. *Magn. Reson. Med.* **57**, 308–318 (2007).
26. De Leener, B., Kadoury, S. & Cohen-Adad, J. Robust, accurate and fast automatic segmentation of the spinal cord. *Neuroimage* **98**, 528–536 (2014).
27. Kearney, H. *et al.* Improved MRI quantification of spinal cord atrophy in multiple sclerosis. *J. Magn. Reson. Imaging* **39**, 617–623 (2014).
28. Fonov, V. S. *et al.* Framework for integrated MRI average of the spinal cord white and gray matter: The MNI-Poly-AMU template. *Neuroimage* **102P2**, 817–827 (2014).
29. Fujimoto, K. *et al.* Quantitative comparison of cortical surface reconstructions from MP2RAGE and multi-echo MPRAGE data at 3 and 7 T. *Neuroimage* **90**, 60–73 (2014).
30. Campbell, J. S. W. *et al.* Promise and pitfalls of g-ratio estimation with MRI. *Neuroimage* **182**, 80–96 (2018).
31. Schmierer, K., Scaravilli, F., Altmann, D. R., Barker, G. J. & Miller, D. H. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Ann. Neurol.* **56**, 407–415 (2004).
32. Callaghan, M. F., Helms, G., Lutti, A., Mohammadi, S. & Weiskopf, N. A general linear relaxometry model of R1 using imaging data. *Magn. Reson. Med.* **73**, 1309–1314 (2015).
33. Jones, D. K. & Basser, P. J. Squashing peanuts and smashing pumpkins: how noise distorts diffusion-weighted MR data. *Magn. Reson. Med.* **52**, 979–993 (2004).
34. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
35. Stikov, N., Trzasko, J. D. & Bernstein, M. A. Reproducibility and the future of MRI research. *Magn. Reson. Med.* **82**, 1981–1983 (2019).
36. Reback, J. *et al.* pandas-dev/pandas: Pandas 1.2.2. *Zenodo* https://doi.org/10.5281/zenodo.3509134 (2021).
37. Yarkoni, T. *et al.* PyBIDS: Python tools for BIDS datasets. *J Open Source Softw.* **4** (2019).

## Acknowledgements

We thank Gerald Moran and Bart Schraa (Siemens Healthcare), Suchandrima Banerjee and Naoyuki Take) (GE Healthcare) for sharing proprietary information and helping with setting up manufacturer-specific protocols, Nicholas Guenther and Alexandru Jora for setting up the git-annex server and procedure for hosting the public database, Charley Gros for helping with the analysis script, Paul Bautin for helping with manual corrections, Noémie Roberge for helping with the interactive plots, Carollyn Hurst, André Cyr, Arnaud Boré and Pierre Bellec (Functional Neuroimaging Unit), Charles Tremblay (Polytechnique Montreal), Antonys Melek and Habib Benali (PERFORM center, Concordia University), Ives Levesque (McGill University), Carol Lien (University of Minnesota) for helping with data acquisitions, Compute Ontario (https://computeontario.ca/) and Compute Canada (www.computecanada.ca) for providing the supercomputer infrastructure and all the volunteers who participated in the Spinal Cord MRI Public Database. This work was funded by the Canada Research Chair in Quantitative Magnetic Resonance Imaging [950-230815], the Canadian Institute of Health Research [CIHR FDN-143263], the Canada Foundation for Innovation [32454, 34824], the Fonds de Recherche du Québec - Santé [28826], the Fonds de Recherche du Québec - Nature et Technologies [2015-PR-182754], the Natural Sciences and Engineering Research Council of Canada [435897-2013], the Canada First Research Excellence Fund (IVADO and TransMedTech), the Quebec BioImaging Network [5886], Spinal Research (UK), Wings for Life (Austria, #169111) and Craig H. Neilsen Foundation (USA) for the INSPIRED project, the National Institutes of Health (NIH) through grants R00EB016689 and R01EB027779 (R.L.B.), the Instituto Investigación Carlos III (Spain, PI18/00823), the Czech Health Research Council grant n. NV18-04-00159, the Ministry of Health, Czech Republic - conceptual development of research organization (FNBr, 65269705), the National Imaging Facility and Queensland NMR Network (UQ), and SpinalCure Australia (M.J.R.), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement n° 616905; Max Planck Society and European Research Council (ERC StG 758974); European Union's Horizon 2020 research and innovation programme under the grant agreement No 681094, and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0137; BMBF (01EW1711A & B) in the framework of ERA-NET NEURON, the European Union's Horizon 2020 research and innovation programme under grant agreement No. 634541, the Engineering and Physical Sciences Research Council (R006032/1, M020533/1) and Rosetrees Trust (UK), UK Multiple Sclerosis Society (892/08, 77/2017), NIHR Biomedical Research Centres, UCLH, the Italian Ministry of Health Young Researcher Grant 2013 (GR-2013-02358177), the FISR Project "Tecnopolo di nanotecnologia e fotonica per la medicina di precisione"(funded by MIUR/CNR, CUP B83B17000010001), TECNOMED project (funded by Regione Puglia, CUP B84I18000540002), Million Dollar Bike Ride from the University of Pennsylvania (MDBR-17-123-MPS), investigator-initiated PREdICT study at the Vall d'Hebron Institute of Oncology (Barcelona) funded by AstraZeneca and CRIS Cancer Foundation, the Wellcome Trust (UK) (203139/Z/16/Z), Systems, Technologies and Applications for Radiofrequency and Communications (STARaCOM) and Swiss National Science Foundation (PCEFP3_181362/1). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## Author contributions

J. Cohen-Adad: study design and conceptualisation; data acquisition and formal data analysis; manuscript preparation and revision; data curation; management of online repositories; coordination of the multi-centre consortium. A. Foias, E. Alonso-Ortiz: study design and conceptualisation; data acquisition and formal data analysis; manuscript preparation and revision; data curation; management of online repositories. M. Abramovic, C. Arneitz, N. Atcheson, L. Barlow, R.L. Barry, M. Barth, M. Battiston, C. Büchel, M. Budde, V.Callot, A. Combes, B. De Leener, M. Descoteaux, P. L. De Sousa, M. Dostal, J. Doyon, A. Dvorak, F. Eippert, K.R. Epperson, K.S. Epperson, P. Freund, J. Finsterbusch, M. Fratini, J. Fukunaga, C.A.M. Gandini Wheeler-Kingshott, G. Germanà, G. Gilbert, F. Giove, C. Gros, F. Grussu, A. Hagiwara, P-G. Henry, T. Horák, M. Hori, J. Joers, K. Kamiya, H. Karbasforoushan, M. Kerkovsky, A. Khatibi, J.-W. Kim, N. Kinany, H.H. Kitzler, S. Kolind, Y. Kong, P. Kudlička, P. Kuntke, N.D. Kurniawan, S. Kusmia, R. Labounek, M.M. Laganà,

C. Laule, C.S. Law, C. Lenglet, T. Leutritz, Y. Liu, S. Llufriu, S. Mackey, E. Martinez-Heras, L. Mattera, I. Nestrasil, K.P. O'Grady, N. Papinutto, D. Papp, D. Pareto, T.B. Parrish, A. Pichiecchio, F. Prados, À. Rovira, M.J. Ruitenberg, R.S. Samson, G. Savini, M. Seif, A.C. Seifert, A.K. Smith, S.A. Smith, Z.A. Smith, E. Solana, Y. Suzuki, G. Tackley, A. Tinnermann, J. Valošek, D. Van De Ville, M.C. Yiannakas, K.A. Weber II, N. Weiskopf, R.G. Wise, P. O. Wyss, J. Xu: study design and conceptualisation; and data acquisition; manuscript editing and revision.

## Competing interests

Guillaume Gilbert is an employee of Philips Healthcare.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41597-021-00941-8.

**Correspondence** and requests for materials should be addressed to J.C.-A.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Julien Cohen-Adad[1,2,3✉], Eva Alonso-Ortiz[1], Mihael Abramovic[4], Carina Arneitz[4], Nicole Atcheson[5], Laura Barlow[4], Robert L. Barry[7,8,9], Markus Barth[10], Marco Battiston[11], Christian Büchel[12], Matthew Budde[13], Virginie Callot[14,15], Anna J. E. Combes[16], Benjamin De Leener[17,18], Maxime Descoteaux[19,20], Paulo Loureiro de Sousa[21], Marek Dostál[22], Julien Doyon[23], Adam Dvorak[24], Falk Eippert[25], Karla R. Epperson[26], Kevin S. Epperson[26], Patrick Freund[27], Jürgen Finsterbusch[12], Alexandru Foias[1], Michela Fratini[28,29], Issei Fukunaga[30], Claudia A. M. Gandini Wheeler-Kingshott[11,31,32], Giancarlo Germani[32], Guillaume Gilbert[33], Federico Giove[34,29], Charley Gros[1,5], Francesco Grussu[11,35], Akifumi Hagiwara[30], Pierre-Gilles Henry[36], Tomáš Horák[37], Masaaki Hori[38], James Joers[39], Kouhei Kamiya[39], Haleh Karbasforoushan[40,41], Miloš Keřkovský[22], Ali Khatibi[23,42], Joo-Won Kim[43], Nawal Kinany[44,45], Hagen H. Kitzler[46], Shannon Kolind[6,24,47], Yazhuo Kong[48,49,50], Petr Kudlička[37], Paul Kuntke[46], Nyoman D. Kurniawan[5], Slawomir Kusmia[51,52,53], René Labounek[54,55], Maria Marcella Laganà[56], Cornelia Laule[57], Christine S. Law[58], Christophe Lenglet[36], Tobias Leutritz[59], Yaou Liu[60,61], Sara Llufriu[62], Sean Mackey[58], Eloy Martinez-Heras[62], Loan Mattera[63], Igor Nestrasil[36,54], Kristin P. O'Grady[16,64], Nico Papinutto[65], Daniel Papp[1,50], Deborah Pareto[66], Todd B. Parrish[40], Anna Pichiecchio[31,32], Ferran Prados[11,12,67], Àlex Rovira[66], Marc J. Ruitenberg[68], Rebecca S. Samson[11], Giovanni Savini[32], Maryam Seif[27,59], Alan C. Seifert[43], Alex K. Smith[50], Seth A. Smith[16,64], Zachary A. Smith[69], Elisabeth Solana[62], Y. Suzuki[39], George Tackley[51], Alexandra Tinnermann[12], Jan Valošek[70], Dimitri Van De Ville[44,45], Marios C. Yiannakas[11], Kenneth A. Weber II[58], Nikolaus Weiskopf[59,71], Richard G. Wise[51,72], Patrik O. Wyss[4] & Junqian Xu[43]**

[1]NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC, Canada. [2]Functional Neuroimaging Unit, CRIUGM, Université de Montréal, Montreal, QC, Canada. [3]Mila - Quebec AI Institute, Montreal, QC, Canada. [4]Department of Radiology, Swiss Paraplegic Centre, Nottwil, Switzerland. [5]Centre for Advanced Imaging, The University of Queensland, Brisbane, Australia. [6]Department of Radiology, University of British

Columbia, Vancouver, BC, Canada. [7]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA. [8]Department of Radiology, Harvard Medical School, Boston, MA, USA. [9]Harvard–Massachusetts Institute of Technology Health Sciences & Technology, Cambridge, MA, USA. [10]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia. [11]NMR Research Unit, Queen Square MS Centre, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, UK. [12]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [13]Department of Neurosurgery, Medical College of Wisconsin, Milwaukee, WI, USA. [14]Aix-Marseille Univ, CNRS, CRMBM, Marseille, France. [15]APHM, Hopital Universitaire Timone, CEMEREM, Marseille, France. [16]Vanderbilt University Institute of Imaging Science, Vanderbilt University Medical Center, Nashville, TN, USA. [17]Department of Computer and Software Engineering, Polytechnique Montreal, Montreal, Canada. [18]CHU Sainte-Justine Research Centre, Montreal, QC, Canada. [19]Centre de Recherche CHUS, CIMS, Sherbrooke, Canada. [20]Sherbrooke Connectivity Imaging Lab (SCIL), Computer Science department, Université de Sherbrooke, Sherbrooke, Canada. [21]Université de Strasbourg, CNRS, ICube, Strasbourg, France. [22]UHB - University Hospital Brno and Masaryk University, Department of Radiology and Nuclear Medicine, Brno, Czech Republic. [23]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada. [24]Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada. [25]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. [26]Richard M. Lucas Center, Stanford University School of Medicine, Stanford, CA, USA. [27]Spinal Cord Injury Center Balgrist, University of Zurich, Zurich, Switzerland. [28]Institute of Nanotechnology, CNR, Rome, Italy. [29]IRCCS Santa Lucia Foundation, Rome, Italy. [30]Department of Radiology, Juntendo University School of Medicine, Tokyo, Japan. [31]Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy. [32]Brain MRI 3T Research Centre, IRCCS Mondino Foundation, Pavia, Italy. [33]MR Clinical Science, Philips Healthcare, Markham, ON, Canada. [34]CREF - Museo storico della fisica e Centro studi e ricerche Enrico Fermi, Rome, Italy. [35]Radiomics Group, Vall d'Hebron Institute of Oncology, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain. [36]Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, MN, USA. [37]Multimodal and functional imaging laboratory, Central European Institute of Technology (CEITEC), Brno, Czech Republic. [38]Department of Radiology, Toho University Omori Medical Center, Tokyo, Japan. [39]Department of Radiology, the University of Tokyo, Tokyo, Japan. [40]Interdepartmental Neuroscience Program, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. [41]Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA, USA. [42]Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, UK. [43]BioMedical Engineering and Imaging Institute (BMEII), Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [44]Institute of Bioengineering/Center for Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne, Geneva, Switzerland. [45]Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland. [46]Institute of Diagnostic and Interventional Neuroradiology, Carl Gustav Carus University Hospital, Technische Universität Dresden, Dresden, Germany. [47]Department Of Medicine (Neurology), University of British Columbia, Vancouver, BC, Canada. [48]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China. [49]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China. [50]Wellcome Centre For Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [51]CUBRIC, Cardiff University, Wales, UK. [52]Centre for Medical Image Computing (CMIC), Medical Physics and Biomedical Engineering Department, University College London, London, UK. [53]Epilepsy Society MRI Unit, Chalfont St Peter, UK. [54]Division of Clinical Behavioral Neuroscience, Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA. [55]Departments of Neurology and Biomedical Engineering, University Hospital Olomouc, Olomouc, Czech Republic. [56]IRCCS Fondazione Don Carlo Gnocchi ONLUS, Milan, Italy. [57]Departments of Radiology, Pathology & Laboratory Medicine, Physics & Astronomy; International Collaboration on Repair Discoveries (ICORD), University of British Columbia, Vancouver, BC, Canada. [58]Division of Pain Medicine, Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, CA, USA. [59]Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. [60]Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China. [61]Tiantan Image Research Center, China National Clinical Research Center for Neurological Diseases, Beijing, China. [62]Center of Neuroimmunology, Laboratory of Advanced Imaging in Neuroimmunological Diseases, Hospital Clinic Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) and Universitat de Barcelona, Barcelona, Spain. [63]Fondation Campus Biotech Genève, 1202, Geneva, Switzerland. [64]Department of Radiology, Vanderbilt University Medical Center, Nashville, TN, USA. [65]UCSF Weill Institute for Neurosciences, Department of Neurology, University of California San Francisco, San Francisco, CA, USA. [66]Neuroradiology Section, Vall d'Hebron University Hospital, Barcelona, Spain. [67]E-health Centre, Universitat Oberta de Catalunya, Barcelona, Spain. [68]School of Biomedical Sciences, Faculty of Medicine, The University of Queensland, Brisbane, Australia. [69]University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA. [70]Department of Neurology, Faculty of Medicine and Dentistry, Palacký University and University Hospital Olomouc, Olomouc, Czech Republic. [71]Felix Bloch Institute for Solid State Physics, Faculty of Physics and Earth Sciences, Leipzig University, Leipzig, Germany. [72]Institute for Advanced Biomedical Technologies, Department of Neuroscience, Imaging and Clinical Sciences, "G. D'Annunzio University" of Chieti-Pescara, Chieti-Pescara, Italy. [EB]e-mail: jcohen@polymtl.ca

Received: 28 November 2021 | Revised: 3 March 2022 | Accepted: 7 March 2022 | Published on: 5 April 2022

DOI: 10.1002/mrm.29249

TECHNICAL NOTE

Magnetic Resonance in Medicine

# Comparison of multicenter MRI protocols for visualizing the spinal cord gray matter

Julien Cohen-Adad[1,2,3] | Eva Alonso-Ortiz[1] | Stephanie Alley[1] | Maria Marcella Lagana[4] | Francesca Baglio[4] | Signe Johanna Vannesjo[5,6] | Haleh Karbasforoushan[7,8] | Maryam Seif[9,10] | Alan C. Seifert[11] | Junqian Xu[11] | Joo-Won Kim[11] | René Labounek[12,13] | Lubomír Vojtišek[14] | Marek Dostál[15] | Jan Valošek[12] | Rebecca S. Samson[16] | Francesco Grussu[16,17] | Marco Battiston[16] | Claudia A. M. Gandini Wheeler-Kingshott[16,18,19] | Marios C. Yiannakas[16] | Guillaume Gilbert[20] | Torben Schneider[21] | Brian Johnson[22] | Ferran Prados[16,23,24]

[1]NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, Canada

[2]Functional Neuroimaging Unit, CRIUGM, University of Montreal, Montreal, Canada

[3]Mila - Quebec AI Institute, Montreal, Canada

[4]IRCCS Fondazione Don Carlo Gnocchi ONLUS, Milan, Italy

[5]Wellcome Center for Integrative Neuroimaging, FMRIB, University of Oxford, John Radcliffe Hospital, Oxford, UK

[6]Department of Physics, Norwegian University of Science and Technology, Trondheim, Norway

[7]Interdepartmental Neuroscience Program, Northwestern University School of Medicine, Chicago, IL USA

[8]Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA USA

[9]Spinal Cord Injury Center, Balgrist University Hospital, University of Zurich, Zurich, Switzerland

[10]Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[11]Biomedical Engineering and Imaging Institute, Department of Radiology, Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY USA

[12]Departments of Neurology and Biomedical Engineering, University Hospital Olomouc, Olomouc, Czech Republic

[13]Division of Clinical Behavioral Neuroscience, Department of Pediatrics, Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, MN USA

[14]Central European Institute of Technology, Masaryk University, Brno, Czech Republic

[15]Department of Radiology and Nuclear Medicine, University Hospital Brno, Brno, Czech Republic

[16]Queen Square MS Centre, UCL Institute of Neurology, Faculty of Brain Sciences, University College London, London, UK

[17]Radiomics Group, Vall d'Hebron Institute of Oncology, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain

[18]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy

[19]Brain MRI 3T Research Center, C. Mondino National Neurological Institute, Pavia, Italy

[20]MR Clinical Science, Philips Canada, Mississauga, Canada

[21]MR Clinical Science, Philips UK, Surrey, UK

[22]MR Clinical Development, Philips North America, Gainesville, FL USA

[23]e-Health Center, Universitat Oberta de Catalunya, Barcelona, Spain

[24]Center for Medical Imaging Computing, Medical Physics and Biomedical Engineering, University College London, London, UK

**Correspondence**
Julien Cohen-Adad, Ecole Polytechnique, Pavillon Lassonde, L5610, 2700, chemin de la Tour, Montréal, QC, H3T 1J4, Canada.
Email: jcohen@polymtl.ca

**Funding information**
The Canada Research Chair in Quantitative Magnetic Resonance Imaging (950-230815), the Canadian Institute of Health Research (CIHR FDN-143263), the Canada Foundation for Innovation (32454, 34824), the Fonds de Recherche du Québec - Santé (28826), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2019-07244), the Canada First Research Excellence Fund (IVADO and TransMedTech), the Courtois NeuroMod project, the Quebec BioImaging Network (5886, 35450), the Czech Health Research Council (NV18-04-00159), the Ministry of Education Youth and Sport of the Czech Republic (LM2015062, Czech-BioImaging project) and the Ministry of Health of the Czech Republic (65269705, project for conceptual development in research organizations), EU Horizon 2020 (CDS-QuaMRI 634541), Engineering and Physical Sciences Research Council (EPSRC EP/R006032/1 and EP/I027084/1), INSPIRED (Spinal Research, UK; Wings for Life, Austria; Craig H. Neilsen Foundation, USA), UK Multiple Sclerosis Society (892/08 and 77/2017), Department of Health's NIHR BRC (R&D 03/10/RAG0449), Guarantors of Brain post-doctoral nonclinical fellowships, the US National Institute of Neurological Disorders and Stroke (NIH/NINDS K01-NS105160), and Beatriu de Pinós postdoctoral fellowships (2020 BP 00117, Secretary of Universities and Research, Government of Catalonia). The National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre.

**Purpose:** Spinal cord gray-matter imaging is valuable for a number of applications, but remains challenging. The purpose of this work was to compare various MRI protocols at 1.5 T, 3 T, and 7 T for visualizing the gray matter.

**Methods:** In vivo data of the cervical spinal cord were collected from nine different imaging centers. Data processing consisted of automatically segmenting the spinal cord and its gray matter and co-registering back-to-back scans. We computed the SNR using two methods (SNR_single using a single scan and SNR_diff using the difference between back-to-back scans) and the white/gray matter contrast-to-noise ratio per unit time. Synthetic phantom data were generated to evaluate the metrics performance. Experienced radiologists qualitatively scored the images. We ran the same processing on an open-access multicenter data set of the spinal cord MRI ($N = 267$ participants).

**Results:** Qualitative assessments indicated comparable image quality for 3T and 7T scans. Spatial resolution was higher at higher field strength, and image quality at 1.5 T was found to be moderate to low. The proposed quantitative metrics were found to be robust to underlying changes to the SNR and contrast; however, the SNR_single method lacked accuracy when there were excessive partial-volume effects.

**Conclusion:** We propose quality assessment criteria and metrics for gray-matter visualization and apply them to different protocols. The proposed criteria and metrics, the analyzed protocols, and our open-source code can serve as a benchmark for future optimization of spinal cord gray-matter imaging protocols.

**KEYWORDS**
acquisition, gray matter, image quality, MRI, protocol, spinal cord

## 1 | INTRODUCTION

Imaging the spinal cord (SC) gray matter (GM) is useful for assessing atrophy in motor-neuron diseases such as amyotrophic lateral sclerosis,[1] for studying dorsal horn atrophy in chronic pain,[2] for better characterizing lesion extent in multiple sclerosis,[3,4] or for improving the interpretation of SC functional MRI[5] or diffusion MRI.[6–9] However, proper imaging of the SC GM is difficult due to its relatively small size, and it requires high spatial resolution at the expense of a lower SNR or longer acquisition times. Moreover, images are hampered by motion (e.g., swallowing, SC motion due to CSF pulsation) and static susceptibility artifacts (induced by the presence of tissues with different susceptibility such as cartilage, bone, parenchyma, and fat), which lead to poor fat saturation, intravoxel dephasing in

gradient-recalled echo (GRE) scans, and image distortions in EPI.[10] In addition to static susceptibility effects, the $B_0$ field varies during respiration due to the change in volume and oxygenation of inhaled air. This effect becomes more prominent with increased magnetic field strength.[11,12]

The imaging protocols that are most commonly used for SC MRI, and rely on $T_1$-weighted and $T_2$-weighted ($T_1$w and $T_2$w) scans, do not provide adequate GM/white-matter (WM) contrast for GM visualization and quantification. Among the preferred sequences[13] are 2D or 3D $T_2$*-weighted ($T_2$*w) GRE and 2D T1w phase-sensitive inversion recovery. In Papinutto and Henry,[13] the authors compared different protocols for GM imaging at 3 T based on 2D phase-sensitive inversion recovery and 2D $T_2$*w sequences across Siemens, Philips and GE vendors, providing the community with a valuable starting point for making informed decisions when it comes to GM imaging. The phase-sensitive inversion-recovery protocols used in that study were based on previous experience,[14–16] and the 2D $T_2$*w protocols were obtained from the 3T cervical SC MRI spine-generic protocol.[17]

The main objective of this study is to compare various imaging protocols at 1.5 T, 3 T, and 7 T for visualizing GM. This article follows the "2018 Spinal Cord Gray Matter Imaging Challenge" that was launched at the 5th Spinal Cord MRI Workshop (http://www.spinalcordmri.org/2018/06/22/workshop.html). More specifically, this study provides (1) evaluation criteria and metrics to assess the quality of SC GM scans, (2) an open-source and automatic analysis framework for computing those metrics, (3) an open-access data set from multiple centers with suggested acquisition protocols for optimal GM visualization, (4) a comparison of those protocols using the proposed criteria and metrics, and (5) a discussion about the pros/cons of various acquisition strategies.

## 2 | METHODS

### 2.1 | Gray-matter imaging challenge: rules and data management

The GM imaging challenge called for MRI protocol design and pioneering data acquisition of SC images with high spatial resolution, minimal acquisition time, and high GM SNR and contrast. Protocol and data submission for the challenge was done on the Niftyweb platform (http://niftyweb.cs.ucl.ac.uk/program.php?p=WMGM), with the acquisition rules listed in Supporting Information Table S1. Submission is now closed, but new participants can still run the evaluation pipeline on the challenge data or on new data using the analysis scripts (see section 3.2).

To facilitate the visualization and processing of the submitted data set, and to promote reusability of open-access material, the submitted data set was anonymized and converted to the Brain Imaging Data Structure[18] and hosted on GitHub: https://github.com/sct-pipeline/gm-challenge-data. Each participant gave consent (at the center where the data were acquired) to have their data publicly accessible.

### 2.2 | Evaluation of imaging protocols

The comparison was divided into quantitative and qualitative assessments. All quantitative assessments were done automatically using the Spinal Cord Toolbox[19] and custom scripts specific to this challenge (https://github.com/sct-pipeline/gm-challenge). Qualitative assessment was done by radiologists.

#### 2.2.1 | Quantitative assessment

Acquisition time, spatial resolution, SNR, and contrast-to-noise ratio (CNR) were evaluated. Due to the difficulty in properly assessing SNR,[20] we opted for two different SNR measures: one based on a single scan (SNR_single) and another based on two scans acquired back-to-back (SNR_diff). The SNR and CNR were computed slice-wise and then averaged across slices.

SNR_single: Although, traditionally, noise is estimated in the background (air), we could not do it here because (1) some images suffered from excessive ghosting in the background, which would lead to overestimation of the noise SD; and (2) some scans were automatically thresholded (zeroed voxels in the background) by the scanner's proprietary reconstruction pipeline. Hence, we opted for computing noise in the WM to obtain a surrogate of SNR in cases where only one image was available. The WM was chosen because it pertains to the region of interest, it includes a sufficient number of voxels per slice, and the signal in this region is assumed to be homogeneous slice-wise (a requirement for spatial SD computation). The steps are:

- Automatically segment the SC[21] and the GM[22] (with manual correction when needed) and compute a WM mask by subtracting the GM from the SC mask.
- A WM mask is eroded by 1 pixel (WMe) to minimize partial volume effect.
- With $S(r)$, the MRI signal in voxel $r$, SNR_single is calculated as

$$\mathrm{SNR}_{\mathrm{single}} = \frac{\mathrm{mean}_{r\in\mathrm{WMe}}\{S(r)\}}{\mathrm{std}_{r\in\mathrm{WMe}}\{S(r)\}}.$$

There is no correction for the Rician distribution, given that the noise is computed in a region largely above the noise floor, where the distribution is closer to a Gaussian function.

SNR_diff is the difference in SNR between two scans as in Deitrich et al,[20] and it was computed as follows:

- Volume 2 is registered to volume 1 (interpolation using nearest neighbor so as to not alter noise properties).
- With $S(r,k_1)$ *and* $S(r,k_2)$, the MRI signal in voxel $r$ for volumes 1 and 2, respectively, the SNR_diff is calculated as

$$\mathrm{SNR}_{\mathrm{diff}} = \frac{\mathrm{mean}_{r\in\mathrm{WM}}\{S(r,k_1)+S(r,k_2)\}}{\sqrt{2}\cdot\mathrm{std}_{r\in\mathrm{WM}}\{S(r,k_1)-S(r,k_2)\}}.$$

The values of CNR_single and CNR_diff were calculated by multiplying the Weber contrast by SNR_single and SNR_diff, respectively. The contrast (in percent) was computed as follows: 100 * |mean(WM) − mean(GM)|/mean(WM). The CNR measures were subsequently divided by the square root of the volume acquisition time (in seconds) and are called CNR_single/$\sqrt{t}$ and CNR_diff/$\sqrt{t}$.

### 2.2.2 | Qualitative assessment

Two experienced radiologists scored four qualitative criteria (see Figure 2) for both acquisitions of each protocol. Images were presented to scorers in randomized order to minimize bias. The scoring integer scale ranged from 1 (worst) through 3 (moderate) to 5 (best). The final score for each protocol was the average of the four qualitative criteria. The median score of the two scorers was computed for each criteria as well as for the final score. The level of agreement over scorers was assessed with Spearman's rank correlation coefficient for each criteria and the final score.

## 2.3 | Comparison with the spine-generic protocol

To compare the protocols submitted to the challenge with the protocol proposed for $T_2$*w SC MRI as part of the spine-generic protocol,[17] we computed SNR_single and CNR_single/$\sqrt{t}$ metrics for $T_2$*w images of the multisubject spine-generic data set ($N = 267$, all acquired at 3 T).[23] The "diff" metrics could not be computed because the spine-generic data set only contains a single $T_2$*w scan for each subject. Due to slight differences in the spine-generic acquisition protocols across GE, Siemens and Philips scanners, the resulting metrics are clustered for each manufacturer.

## 2.4 | Simulations to assess the relevance of the evaluation metrics

To assess the relevance of the proposed metrics, we generated synthetic data of the spinal cord with varying WM/GM contrasts, noise levels and smoothing factors, as done in Levy et al.[24] Each phantom consisted of 10 slices extracted from the PAM50 template[25] centered at the mid-C4 vertebral level. The effect of spatial resolution was assessed by smoothing the phantom with a kernel of 1-mm SD. Different noise levels were then added to each phantom (additive Gaussian noise with zero mean), leading to SDs in the WM of 20, 5, and 1 and resulting in theoretical SNR_single levels of 10, 20, and 100. For both smoothed and unsmoothed phantoms, each simulated SNR level was modified so as to simulate different WM/GM contrast levels. This was done by fixing the signal value in WM to [100], while varying values in GM [120, 140, 160, 180], yielding contrasts of 20%, 40%, 60% and 80%. The signal in WM was fixed so that the SNR would be insensitive to the contrast (SNR was computed in the WM only). We then used these phantoms to assess the sensitivity and specificity of the evaluation criteria. We also assessed whether the measured contrast was insensitive to SNR and the other way around.

## 2.5 | Optimal combination of TEs in multi-echo GRE acquisitions

To test whether CNR is optimized at or near $T_2$*, we evaluated SNR, contrast, and the product of these two values (which serves as an indirect measure of CNR) in 7T GRE images from the Mount Sinai submission (9605). The TEs varied between 3 and 19 ms, at which point localized signal dropouts due to magnetic-field inhomogeneities began to encroach on the SC and in root-sum-square combinations of these images. $T_2$* values of 21.4 and 25.5 ms were calculated in WM and GM, respectively. Voxelwise maps of $T_2$* had extremely high noise, and therefore could not be accurately segmented for analysis.

The root-sum-square combination of echo images weights the contribution from each echo image by its signal intensity at each voxel, thereby maximizing SNR. However, the criteria that we intend to maximize is rather the CNR. The CNR-optimal weighting scheme would instead use the contrast or CNR of each individual echo image as the weights in a weighted sum. Four weighting schemes

were evaluated: (1) the theoretical contrast ratio, calculated as the ratio of two exponential decays having time constants equal to the $T_2$* values of WM and GM; (2) the observed contrast in the individual echo images; (3) the theoretical signal difference, calculated as the difference of the aforementioned exponential decays; and (4) the observed CNR (SNR × contrast product).

## 3 | RESULTS

The results presented here can be reproduced with the following code/data versions:

- https://github.com/sct-pipeline/gm-challenge/releases/tag/v0.5
- https://github.com/sct-pipeline/gm-challenge-data/releases/tag/r20220125
- https://github.com/spine-generic/data-multi-subject/releases/tag/r20220125

### 3.1 | Designed imaging protocols

Participating researchers designed, optimized, and submitted 13 different protocols whose data were acquired over nine MRI imaging centers. All protocols used 2D $T_2$*w imaging, except for one that made use of a 2D $T_2$*w scan with an additional inhomogeneous magnetization transfer (ihMT) prepulse to further suppress WM signal (Philips 9604). Two protocols were optimized for 1.5T MRI, six for 3T MRI, and five for 7T MRI. Each fully detailed protocol is available on the GitHub's "gm-challenge-data" repository under each subject (file name: sub-XXXX/anat/sub-XXXX_acq_params.pdf).

### 3.2 | Interprotocol comparison

For each protocol, Figure 1 shows a representative axial slice of an acquired image and its quantitative characteristics. The shortest scan time, highest SNR, contrast and CNR, per field strength, is indicated in bold.

Supporting Information Figure S1 shows a pairwise comparison of both SNR methods used in this study. On average, SNR_single is 30% smaller than SNR_diff.

Final qualitative assessment scores identified that the overall image quality is highly comparable between 3T and 7T protocols (Figure 2). The GM/WM contrast and sharpness were mostly higher for 7T scans, but increased artifacts devalued their overall image quality (Figure 2). The image quality of 1.5T protocols was less than moderate

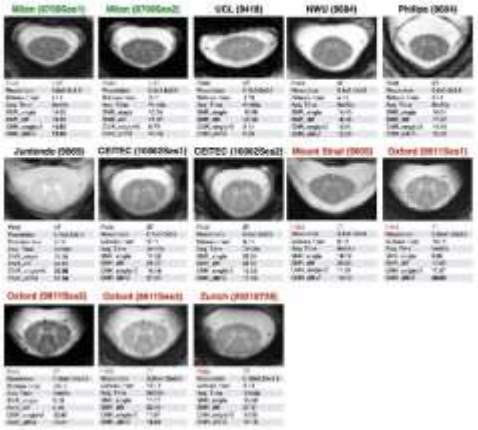


FIGURE 1 Representative images for each protocol with its quantitative assessment. Protocols are ordered by field strength: 1.5 T (green), 3 T (black), 7 T (red), and by submission ID (in brackets, next to the center). Contrast-to-noise ratios (CNRs) are expressed in percent. The best CNR per unit time, per field strength, is indicated with bold font. Each image corresponds to an axial slice centered at the C2/C3 intervertebral disc. Resolution is in millimeters. "Echoes/nav" corresponds to the number of echoes and the number of averages (combined with root sum squared except for site "Philips," where all echoes were summed). Additional quantitative metrics (contrast, CNR) can be downloaded from GitHub (https://github.com/sct-pipeline/gm-challenge/releases/download/v0.5/gmchallenge_20220114_204833.zip)

over most of qualitative assessments (Figure 2). Spearman rank correlation coefficients assessed that both scorers agreed in trends of scores over acquisitions in all qualitative assessments ($p \leq 0.009$) except the sharpness of the WM/GM border ($p = 0.114$).

### 3.3 | Comparison with the spine-generic protocol

Figure 3 shows SNR_single and contrast measured on the $T_2$*w images of the spine-generic multisubject data set. Because the gm-challenge protocols and the spine-generic protocol focused on different FOVs of different sizes, a direct and fair comparison is not fully possible. Moreover, because only one $T_2$*w scan was acquired in the spine-generic protocol, we could not compute SNR_diff. When looking at the Siemens protocol, SNR_single and contrast of the gm-challenge results mostly overlap with the Q1–Q3 interval of the spine-generic results. The Philips gm-challenge result shows better SNR_single (15.04,

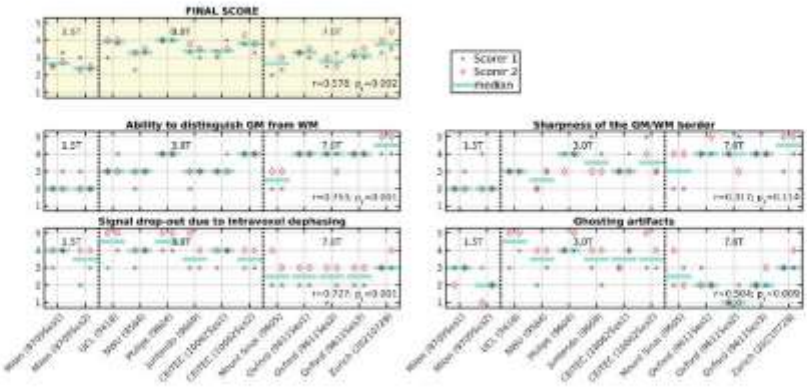



FIGURE 2 Qualitative assessment of MRI protocols. The top plot indicates the final scores for the qualitative assessment, which are taken to be the average of the four qualitative criteria shown in the remaining four plots. The y-axis is the integer score of the scale from 1 (worst) through 3 (moderate) to 5 (best). For the criteria "Signal dropout due to intravoxel dephasing," a low score means "strong signal dropout" (i.e., less signal). Criteria were assessed by two independent scorers, whose scores are indicated with unique markers at the left side for test and at right side for retest scans around the cyan-line median of all scores per data submission (i.e., four scores per submission). The value "$r$" represents the Spearman rank correlation coefficient assessing a level of agreement between two scorers. The value "$p_r$" represents the $p$-value of the correlation coefficient

above the Q1 percentile) and similar contrast (15.32, within the Q1–Q3 interval).

### 3.4 | Validation of the quality-assessment metrics

Figure 4 shows the synthetic phantoms (upper panel) and the measured contrast and SNR (lower panels). The contrast measured on the synthetic phantom showed values similar to the simulated contrast, regardless of the SNR value (Figure 4A, left). For the smoothed phantom, higher differences between simulated and measured contrast were obtained (Figure 4B, left). The measured SNR_diff was similar to the simulated SNR for each contrast (Figure 4A, middle), with a negligible difference for the smoothed phantom (Figure 4B, middle). However, the SNR_single lacked accuracy (Figure 4A, right), especially with smoothing (Figure 4B, right). This is likely due to the strong impact of partial volume effect (mixed tissue within WM mask).

### 3.5 | Optimal combination of TEs

Figure 5 shows the results of the simulation that investigated SNR, contrast, and pseudo-CNR as a function of TE. As expected, for individual images, SNR decreases and contrast increases rapidly with increasing TE. The SNR × contrast product has a broad plateau between 10 and 15 ms. For root-sum-square combinations of echo images up to a given TE (i.e., cumulative echo images), SNR is maximized at approximately 10–12 ms, whereas contrast increases with increasing TE. The SNR × contrast product for cumulative echoes also increases with increasing TE, but appears to plateau at approximately 17–19 ms. The plateaus in the SNR × contrast product for both individual and cumulative echo images suggests that factors besides $T_2^*$ and thermal noise degrade images at TEs exceeding 15–17 ms.

All four of the weighted schemes produce greater contrast than a root-sum-square combination with uniform weighting, but the root-sum-square combination with uniform weighting yields the highest CNR.

## 4 | DISCUSSION

In this article we suggest a number of criteria for evaluating SC GM MRI, and we use those criteria to assess image protocols that were submitted to the 2018 Spinal Cord Gray Matter Imaging Challenge. The imaging criteria, the analyzed protocols, and the open-source code that was developed for assessing image quality can serve the community as a benchmark for future protocol optimization. The following discussion expands on some of the strategies for helping the imaging community further optimize such protocols.

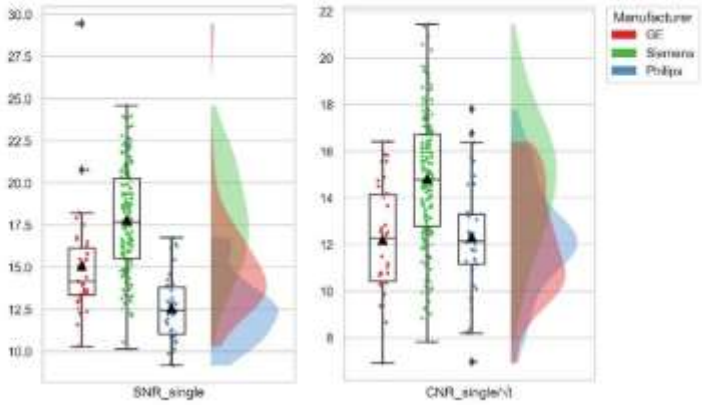



FIGURE 3 The SNR_single (left) and CNR_single/$\sqrt{t}$ (right) computed on the $T_2^*$-weighted data from the multisubject data set of the spine-generic project ($N = 267$, all acquired at 3 T). Each panel shows the individual data (plot), the median and quartiles (box plots), the mean (triangle), and the distribution (violin plot). Outliers (diamonds) are defined as being outside the $1.5 \times$ IQR. Abbreviation: IQR, interquartile range

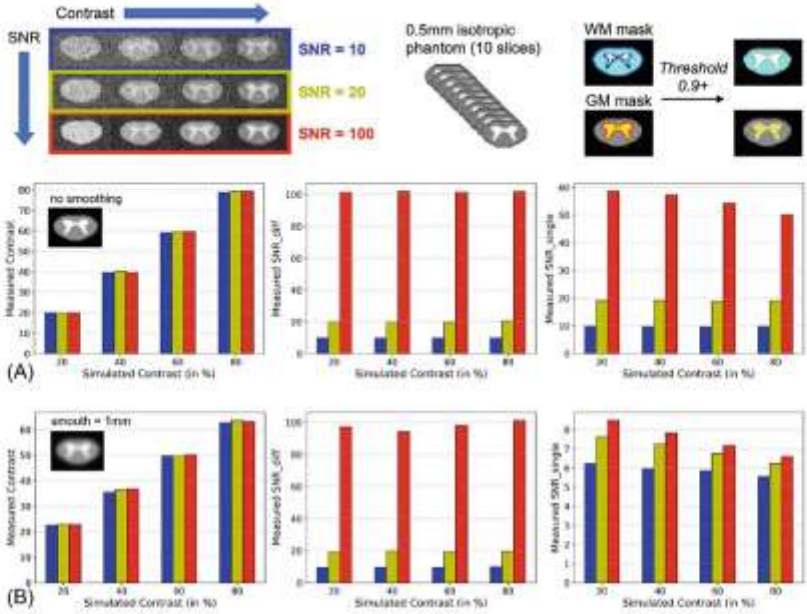
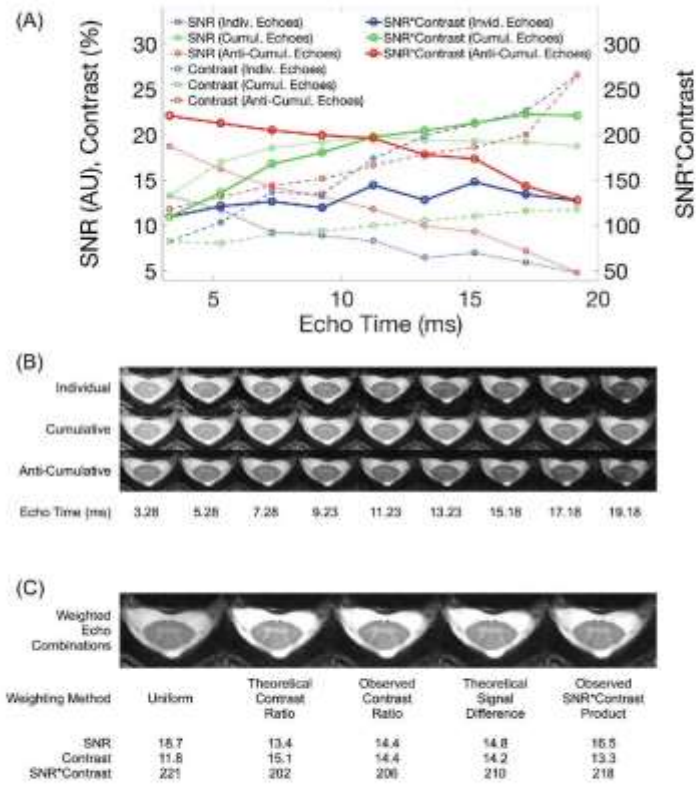



FIGURE 4 Interplay between evaluation metrics. These simulations are based on phantoms constructed with various levels of contrast (20%, 40%, 60%, and 80%) and SNR (5 [blue], 20 [yellow], and 100 [red]). White-matter (WM) and gray-matter (GM) masks derived from the PAM50 template[23] are thresholded at 0.9 to be used with the weighted-average method to extract signal in the WM and GM, respectively. Results show the evaluation metrics: "Measured Contrast," "Measured SNR_diff," and "Measured SNR_single" as a function of the simulated contrast and SNR, without (A) and with (B) 1-mm kernel smoothing

## 4.1 | Evaluation criteria

One of the difficulties in organizing this challenge was to find the right balance between harmonization/simplicity (e.g., finding a set of evaluation criteria that can apply to all participants) and exhaustiveness/rigor (i.e., making sure evaluation is accurate and fair). We acknowledge there are limitations in the current design, which are discussed subsequently.




| Weighting Method | Uniform | Theoretical Contrast Ratio | Observed Contrast Ratio | Theoretical Signal Difference | Observed SNR*Contrast Product |
|---|---|---|---|---|---|
| SNR | 18.7 | 13.4 | 14.4 | 14.8 | 16.5 |
| Contrast | 11.8 | 15.1 | 14.4 | 14.2 | 13.3 |
| SNR*Contrast | 221 | 202 | 206 | 210 | 218 |

FIGURE 5 The SNR, contrast, and pseudo-CNR (SNR*Contrast) as a function of TE. A, The values of SNR_single, WM-GM contrast, and their product (an indirect metric of CNR, which we call pseudo-CNR) are plotted against TE for individual images at given TEs (blue), for root-sum-square combinations of echo images up to a given TE ("cumulative"; green), and for root-sum-square combinations of echo images beginning with a given TE ("anti-cumulative"; red). B, A montage of the underlying individual, cumulative, and anticumulative echo images illustrates the changes in SNR and contrast as echoes are added. C, An additional montage of weighted echo combinations and their SNR, contrast, and pseudo-CNR (SNR*Contrast) is also shown

### 4.1.1 | Signal-to-noise ratio

In this study we used two different methods to compute SNR: the "diff" method, which uses the subtraction of two scans acquired back-to-back, and the "single" method, which uses a single scan in which the noise variance is computed inside the WM. On average, SNR_single was 30% smaller than SNR_diff, which is likely caused by the fact that we measured the SD of the signal within the WM, and not in a background region that contains pure noise. A region of interest within the WM may have sources of signal variance other than noise, including partial volume effects with the CSF and the GM. From Dietrich et al,[20] SNR_diff is closer to the "true" SNR (i.e., the "mult" or "nema" approach), which is also confirmed by the simulation results (Figure 4). Therefore, we considered the "diff" results from the present study to be more reliable. The "single" method has the advantage of being computed with only one scan; hence, we were able to compute SNR from a retrospective database of 267 individuals from the spine-generic project.

The SNR is directly related to the average of the magnitude image in the region of interest (in our case, the WM). Therefore, if a sequence yields low signal in the WM, the SNR will consequently be low (assuming constant noise variance). For example, let us consider two data sets (A and B) with the same noise amplitude everywhere in the image, the same mean signal in the GM, but the mean signal in the WM being lower in data set A. The SNR calculated in the GM would be the same in data sets A and B, but the SNR calculated in the WM would be lower in data set A, while the WM-GM contrast would be higher in data set A. The contrast, on the other hand, will be increased by a low value in the denominator. This is observed in the Oxford (9611Ses2) submission, which shows a relatively low SNR value in the WM, but high contrast. If SNR was measured in another region, the apparent relative performance across protocols would likely differ.

Another (related) consideration is that $T_2^*$ is driven by the field strength and the orientation of myelinated fibers.[26] Therefore, it is not surprising that some of the 7T scans show a relatively lower SNR compared with 1.5T

and 3T scans, even though higher field strength should *in principle* yield higher SNR. Moreover, to compare SNR between field strengths, one should also account for voxel volume and acquisition time. An SNR efficiency measure that corrects for those would be interesting to include in the future.

When the scanner saves its "magnitude" data, it may already be slightly filtered (e.g., using a Fermi filter to reduce ringing), which would change the inherent noise profile before the SNR is calculated. Also, the use of a multichannel coil induces spatially variant noise properties; hence, there is a bias when computing noise SD across space, as was done here. Other methods exist that are more accurate than the ones used here (e.g., acquiring two scans back-to-back, one with and the other without transmit voltage, to estimate noise SD without any bias from coil combination.[20] This method requires collecting and processing raw data, which was not done for the sake of simplicity.

#### 4.1.2 | Contrast and CNR

One of the difficulties in estimating contrast is obtaining a reliable measure of the average signal within each region, in this case WM and GM. To minimize partial volume effects, we eroded the WM mask by 1 voxel. We decided not to do the erosion for the GM mask, because this would have resulted in a very low number of remaining voxels, and hence low statistical power. If we had access to partial volume information, we could have used Gaussian mixture modeling to account for partial volume effect at the CSF/WM/GM interfaces, as was done in Levy et al.[24] Such information could be derived from a high-resolution atlas registered to each data set, and then downsampled at the native resolution of the data. This was not done here because such registration is critical, and any misregistration would yield other errors that we preferred not to address within the scope of this study. Contrast is also influenced by slice orientation. This is mostly due to partial volume effects; however, $B_0$ inhomogeneity and susceptibility differences among discs, bones, and air degrade the contrast in GRE-based sequences as well.

A study by Papinutto and Henry[13] reported an average CNR(GM/WM) of 1.56 on Siemens 3T data sets. To be able to compare this value with our results, we computed the CNR_single without normalizing by the square root of the acquisition time and without converting it into percent value. We considered only the Siemens 3T results, yielding a CNR_single(GM/WM) of $2.81 \pm 0.56$ (mean $\pm$ SD). This is slightly higher than the average value reported by Papinutto and Henry.

#### 4.1.3 | Resolution

The spatial resolution impacts the "sharpness" of an image, or our ability to distinguish between two small objects. A measure of sharpness can be obtained by computing the Laplacian of the image, then computing the mean of the Laplacian inside the SC. However, this measure is also sensitive to the noise level: the higher the noise, the higher the Laplacian. For this reason, we only considered the acquired spatial resolution (FOV divided by matrix size), although we should keep in mind that the effective resolution is also affected by the use of partial Fourier and additional filtering done by each manufacturer, even though one criteria of the challenge was specifically not to add reconstruction filters (e.g., Hanning windowing).

### 4.2 | Choice of sequence parameters

Below are some useful considerations when optimizing SC GM imaging. More details are given in the spine-generic protocol study.[17]

#### 4.2.1 | Two-dimensional versus three-dimensional

Compared with 3D imaging, multislice (2D) imaging is more robust to subject motion (if the subject moves, this will not affect the entire image), has no aliasing at the edges, and there are no issues with the $B_1^+$ profile (3D images have imperfect slab profiles creating lower flip angles at the edges, which requires one to discard 2–3 slices at the edge). On the other hand, 3D acquisitions are more SNR efficient.

#### 4.2.2 | Phase-encoding direction

Because motion is predominantly along the anterior–posterior direction, when possible, it would be preferable to phase-encode along the right–left direction. However, when imaging below the cervical cord, this becomes difficult because the shoulders and arms will alias onto the image.

#### 4.2.3 | Saturation band

The traditional purpose of saturation bands is to suppress unwanted signals, to avoid wrap-around artifacts. Because these spatial saturation pulses are usually transmitted at a

different carrier frequency, they produce a slight magnetization transfer effect, which in turn alters WM/GM contrast. Therefore, they could be used to enhance WM/GM contrast, assuming that the magnetization-transfer effect suppresses signal from WM more than from GM, and that the main contrast is $T_2$*-like (i.e., brighter GM).

### 4.2.4 | Optimal combination of TEs

Most of the submitted protocols relied on $T_2$*w imaging with multiple TEs. In $T_2$*w image acquisitions, knowledge of the $T_2$* relaxation times of the two tissue types whose contrast is to be optimized can aid in the creation of an imaging protocol. Although SNR is highest at the shortest TE, $T_2$* contrast increases with increasing TE. However, in practice, neither SNR nor contrast should be optimized in isolation. Instead, efforts should be made to optimize the CNR or CNR per unit time. Under a simplistic assumption of pure thermal noise, CNR was shown to be optimized at TE = $T_2$*.[27] Other factors, such as magnetic-field inhomogeneity and limitations on total scan time, may favor shorter TE, as does the increased physiological noise at higher TE.[12] The latter factor may explain why the root-sum-square echo combination, which upweights early echoes, was here observed to have higher CNR than contrast-weighted echo combinations, which should theoretically be optimal under pure thermal noise.

An additional consideration in multi-echo GRE sequences is the choice of monopolar versus bipolar readout. Bipolar readouts allow for TEs to be spaced more closely, yielding increased SNR, but may result in different patterns of spatial distortion between even and odd echoes (positive and negative readouts) due to background magnetic-field inhomogeneities (https://raw.githubusercontent.com/sct-pipeline/gm-challenge/master/doc/fig_monopolar_bipolar.gif). The misregistration between the even and odd echoes would introduce blurring when combining all echoes. Monopolar readouts produce a set of echoes with compatible patterns of spatial distortion at some cost to SNR and CNR.

## ACKNOWLEDGMENT

The authors thank Alexandru Foias and Nicolas Pinon for helping with the generation of figures, and Pavla Hanzliková (from the Department of Radiology, University of Ostrava, Czechia) for helping with the qualitative assessment.

## ORCID

*Julien Cohen-Adad* https://orcid.org/0000-0003-3662-9532
*Eva Alonso-Ortiz* https://orcid.org/0000-0001-6590-7234
*Signe Johanna Vannesjo* https://orcid.org/0000-0003-2432-4192
*Haleh Karbasforoushan* https://orcid.org/0000-0003-2032-9703
*Alan C. Seifert* https://orcid.org/0000-0001-7877-4813
*René Labounek* https://orcid.org/0000-0003-0439-1304
*Jan Valošek* https://orcid.org/0000-0002-7398-4990
*Francesco Grussu* https://orcid.org/0000-0002-0945-3909
*Marco Battiston* https://orcid.org/0000-0003-2231-2251

## TWITTER

*Julien Cohen-Adad* @jcohenadad

## REFERENCES

1. Paquin M-È, El Mendili MM, Gros C, Dupont SM, Cohen-Adad J, Pradat P-F. Spinal cord gray matter atrophy in amyotrophic lateral sclerosis. *AJNR Am J Neuroradiol*. 2018;39:184-192.
2. Jutzeler CR, Huber E, Callaghan MF, et al. Association of pain and CNS structural changes after spinal cord injury. *Sci Rep*. 2016;6:18534.
3. Calabrese M, Favaretto A, Martini V, Gallo P. Grey matter lesions in MS: from histology to clinical implications. *Prion*. 2013;7:20-27.
4. Agosta F, Pagani E, Caputo D, Filippi M. Associations between cervical cord gray matter damage and disability in patients with multiple sclerosis. *Arch Neurol*. 2007;64:1302-1305.
5. Kornelsen J, Mackey S. Potential clinical applications for spinal functional MRI. *Curr Pain Headache Rep*. 2007;11:165-170.
6. Wheeler-Kingshott CA, Stroman PW, Schwab JM, et al. The current state-of-the-art of spinal cord imaging: applications. *Neuroimage*. 2014;84:1082-1093.
7. By S, Smith AK, Dethrage LM, et al. Quantifying the impact of underlying measurement error on cervical spinal cord diffusion tensor imaging at 3T. *J Magn Reson Imaging*. 2016;44:1608-1618.
8. Massire A, Rasoanandrianina H, Taso M, et al. Feasibility of single-shot multi-level multi-angle diffusion tensor imaging of the human cervical spinal cord at 7T. *Magn Reson Med*. 2018;80:947-957.
9. Labounek R, Valošek J, Horák T, et al. HARDI-ZOOMit protocol improves specificity to microstructural changes in presymptomatic myelopathy. *Sci Rep*. 2020;10:17529.
10. Cohen-Adad J, Wheeler-Kingshott C. *Quantitative MRI of the Spinal Cord*. Cambridge, Massachusetts: Academic Press; 2014.
11. Verma T, Cohen-Adad J. Effect of respiration on the B0 field in the human spinal cord at 3T. *Magn Reson Med*. 2014;72:1629-1636.
12. Vannesjo SJ, Miller KL, Clare S, Tracey I. Spatiotemporal characterization of breathing-induced B0 field fluctuations in the cervical spinal cord at 7T. *Neuroimage*. 2018;167:191-202.
13. Papinutto N, Henry RG. Evaluation of intra- and interscanner reliability of MRI protocols for spinal cord gray matter and total cross-sectional area measurements. *J Magn Reson Imaging*. 2019;49:1078-1090.

14. Papinutto N, Schlaeger R, Panara V, et al. Age, gender and normalization covariates for spinal cord gray matter and total cross-sectional areas at cervical and thoracic levels: a 2D phase sensitive inversion recovery imaging study. *PLoS One*. 2015;10:e0118576.
15. Papinutto N, Datta E, Zhu AH, et al. Multisite feasibility study of spinal cord gray matter and total cord areas measurements on 2D phase sensitive inversion recovery images. *Proceedings of the 24th Annual Meeting ISMRM*. Singapore; 2016; 698-708.
16. Papinutto N, Schlaeger R, Panara V, et al. 2D phase-sensitive inversion recovery imaging to measure in vivo spinal cord gray and white matter areas in clinically feasible acquisition times. *J Magn Reson Imaging*. 2015;42:698-708.
17. Cohen-Adad J, Alonso-Ortiz E, Abramovic M, et al. Generic acquisition protocol for quantitative MRI of the spinal cord. *Nat Protoc*. 2021;16:4611-4632.
18. 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.
19. De Leener B, Lévy S, Dupont SM, et al. SCT: spinal cord toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage*. 2017;145:24-43.
20. Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *J Magn Reson Imaging*. 2007;26:375-385.
21. Gros C, De Leener B, Badji A, et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage*. 2019;184:901-915.
22. Perone CS, Calabrese E, Cohen-Adad J. Spinal cord gray matter segmentation using deep dilated convolutions. *Sci Rep*. 2018;8:5966.
23. Cohen-Adad J, Alonso-Ortiz E, Abramovic M, et al. Open-access quantitative MRI data of the spinal cord and reproducibility across participants, sites and manufacturers. *Sci Data*. 2021;8:219.
24. Levy S, Benhamou M, Naaman C, Rainville P, Callot V, Cohen-Adad J. White matter atlas of the human spinal cord with estimation of partial volume effect. *Neuroimage*. 2015;119:262-271.
25. De Leener B, Fonov VS, Collins DL, Callot V, Stikov N, Cohen-Adad J. PAM50: unbiased multimodal template of the brainstem and spinal cord aligned with the ICBM152 space. *Neuroimage*. 2018;165:170-179.
26. Cohen-Adad J, Polimeni JR, Helmer KG, et al. T2* mapping and B0 orientation-dependence at 7T reveal cyto- and myeloarchitecture organization of the human cortex. *Neuroimage*. 2012;60:1006-1014.
27. Van de Moortele P-F, Ugurbil K, Lehericy S. Is T2* always the optimum echo time in BOLD fMRI? Challenging a common concept with a new contrast to noise ratio BOLD model. *Proceedings of the 16th Annual Meeting of ISMRM*; ISMRM; 2008; 133-139.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**Table S1.** Acquisition criteria for submitting data to the challenge. Acronyms: Signal-to-noise ratio (SNR), T2-weighted (T2w), T2*-weighted (T2*w), field of view (FOV), superior-inferior (S-I), difference (diff). (*): The submission "Juntendo (9669)" used 5 mm slices.

**Figure S1.** Pairwise comparison of the two SNR methods used in this study, showing data acquired at 1.5 T (triangle), 3 T (circle) and 7T (cross). The SNR_diff method uses the subtraction of two scans acquired back-to-back, and the SNR_single method uses a single scan where the noise variance is computed inside the WM. The dashed line corresponds to no difference between the two methods.

**How to cite this article:** Cohen-Adad J, Alonso-Ortiz E, Alley S, et al. Comparison of multicenter MRI protocols for visualizing the spinal cord gray matter. *Magn Reson Med*. 2022;88:849-859. doi: 10.1002/mrm.29249

ORIGINAL RESEARCH

# Analysis of Diffusion Tensor Measurements of the Human Cervical Spinal Cord Based on Semiautomatic Segmentation of the White and Gray Matter

Marek Dostál, MS,[1,2] Miloš Keřkovský, PhD,[2*] Eva Koriťáková, PhD,[3] Eva Němcová, MD,[2] Jakub Stulík, MD,[2] Monika Staňková, MD,[2] and Vladan Bernard, PhD[1]

**Background:** Segmentation of the gray and white matter (GM, WM) of the human spinal cord in MRI images as well as the analysis of spinal cord diffusivity are challenging. When appropriately segmented, diffusion tensor imaging (DTI) of the spinal cord might be beneficial in the diagnosis and prognosis of several diseases.
**Purpose:** To evaluate the applicability of a semiautomatic algorithm provided by ITK-SNAP in classification mode (CLASS) for segmenting cervical spinal cord GM, WM in MRI images and analyzing DTI parameters.
**Study Type:** Prospective.
**Subjects:** Twenty healthy volunteers.
**Sequences:** 1.5T, turbo spin echo, fast field echo, single-shot echo planar imaging.
**Assessment:** Three raters segmented the tissues by manual, CLASS, and atlas-based methods (Spinal Cord Toolbox, SCT) on $T_2$-weighted and DTI images. Masks were quantified by similarity and distance metrics, then analyzed for repeatability and mutual comparability. Masks created over $T_2$ images were registered into diffusion space and fractional anisotropy (FA) values were statistically evaluated for dependency on method, rater, or tissue.
**Statistical Tests:** *t*-test, analysis of variance (ANOVA), coefficient of variation, Dice coefficient, Hausdorff distance.
**Results:** CLASS segmentation reached better agreement with manual segmentation than did SCT ($P < 0.001$). Intra- and interobserver repeatability of SCT was better for GM and WM (both $P < 0.001$) but comparable with CLASS in entire spinal cord segmentation ($P = 0.17$ and $P = 0.07$, respectively). While FA values of whole spinal cord were not influenced by choice of segmentation method, both semiautomatic methods yielded lower FA values ($P < 0.005$) for GM than did the manual technique (mean differences 0.02 and 0.04 for SCT and CLASS, respectively). Repeatability of FA values for all methods was sufficient, with mostly less than 2% variance.
**Data Conclusion:** The presented semiautomatic method in combination with the proposed approach to data registration and analyses of spinal cord diffusivity can potentially be used as an alternative to atlas-based segmentation.
**Level of Evidence:** 1
**Technical Efficacy:** Stage 2

J. MAGN. RESON. IMAGING 2018;48:1217–1227.

Using such conventional magnetic resonance imaging (MRI) techniques as $T_1$- and $T_2$-weighted images, radiologists are able to discern gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). These conventional methods are insufficient, however, for imaging WM's internal structure.[1] WM is an organized fibrous structure that results in water diffusion becoming anisotropic and with preferred diffusion along the direction of the fibers. One

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.26166

Received Jan 18, 2018, Accepted for publication Apr 10, 2018.

*Address reprint requests to: M.K., Department of Radiology and Nuclear Medicine, University Hospital Brno and Masaryk University, Jihlavská 340/20, 625 00 Brno, Czech Republic. E-mail: kerkovsky.milos@fnbrno.cz

From the [1]Department of Biophysics, Faculty of Medicine, Masaryk University, Brno, Czech Republic; [2]Department of Radiology, University Hospital Brno and Masaryk University, Brno, Czech Republic; and [3]Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Brno, Czech Republic

**TABLE 1. Imaging Protocol**

| Type | Orientation | TR [msec] | TE [msec] | FOV [mm] | Matrix | Slice thickness [mm] | Other information |
|---|---|---|---|---|---|---|---|
| $T_1$-w | Sagittal | 400 | 7.8 | 255×255×33 | 528×528 | 3.3 | TSE, Avg = 4 |
| $T_2$-w | Sagittal | 3,500 | 120 | 255×255×33 | 528×528 | 3.3 | TSE, Avg = 8 |
| $T_2$-w* | Axial | 334 | 9.21 | 170×170×56 | 432×432 | 4 | FFE, Avg = 4 |
| DTI* | Axial | 3,200 | 92.3 | 170×170×56 | 192×192 | 4 | SS EPI, 1 b = 0, 15 dir (b = 800s/mm$^2$), Avg = 6 |

T1-weighted (T1-w), T2-weighted (T2-w), diffusion tensor imaging (DTI), repetition time (TR), echo time (TE), field of view (FOV), turbo spin echo (TSE), fast field echo (FFE), single shot echo planar imaging (SS EPI), average (Avg).
*Performed in two separate continuing acquisitions from C1 to C3/C4 and from C3/C4 to C7 with emphasis on maximal perpendicularity to the spinal cord and while being attentive to minimizing gap or minimal overlap of these FOVs. T2-w and DTI had the same center of FOV (Fig. 1).

method currently used for assessing the quality of WM's internal structure is diffusion tensor imaging (DTI). In pathological states, WM fibers are degraded, thereby disturbing their structural integrity and resulting in more isotropic water diffusion (decrease in anisotropy). DTI is used, for example, in studies of brain tumors, multiple sclerosis (MS), epilepsy, ischemic stroke, as well as tumors and other lesions of the spinal cord.[2]

A necessary and frequently crucial step in analyzing DTI images of the central nervous system is segmentation of WM, GM, and CSF. There exist many methods for segmenting brain images,[3] but there are far fewer methods for spinal cord segmentation.[4] Most of these methods segment the entire spinal cord (ESC) and CSF at various levels of automation (semiautomatic or fully automatic) and from various image modalities ($T_1$, $T_2$, DTI, and others).[5–11] Only a few methods are able to segment not just the ESC but also WM and GM.[7,9,12–19]

The main objective of this study was to evaluate the possibilities for using a semiautomatic segmentation method (CLASS) based on a semisupervised machine-learning technique implemented in ITK-SNAP[20–24] for segmentation of MRI data of the cervical spinal cord and to utilize this technique for analyzing diffusion parameters of different spinal cord tissues.

## Materials and Methods

The study group consisted of 20 volunteers (16 women and 4 men) aged 23 to 40 years (mean age of 28.05 and standard deviation of 4.6 years) in whom no pathological spinal cord changes were found by an experienced neuroradiologist (M.K.). All volunteers signed informed consent agreements to participate in the study, which was approved by the University Hospital's Ethics Committee.

MRI data were acquired using a 1.5T MR (Philips Achieva, Best, Netherlands) with a 16-channel head and cervical coil. The scanning protocol is shown in Table 1. $T_2$-weighted fast field echo and DTI sequences had the same geometry covering in two parts spinal cord segments C1–C3/C4 and C3/C4–C7, while taking into consideration the overlap of areas around the C3/C4 disc and the best perpendicularity to the spinal cord of both parts (Fig. 1).

The first step in data analysis was to join upper and lower $T_2$ and DTI images into the same space orientation. Coordinates of the spinal canal were manually set on the two images (upper and lower), rigid registration was performed, and one image of the entire uninterrupted cervical spine was created. Based on sagittal images, only axial slices ranging from the level of the posterior arch of C1 to the cranial endplate of the C7 vertebral body were used for all other steps. Functional MRI of the Brain Software Library (FSL) was used for processing all images.[25] ITK-SNAP v. 3.4 for manual and CLASS segmentation[20,23,24] (http://www.

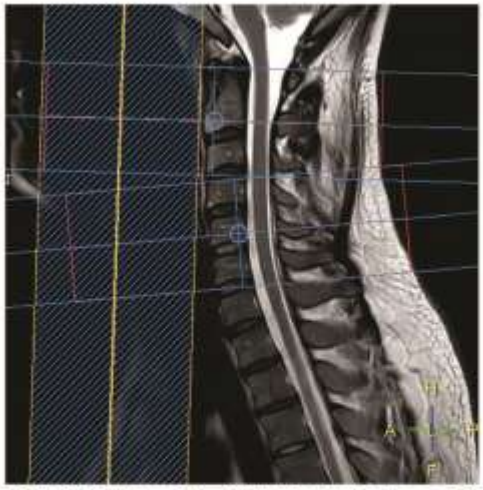


**FIGURE 1: Example of $T_2$-weighted fast field echo and DTI sequences planning on the $T_2$-weighted image in sagittal plane. Axial $T_2$ and DTI sequences had the same geometry covering in two parts spinal cord segments C1–C3/C4 and C3/C4–C7, while taking into consideration the overlap of areas around the C3/C4 disc and the best perpendicularity to the spinal cord of both parts.**

itksnap.org/pmwiki/pmwiki.php) and Spinal Cord Toolbox (SCT)[26] (https://sourceforge.net/projects/spinalcordtoolbox/) as a representative of atlas-based segmentation methods.

All segmentations were done by three raters and three methods, as shown below. Rater 1 (M.D.) trained segmentations with ITK-SNAP for at least 20 hours. Raters 2 (E.N.) and 3 (J.S.), both with 5 years of practice in radiology, trained for 3 hours and 1 hour, respectively. Rater 1's manual segmentations were supervised by a neuroradiologist with more than 12 years of MRI practice (M.K.) and were regarded as a reference standard. All raters are coauthors of this article.

Manual segmentation of the ESC, GM, and WM on $T_2$ images was performed independently by three raters using ITK-SNAP software for the entire group and then three more times on seven randomly selected subjects (two men and five women) at minimum 2-day intervals between these assessments. Only ESC was manually segmented on DTI (FA) images in the same design as $T_2$ images. During manual segmentation, raters paid attention to the best possible tissue separation to minimize the contamination with CSF. $T_2$ images were registered into DTI space by identical matrix. The center of gravity of each slice of the binary mask was calculated for both $T_2$ and DTI masks, and the centers of gravity of both images were registered by a simple 2D translation algorithm. Then 3D nonrigid registration was applied. Each step of the registration was repeated for GM and WM binary masks using exactly the same registration matrices as were calculated when ESC was registered. This method enables determination of FA values for ESC, GM, and WM.

ESC segmentation by SCT (sct_propseg algorithms) was performed in three different initialization settings with constant parameters (radius 4; detect-n 4; detect-gap 4; nbiter 200; max-area 120; max-deformation 2.5; min-contrast 50). The first setting was in default mode, which is fully automated (Default); the second was in three-point mode, wherein raters manually set three points within the central canal of the spinal cord on three different axial slices (3 Point); and the third was in CenterLine mode, where raters manually marked the position of the spinal cord central canal by one voxel on all axial slices (CenterLine). WM and GM were segmented using the sct_segment_graymatter algorithm under default settings, and the threshold of the obtained probabilistic tissue masks was set to 0.5.

ESC segmentation by ITK-SNAP was performed by a classification method (CLASS), which consists of three steps. The first step (presegmentation) includes manual labeling of voxels of two or more different classes (in our case this means ESC as the first class and the nearest surrounding like CSF or vertebrae as the second class) on three slices (second, middle, and penultimate). Special care was taken to avoid contamination of the ESC area with voxels of surrounding tissues at the borderline zone to decrease the partial volume effect. These manually defined volumes were subsequently used as a training set for classification of the tissues within an individual subject. Based on this training the contextual information about intensity of neighboring voxels and coordinates of voxels from multiple image layers were derived. These data, which allow for correctly classifying structures without marked image contrast due to their different texture, are used as an input for a random forest classifier[24] and geodesic active contour method.[27,28] The contour represents closed surface, which is evolving according to partial differential equation, where internal forces (derived from contour's geometry) and external forces (given by the image information) affect the contour evolution. For more mathematical background concerning the methods used, see Yushkevich et al,[20] Yushkevich and Gerig,[25] Yushkevich et al,[24] and Caselles et al.[27] The number of trees (100 trees) as well as tree depth (50 trees) was constant for each run. An output of the presegmentation step is "speed image," which is a product of the speed function as one component of evolving force. Its value is close to 0 at the edges of intensity in the input image or close to 1 in regions where intensity is homogeneous. The second step initializes the segmentation by manual placement of the seed points over the spinal cord area on every even-numbered slice. In the last step (evolution), the user sets the weights from the active contour equation, which affects the expansivity and smoothness of segmentation masks, and iteratively runs the evolution. The number of iteration steps varied around 100, and each iteration took around 30 seconds to complete.

Several preprocessing steps were performed before segmenting GM and WM by ITK-SNAP. The ratio of GM and WM areas on axial scans is ~20:80,[29,30] and the signal intensity of GM is higher than that of WM on $T_2$-weighted images. We therefore assumed that 20% of ESC voxels with the highest intensity corresponded mainly in GM. Voxels with extremely high and low intensity were "homogenized" by applying an upper threshold and threshold, respectively (Eq. 1):

$$I_{new} = \begin{cases} I_{Q15} & I \le I_{Q15} \\ I & I_{Q15} < I < I_{Q90} \\ I_{Q90} & I \ge I_{Q90} \end{cases} \quad (1)$$

where $I_{new}$ is the voxel value after homogenization; $I_{Q15}$ and $I_{Q90}$ are values of the 15% and 90% intensity quartiles, respectively; and I is the current value of the given voxel. This step ensured that 15% of the least intensive and 10% of the most intensive voxels have the same respective intensity values (ie, they are homogenous). Subsequently, all voxels were modified according to the formula:

$$I_{new} = \left(\frac{I - I_{Q82.5}}{10}\right)^2 \quad (2)$$

where $I_{new}$ is the calculated voxel value, $I_{Q82.5}$ is the value of the 82.5% intensity quartile, and I is the current value of the given voxel. This step ensured that the intensity of voxels with a large probability of being GMs is close to zero and all others have higher values (ie, the background has a value equal to zero). This step makes segmentation easier because the algorithm does not need to recognize three different tissues (WM, GM, background) with similar intensities but only two tissues (WM, the rest) (Fig. 2B).

The segmentation of GM and WM was similar to that of ESC. Raters manually segmented GM, WM, and also a small area of background around ESC on three slices of the original $T_2$ image. The algorithm, however, uses the information from both images (original $T_2$ and the modified one, which was described in a previous paragraph) as a teaching feature. Iterated segmentation was done afterwards, and the GM and background masks were

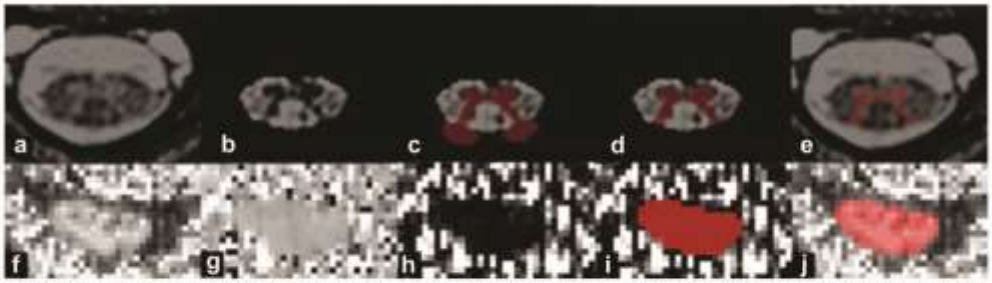



**FIGURE 2: Original $T_2$ image of spinal cord (A), preprocessed image before GM segmentation by ITK-SNAP (B), result of segmentation by classification algorithm for two tissues (C), final GM segmentation on preprocessed image (D), and on original image (E). FA image of spinal cord in different subjects (F), projection image of the main diffusion tensor vector onto the x-axis (G), squared projection image of the main diffusion tensor onto the x-axis (H), final entire spine cord segmentation on image H (I), and on FA image (J) with good result including blurred area.**

obtained (Fig. 2C). Only voxels included within the ESC mask were considered to be GM, however (Fig. 2D). ESC voxels, which are not classified as GM, represent WM. WM and GM masks were registered into the DTI space in the same way as described for the manual segmentation method.

When preprocessing DTI data, we applied eddy current correction (FSL) and diffusion tensor was fitted by dtifit script (FSL) for upper and lower images. After that, both parts of the cervical spinal cord were connected with the same transformation matrix as in the case of the $T_2$ image, and thus a whole cervical spinal cord DTI image was created. For manual segmentation, an image of FA values was used (Fig. 2F). The FA image and squared projection image of the main diffusion tensor vector (MDTV) onto the x-axis (marked by FSL as $V_1$) were used for segmenting ESC and CSF using the CLASS method.

The original $V_1$ image (Fig. 2G) was squared to remove the negative sign and sharpen edges between ESC and CSF (Fig. 2H). Raters manually segmented ECS on the FA image, while the CLASS algorithm uses both FA and $V_1$ images for teaching, which improves the segmentation accuracy. The segmentation mask was shown in parallel on both the FA and $V_1$ images (Fig. 2I,J) for visual examination.

Two metrics were used to verify segmentation agreement: similarity and distance. The 3D Dice coefficient (DC) was used to represent the similarity metric:

$$DC = \frac{2(S_1 \cap S_2)}{|S_1| + |S_2|} \quad (3)$$

where $S_1$ and $S_2$ are counts of voxels acquired by segmentations 1 and 2, respectively. The numerator denotes double the number of voxels which have both segmentations in common, and the denominator defineates the total combined voxel count of the two segmentations. The result ranges from 0 to 1, where 1 indicates identical segmentation and 0 indicates absolutely different segmentation. The DC coefficient is highly sensitive to the total number of voxels, which has a larger impact on evaluations of small structures like GM than on large structures.[31]

The distance metric is described by the Hausdorff distance (HD):

$$HD = \max\Big(h(A,B), h(B,A)\Big) \qquad h(A,B) = \max_{a \in A} \min_{b \in B} \|a-b\| \quad (4)$$

where h(A,B) is the direct Hausdorff distance between finite point sets A and B and $\|a-b\|$ is the Euclidean distance of two points a and b from point sets A and B, respectively.[32] HD is expressed in millimeters and describes maximal imprecision of the two segmentations. Both metrics were performed using EvaluateSegmentation script.[31]

The first step in evaluating the segmentations was a mutual comparison of the CLASS and SCT methods (in three different modes) with manual segmentation of the most experienced (gold standard) operator for ESC segmented on the $T_2$ image. The two best modalities, SCT CenterLine mode and CLASS method, were then statistically evaluated and compared with manual segmentation in more detail for both image contrasts ($T_2$ and FA) and for different tissues (WM and GM). Furthermore, intra- and interobserver repeatability were evaluated for segmentation of all tissues using manual and both semiautomatic segmentation techniques (SCT, CLASS). Finally, the influences of rater and segmentation method on FA values of different spinal cord tissues and their repeatability were appraised.

For evaluating intraobserver repeatability, images of seven randomly chosen subjects were segmented four times by all three raters using different methods: manual segmentation, SCT (CenterLine), and CLASS. Masks obtained by a single rater from a single subject using a particular segmentation method were paired off in all possible combinations, representing six independent pairs (combination number $\binom{4}{2}=6$). For each method, 126 DC and HD were calculated (seven subjects, three raters, six combinations) and mutually evaluated by *t*-test separately for every tissue.

In the case of interobserver repeatability, masks obtained by segmentation in all 20 subjects by all three raters were paired off in all three possible combinations. All in all, we obtained 60 DC and HD for every method and tissue.

Binary masks of ESC, GM, and WM created based on images of 20 subjects by CLASS and manual segmentation methods were registered into the DTI space. Because SCT masks were

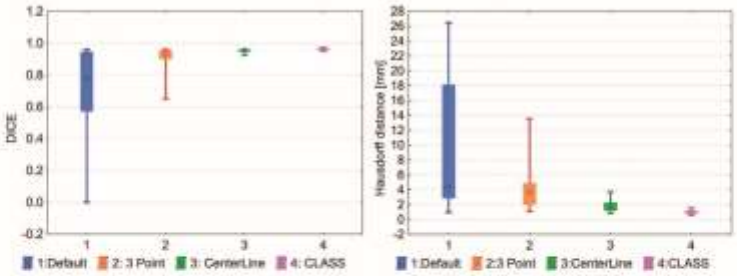



**FIGURE 3: Mutual comparison of four semiautomatic segmentations of entire spinal cord on $T_2$-w image with manual segmentation represented by 3D Dice coefficient and Hausdorff distance on box-and-whisker plots (min, 25% quartile, median, 75% quartile, max). Boxes of different colors represent agreement of different semiautomatic methods with manual segmentation from the perspective of 3D Dice coefficient and Hausdorff distance values. Default: Spinal Cord Toolbox (SCT) in Default mode, 3 Point: SCT three-point initialization method, CenterLine: SCT CenterLine method, and CLASS: ITK-SNAP classification method.**

generated directly in DTI space, they did not need registration. Analysis of variance (ANOVA) was used to compare how application of different segmentation methods performed by various raters impacts FA values of different tissues. Factorial ANOVA was performed with FA values as dependent variables and with raters, methods, and tissues as categorical factors.

FA's intraobserver repeatability was verified on seven subjects, each segmented four times by three raters and three methods. An intraobserver coefficient of variation (CoV) was calculated based on four median FA values of the same subject, method, and rater. In the case of interrepeatability, median FA values obtained from 20 subjects, three raters, and three methods without repetition were used. Inter-CoV was calculated based on three median FA values for the same subject, method, and different rater. CoV values of different methods for individual tissues were mutually compared using *t*-tests.

### Statistics

Student's *t*-test was used for statistical comparison of different segmentation methods quantified by DC and HD values as well as for comparison of repeated segmentations in terms of inter- and intraobserver repeatability evaluations. ANOVA and Tukey's post-hoc test were used to compare FA values of various tissues obtained by different methods and raters and CoV was used to quantify the repeatability of FA values obtained by means of different segmentation techniques.

$$CoV(x) = SD(x)/\bar{x} \quad (5)$$

CoV is a ratio of the standard deviation (SD) of quantity x and the mean value of this quantity ($\bar{x}$). It is expressed as a percentage. The lower the value, the better is the consistency.

Statistical tests were computed using Statistica 12 (StatSoft, Tulsa, OK) software, graphs were created with the same software, DC and HD were performed using EvaluateSegmentation script[31] and for all statistical tests the significance level was set to 0.05.

## Results

### Mutual Comparison of the Methods

Rater 1 segmented the ESC of all 20 subjects by all five methods (manually, CLASS, and SCT in three different settings). DC and HD were calculated for semiautomated methods compared to manual segmentation (Fig. 3). Based on these results, and inasmuch as the other two modes corresponded unsatisfactorily with manual segmentation, for further segmentations we used SCT only in the CenterLine mode.

A detailed comparison of the two most promising semiautomatic methods with manual segmentation done by Rater 1 was determined for ESC segmented on $T_2$ and FA images and for GM and WM segmented on $T_2$ images in all 20 subjects. CLASS segmentation performed by a skilled rater generally resulted in statistically significantly more masks being similar to manual segmentation than did SCT segmentation in CenterLine mode for both $T_2$ and FA images (Fig. 4).

### Intraobserver Repeatability

This section evaluates the consistency of different segmentation methods repeated by the same rater on the same subject (Table 2). For ESC and WM segmentation, both semiautomated methods are more consistent than is manual segmentation performed by all raters (Fig. 5). This is demonstrated by the significantly higher DC values and lower HD values achieved by the semiautomatic methods as compared to the manual method. In the case of GM segmentation, semiautomated methods also yielded higher DC values, while CLASS also produced higher HD values than both manual and SCT. In nine cases, the SCT method segmented ESC on the FA image inaccurately, thereby resulting in lower DC and higher HD values (outliers in Fig. 5).

### Interobserver Repeatability

The consistency of segmentation methods across different raters is shown here (Table 2). In the case of ESC segmentation, both semiautomated methods produce more consistent results across raters (Fig. 6). SCT is the most consistent method for GM and WM segmentation across raters. The CLASS method is slightly poorer than manual segmentation for WM and poorer when GM is segmented.

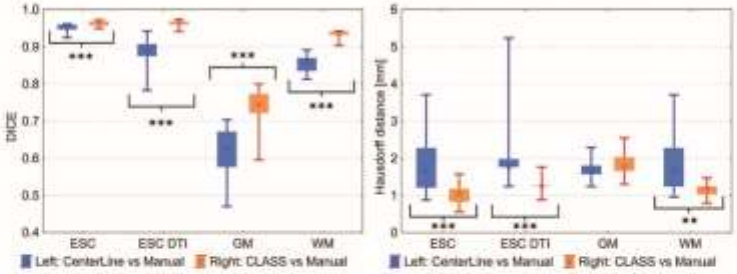



FIGURE 4: Comparison of entire spinal cord (ESC) segmented on $T_2$ data, ESC on DTI data, gray (GM) and white matter (WM) segmented on $T_2$ data in all 20 subjects is represented by 3D Dice coefficient and Hausdorff distance on box-and-whisker plots (min, 25% quartile, median, 75% quartile, max). The blue boxes represent a comparison of the Spinal Cord Toolbox (SCT) method with the CenterLine seed mask against manual segmentation as the gold standard. The orange boxes compare ITK-SNAP classification method (CLASS) with manual segmentation. A *t*-test was performed to evaluate statistical differences between agreements of different methods. $^{**}P<0.001$ and $^{***}P<0.0001$. The pair with no asterisk do not differ statistically ($P>0.05$).

**TABLE 2. Overview of Intra- and Interrepeatability of Three Segmentation Methods and Four Different Segmentations Described by 3D Dice Similarity Coefficient and Hausdorff Distance**

| | | Median DICE (IQR) | | | | Median HD [mm] (IQR) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | ESC | ESC DTI | GM | WM | ESC | ESC DTI | GM | WM |
| Intrarepeatability | Man | 96.5(2.0) | 96.6(2.3) | 80.3(8.9) | 91.9(5.1) | 0.79(0.3) | 0.89(0.4) | 1.15(0.4) | 0.88(0.2) |
| | SCT | 98.7(0.7) | 99.0(0.4) | 87.3(4.5) | 95.9(1.6) | 0.96(0.4) | 0.89(0) | 0.96(0.4) | 0.96(0.4) |
| | CLASS | 99.0(1.2) | 97.6(2.1) | 82.3(17.3) | 95.9(2.9) | 0.56(0.5) | 0.89(0.4) | 1.55(1.4) | 0.92(0.3) |
| Interrepeatability | Man | 95(2.1) | 95.5(1.7) | 70.1(6.5) | 86.8(3.1) | 0.88(0.2) | 1.25(0.4) | 1.62(0.3) | 0.96(0.2) |
| | SCT | 98.7(0.9) | 98.8(0.8) | 87(5.8) | 95.8(1.6) | 0.96(0.9) | 0.88(0.4) | 1.04(0.4) | 0.96(0.6) |
| | CLASS | 98.7(1.3) | 96.7(2.4) | 65.4(19.3) | 87.5(3.5) | 0.57(0.3) | 0.88(0.4) | 2.56(1.1) | 1.18(0.2) |

The interquartile range represents the distance between the 25% and 75% quartiles. Manual method (Man), Spinal Cord Toolbox in CenterLine mode (SCT), ITK-SNAP in classification mode (CLASS), entire spinal cord (ESC), gray (GM), white matter (WM), entire spinal cord segmented in DTI data (ESC DTI), 3D Dice similarity coefficient (DICE), Hausdorff distance (HD), and interquartile range (IQR).

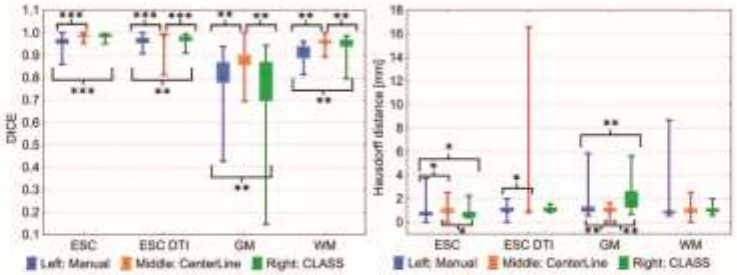



FIGURE 5: Intrarepeatability of three segmentation methods performed by three raters, four times on seven subjects and three tissues (entire spinal cord on $T_2$-w [ESC] and on DTI [ESC DTI], gray [GM] and white matter [WM]). All possible pair-combinations of segmented masks were created, and 1134 3D Dice coefficients (DICE) and Hausdorff distances (HD) were calculated. Box-and-whisker plots show the results (min, 25% quartile, median, 75% quartile, max) and each box represents 126 coefficients. Blue boxes represent intrarepeatability of manual segmentation; orange boxes show results of Spinal Cord Toolbox (SCT) method with CenterLine seed mask; and green boxes represent ITK-SNAP classification method (CLASS). A *t*-test was performed to evaluate statistical differences between agreements of different methods. $^{*}P<0.01$, $^{**}P<0.001$, and $^{***}P<0.0001$. Pairs with no asterisk did not differ significantly ($P>0.05$).

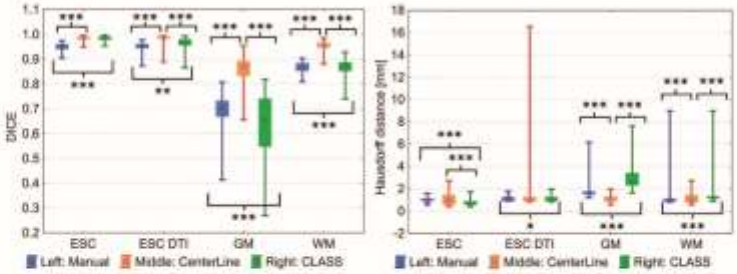



FIGURE 6: Interrepeatability of three segmentation methods performed by three raters on 20 subjects and four tissues (entire spinal cord on $T_2$-w [ESC] and DTI [ESC DTI], gray [GM] and white matter [WM]). All possible pair combinations of segmented masks were created and 720 3D Dice coefficients (DICE) and Hausdorff distances were calculated. Box-and-whisker plots show results (min, 25% quartile, median, 75% quartile, max), and each box is assembled from 60 coefficients. Blue boxes represent interrepeatability of manual segmentation; orange boxes show results of the Spinal Cord Toolbox (SCT) method with CenterLine seed mask; and green boxes indicate the ITK-SNAP classification method (CLASS). A *t*-test was performed to evaluate statistical differences between agreements of different methods. *$P<0.01$, **$P<0.001$, and ***$P<0.0001$. Pairs with no asterisk did not differ significantly ($P>0.05$).

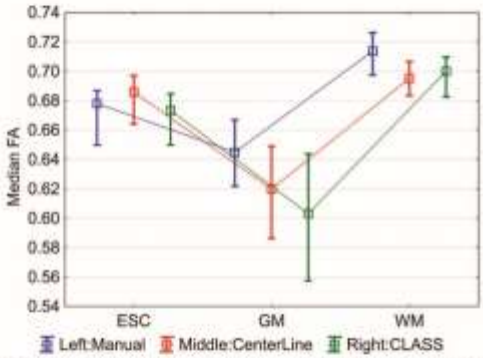



FIGURE 7: Tukey's post-hoc test of method–tissue interaction term shows significantly higher median FA value obtained by manual segmentation of gray matter (GM) in comparison with both semiautomated methods ($P<0.005$). FA value differences equal 0.02 (CenterLine) and 0.04 (CLASS). Boxes indicate median FA values and whiskers mark the 25% and 75% quartiles. The FA values of entire spinal cord (ESC) and white matter (WM) are not significantly dependent on segmentation method or rater.

### *Evaluation of FA Values*

We showed no dependence of FA values on rater ($F[2,513] = 0.22$, $P = 0.806$), but the effects of method, tissue, and their interaction term are statistically significant ($F[2,513] = 13.68$, $P < 0.0001$; $F[2,513] = 198.49$, $P < 0.0001$; and $F[4,513] = 5.92$, $P < 0.005$, respectively). No other effect was statistically significant. Tukey's post-hoc test was performed on the method–tissue interaction term (Fig. 7) and shows that the manual segmentation of GM was significantly different from both other methods ($P < 0.005$). Divergence between median FA values of manual and semiautomated methods were equal to 0.02 (CenterLine) and 0.04 (CLASS). The semiautomated methods show very good agreement for all tissues.

Repeatability of FA values for all methods was sufficient, with mostly less than 2% variance. SCT produced the most consistent results and the repeatability of the manual method was similar to that of the CLASS method (Table 3).

**TABLE 3. Median Values of Coefficient of Variation Describe Intrarepeatability of Three Segmentation Methods on Three Tissues Done by Three Different Raters in Four Repetitions on Seven Subjects and Interrepeatability on 20 Subjects Without Repetition**

| CoV [%] Intrarepeatability (IQR) | | | | CoV [%] Interrepeatability (IQR) | | | |
|---|---|---|---|---|---|---|---|
| Tissue\Method | Manual | CenterLine | CLASS | Tissue\Method | Manual | CenterLine | CLASS |
| ESC | 1.4(0.5)[@] | 0.1(0.04) | 1(0.6)[@] | ESC | 2.2(1.7)* | 0.1(0.1) | 1.3(1.5)* |
| WM | 1(0.7) | 0.1(0.04) | 0.4(0.2) | WM | 1.3(1.2)[+] | 0.1(0.2) | 1.6(1.0)[+] |
| GM | 1.2(0.4) | 0.2(0.3) | 1.8(1.1) | GM | 1.8(1.2) | 0.4(0.3) | 3.8(3.4) |

Paired CoV values marked with *, +, and @ symbols are not significantly different (*t*-test, $P > 0.05$); other pairs are significantly different ($P < 0.05$). Coefficient of variation (CoV), entire spinal cord (ESC), gray matter (GM), white matter (WM), Spinal Cord Toolbox segmentation in CenterLine mode (CenterLine), ITK-SNAP in classification mode (CLASS), interquartile range (IQR).

## Discussion

In this article we introduced the application of ITK-SNAP segmentation in classification mode on the structural and DTI MRI data of cervical spinal cord. Requirements for this method included a capability for accurate segmentation of the ESC area, separate segmentation of WM and GM from $T_2$-weighted images, as well as segmentation of the ESC area from DTI images and differentiation of its contours from CSF. We focused on the clinical applicability and employment of the method for analysis of the diffusion parameters of spinal cord.

There exist several approaches to spinal cord segmentation. Basically, three main groups of segmentation methods further divided into various subgroups can be differentiated.[4] In older studies, there generally predominated use of intensity-based methods such as thresholding, edge detection, or intensity-based classifiers.[31,33–35] Other authors used surface-based segmentation like active contour or deformable models[8,10,17,36,37] and the last main group adopted such image-based methods as graph-cut, atlas deformation, or classifiers.[7,15,16] From the perspective of this classification we may consider ITK-SNAP as a hybrid method using some of the components from all three of these groups due to its employment of active contour algorithms, random forest classifier, and texture information about intensity and voxel coordinates. This approach has the advantage that it does not use atlas deformable algorithms. That makes this method more personalized, eliminates the risk of misregistration, and may also make it more accurate in specific pathological conditions when the spine is deformed or abnormal, eg, in case of severe spinal cord compression. Work-flow of the segmentation in ITK-SNAP also enables easy separate segmentation of the visible spinal cord lesions, which can further be used for separate analysis of diffusion parameters within these areas. On the other hand, this method may be more sensitive to image quality compared to atlas-based models and could yield poorer results in situations where the contrast- or signal-to-noise ratio is poor. In such conditions even an experienced radiologist may find it difficult to perform manual segmentation and the use of an atlas-based method may be beneficial.

ITK-SNAP was introduced as a general segmentation tool with a user-friendly interface and was originally evaluated on segmentation of caudate nucleus.[20] Since that time, its use has been reported in more than 1700 articles.[24] The first classification method based on random forests is available from ITK_SNAP v. 3.2 (January 2015) without implementation of texture information. This additional technique was added into v. 3.4 (January 2016) and it allows advanced learning methods such as multimodal segmentation, whereby learning algorithms take information from several different images for better classification.[23,24] As far as we are aware, these semiautomatic techniques have not yet been systematically evaluated for the segmentation of human spinal cord.

Older studies[33,34,36,37] and studies dealing with DTI[11,15,35] mostly use MR with 1.5T induction, taking advantage of lower susceptibility artifacts in DTI acquisition. Of the aforementioned studies, the CLASS method presented in this article achieves the most accurate results in segmenting the entire spinal cord from a $T_2$ image. The more recent studies and studies dealing with WM and GM segmentation[5,7,9,14–19,38] use mostly MR with 3T induction. $T_2$-weighted images from such MR devices have better signal-to-noise ratios, and thus the contrast between GM and WM is sharper and the segmentation is more accurate. This may be an advantage especially in examining lower cervical spine, where the image quality is usually poorer than in the case of upper cervical segments. On the other hand, DTI data are more affected by susceptibility artifacts. Our presented method is therefore not the most accurate in segmenting GM and WM, but it is still among the most accurate methods published to date.

It should be noted, however, that the comparison of different segmentation methods based on DC and HD coefficients with literature data is complicated due to the varying acquisition parameters and image quality. Therefore, using the same dataset, we decided to perform direct comparison of the newly applied CLASS method with the better-established and commonly published SCT technique. A similar approach was used in a study by Prados et al,[38] who compared the capabilities of different segmentation methods on dedicated MR data. That study provides a broader overview of different up-to-date techniques. The methods generally reached median DC values for GM segmentation between 0.6 and 0.85 and median HD values ranged from 1.5 to 7 mm in comparison to manual segmentation. Compared to this, the benefit of our study may be seen especially in the extension of the segmentations also on the DTI data to connect anatomical information with diffusion. Thus, we can introduce a comprehensive and clinically usable methodological approach of cervical cord DTI data analysis. In any case, the CLASS method may merit further evaluation as part of the ongoing GM segmentation challenge project[38] in order to be compared with more techniques than just SCT.

Some publications[10,14,15,18,33,36,37] quantify segmentation repeatability by comparing areas or volumes of segmentation masks or on the basis of various derived coefficients (coefficient of variation or intraclass correlation coefficients [ICC]). We do not consider such approaches to be optimal for quantifying the segmentation method, because congruence of areas or volumes does not automatically imply congruent segmentations. Therefore, we believe it is more suitable to use a similarity metric (Dice coefficient, Jaccard

coefficient, global consistency error, or the like) together with a distance metric (such as Hausdorff distance or Mahalanobis distance), which correlate mutually as little as possible and thus do not yield redundant information.[31]

As for DTI data analysis, we evaluated the reproducibility and influence of different segmentation techniques on FA, which we chose as a representative scalar parameter quantifying anisotropic diffusion. FA values measured within ESC and WM appeared to be very consistent among different raters and methods, as ANOVA analysis revealed no significant differences. Some inconsistency was observed in the case of FA values of GM, where manual segmentations provided slightly, but statistically significantly, higher FA values as compared to both semiautomatic methods. Given that FA of normal GM is physiologically lower than that of WM,[39] the higher FA values of GM measured within the masks of manual segmentation may be attributed to incorrect inclusion of a greater number of WM voxels by less experienced raters. From this perspective, measurement of lower GM FA by both semiautomatic methods may be considered more plausible.

To measure FA within WM and GM, we used a technique of 2D registration of segmentations on $T_2$ and FA images using a spinal cord contour as a landmark. Due to the large variability in spinal cord curvature, lack of unambiguous landmarks, and small dimensions, affine or nonrigid 3D registration is almost impossible.[5] Therefore, we had to apply 2D rigid registration for the entire spinal cord contour segmented from particular axial scans and applied the resulting transformation matrix on the masks of individual tissues. For the lowest possible deviations of $T_2$ and DTI images, emphasis was given to congruence of their geometric acquisition parameters. Visual examination was also performed by overlapping the images along the z-axis in the sagittal plane. Perpendicularity of axial sections is necessary for maximizing contrast between WM and GM in a $T_2$ image and minimizing the partial volume effect in a DTI image.[40] The acquisition was therefore divided into two parts, with emphasis on maximum perpendicularity of the individual sections and avoiding mutual overlap of these acquisitions.

The technique described above did not reveal significant differences of measured FA values for GM, WM, and ESC between the CLASS and CenterLine SCT methods. This result indicates that our approach may provide a functional alternative to a more established atlas-based technique for quantifying diffusion parameters. Both methods revealed statistically significant differences in FA between WM and GM. This supports the accuracy of the measurements, as it is difficult to establish the gold standard of WM and GM segmentations in DTI data by visual control due to comparatively low contrast and resolution.

The inter- and intraobserver variability of FA expressed by CoV was generally low among all techniques, which denotes the generally good reproducibility of this biomarker. However, the variability of FA measured using the CenterLine method showed the lowest rate. Minor manual input combined with the atlas-based technique is probably the reason for similarly better agreement of the segmentation masks themselves between different raters and repeated measurements compared to the CLASS method.

This work has several limitations that originate in part from the intention to examine the clinical applicability of the method. One of those limitations is the setting of fixed SCT parameters, which may influence the results. We decided on a constant setting in order to maximize the benefits of the automatic method with relatively minimal input from the operator. Moreover, the degree to which a common user will comprehend all of the parameters may not be sufficiently high to enable precise optimization of all parameters for individual segmentations. As discussed above, the use of a 1.5T magnetic field could constitute another limitation.

The time aspect and labor intensity could also play a role in determining the practical applicability of this method. The mean duration of segmentation in one subject by rater MD performed on seven subjects was estimated at 635 seconds for SCT and 644 seconds for the CLASS method. Although the segmentation of all spinal cord tissues (with visual inspection of each segmentation mask, but without any postprocessing or manual corrections of segmentation masks) by SCT requires less time than does the CLASS method; the difference is not so great as one might expect when considering the higher degree of automatization in the case of SCT. This may be caused by the need for extra checking of the particular steps in SCT processing, which are done automatically by the CLASS method during the segmentation procedures. Moreover, creating CenterLine masks for $T_2$ and FA images in SCT is quite time-consuming. Although there can be no doubt that the CLASS method is more laborious than is SCT, as it uses more manual input from the operator, this disadvantage may be offset in part by creating scripts to increase the efficiency of the method. On the other hand, when the segmentation is not sufficiently accurate, the CLASS method allows straightforward and prompt manual correction immediately after the iteration step. That is in contrast to SCT, which requires setting up appropriate threshold values for cutting off less probable voxels of GM and WM, and uses an additional software tool to make the appropriate corrections manually. In order more objectively to evaluate the performance of the two semiautomatic techniques in this study, however, no manual corrections to either CLASS or SCT segmentations were made.

Although another potential limitation may be seen in the relatively low number of subjects, the power of statistical analyses evaluating the accuracy of segmentations was generally sufficient due to multiple repetitions by several operators. As the machine-learning algorithm implemented in the CLASS method is individually based, the total number of subjects analyzed is not really relevant from this perspective. Inasmuch as spine curvature is relatively variable, however, a larger study group could cover a wider spectrum of anatomical configurations and potentially provide more precise results.

The choice as to the number of voxels marked for the learning algorithm could also constitute a source of some inaccuracy. In this study, we used information from three defined slices at upper, middle, and lower cervical levels. This number of slices was set empirically based on preliminary testing, and it was chosen as the best compromise between accuracy and the time needed for processing. Nevertheless, in patients with severe spinal deformity or alteration of the spinal cord signal intensity, it may be beneficial to provide a higher number of teaching masks or to place them at different spinal cord segments to achieve the best results from the semiautomatic segmentation. Inasmuch as the algorithm uses voxel coordinates and intensity of neighboring voxels as a teaching feature, a small spinal cord lesion or mild compression at the level of the teaching masks should not have substantial negative impact on the resulting segmentation, but the performance of the CLASS technique in such pathological conditions is yet to be established.

To conclude, the ITK-SNAP semiautomatic segmentation technique using machine learning is exploitable for the segmentation of a human cervical spinal cord in structural and DTI MRI data and constitutes a convenient alternative to an atlas-based segmentation method. While SCT provides rather more reproducible segmentations, the CLASS technique, on the other hand, revealed better agreement of segmentation masks with manual segmentation by the most experienced rater. Furthermore, the introduced technique for registration of DTI and $T_2$ images appears to be applicable for measurements of diffusion scalar parameters within different tissues of the spinal cord, thus providing results corresponding to those of a more established atlas-based technique used by SCT.

## Acknowledgments

Contract grant sponsor: Czech Health Research Council; contract grant number: AZV-15-32133A; Contract grant sponsor: Grant Agency of Masaryk University; contract grant number: MUNI/A/1464/2014

## References

1. Dong Q, Welsh RC, Chenevert TL, et al. Clinical applications of diffusion tensor imaging. J Magn Reson Imaging 2004;19:6–18.
2. Lemer A, Mogensen MA, Kim PE, Shimishi MS, Hwang DH, Law M. Clinical applications of diffusion tensor imaging. World Neurosurg 2014;82:96–109.
3. Balafar MA, Ramli AR, Saripan MI, Mashohor S. Review of brain MRI image segmentation methods. Artif Intell Rev 2010;33:261–274.
4. Leener BD, Taso M, Cohen-Adad J, Callot V. Segmentation of the human spinal cord. Magn Reson Mater Phys Biol Med 2016;29:125–153.
5. Asman AJ, Bryan FW, Smith SA, Reich DS, Landman BA. Groupwise multi-atlas segmentation of the spinal cord's internal structure. Med Image Anal 2014;18:460–471.
6. Chen M, Carass A, Cuzzocreo J, Bazin P-L, Reich DS, Prince JL. Topology preserving automatic segmentation of the spinal cord in magnetic resonance images. IEEE 2011:1737–1740.
7. Chen M, Carass A, Oh J, et al. Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. NeuroImage 2013;83:1051–1062.
8. De Leener B, Kadoury S, Cohen-Adad J. Robust, accurate and fast automatic segmentation of the spinal cord. NeuroImage 2014;98:528–536.
9. El Mendili M-M, Chen R, Tiret B, et al. Fast and accurate semi-automated segmentation method of spinal cord MR images at 3T applied to the construction of a cervical spinal cord template. PLoS One 2015;10:e0122224.
10. Horsfield MA, Sala S, Neema M, et al. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: Application in multiple sclerosis. NeuroImage 2010;50:446–455.
11. Tang L, Wan Y, Zhou Z, von Deneen KM, Huang D, Ma L. Reduced field-of-view DTI segmentation of cervical spine tissue. Magn Reson Imaging 2013;31:1507–1514.
12. Asman AJ, Smith SA, Reich DS, Landman BA. Robust GM/WM segmentation of the spinal cord with iterative non-local statistical fusion. Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv 2013;16(Pt 1):759–767.
13. Ellingson BM, Ulmer JL, Schmit BD. Gray and white matter delineation in the human spinal cord using diffusion tensor imaging and fuzzy logic. Acad Radiol 2007;14:847–858.
14. Yiannakas MC, Kearney H, Samson RS, et al. Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: A pilot study with application to magnetisation transfer measurements. NeuroImage 2012;63:1054–1059.
15. Bergo FPG, França MC, Chevis CF, Candela F. SpinaSeg: A segmentation and measurement tool for evaluation of spinal cord atrophy. In: 2012 7th Iber Conf Inf Syst Technol CISTI; 2012:1–4.
16. Fonov VS, Le Troter A, Taso M, et al. Framework for integrated MRI average of the spinal cord white and gray matter: The MNI–Poly-AMU template. NeuroImage 2014;102:817–827.
17. Koh J, Scott PD, Chaudhary V, Dhillon G. An automatic segmentation method of the spinal canal from clinical MR images based on an attention model and an active contour model. In: 2011 IEEE Int Symp Biomed Imaging Nano Macro; 2011:1467–1471.
18. Taso M, Le Troter A, Sdika M, et al. A reliable spatially normalized template of the human spinal cord — Applications to automated white matter/gray matter segmentation and tensor-based morphometry (TBM) mapping of gray matter alterations occuring with age. NeuroImage 2015;117:20–28.
19. Yiannakas MC, Mustafa AM, De Leener B, et al. Fully automated segmentation of the cervical cord from T1-weighted MRI using PropSeg: Application to multiple sclerosis. NeuroImage Clin 2016;10:71–77.
20. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. NeuroImage 2006;31:1116–1128.
21. Breiman L. Random forests. Mach Learn 2001;45:5–32.
22. Criminisi A, Shotton J. Semi-supervised classification forests. In: Decis For Comput Vis Med Image Anal. Criminisi A, Shotton J (eds.).

Advances in Computer Vision and Pattern Recognition. London: Springer; 2013:95–107.

23. Yushkevich PA, Gerig G. ITK-SNAP: An interactive medical image segmentation tool to meet the need for expert-guided segmentation of complex medical images. IEEE Pulse 2017;8:54–57.

24. Yushkevich PA, Gao Y, Gerig G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: 2016 38th Annu Int Conf IEEE Eng Med Biol Soc EMBC; 2016: 3342–3345.

25. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. NeuroImage 2012;62:782–790.

26. De Leener B, Lévy S, Dupont SM, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. NeuroImage 2017;145(Pt A):24–43.

27. Caselles V, Kimmel R, Sapiro G. Geodesic active contours. Int J Comput Vis 1997;22:61–79.

28. Zhu SC, Yuille A. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans Pattern Anal Mach Intell 1996;18:884–900.

29. Goto N, Otsuka N. Development and anatomy of the spinal cord. Neuropathology 1997;17:25–31.

30. Fradet L, Arnoux P-J, Ranjeva J-P, Petit Y, Callot V. Morphometrics of the entire human spinal cord and spinal canal measured from in vivo high-resolution anatomical magnetic resonance imaging. Spine 2014; 39:E262–E269.

31. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015;15: 29.

32. Taha AA, Hanbury A. An efficient algorithm for calculating the exact Hausdorff distance. IEEE Trans Pattern Anal Mach Intell 2015;37: 2153–2163.

33. Lossseff NA, Webb SL, O'Riordan J, et al. Spinal cord atrophy and disability in multiple sclerosis. Brain 1996;119:701–708.

34. Tench CR, Morgan PS, Constantinescu CS. Measurement of cervical spinal cord cross-sectional area by MRI using edge detection and partial volume correction. J Magn Reson Imaging 2005;21:197–203.

35. Zivadinov R, Banas AC, Yella V, Abdelrahman N, Weinstock-Guttman B, Dwyer MG. Comparison of three different methods for measurement of cervical cord atrophy in multiple sclerosis. Am J Neuroradiol 2008;29:319–325.

36. Coulon O, Hickman SJ, Parker GJ, Barker GJ, Miller DH, Arridge SR. Quantification of spinal cord atrophy from magnetic resonance images via a B-spline active surface model. Magn Reson Med 2002; 47:1176–1185.

37. Van Uitert R, Bitter I, Butman JA. Semi-automatic spinal cord segmentation and quantification. CARS 2005: Computer Assisted Radiology and SurgeryProceedings of the 19th International Congress and Exhibition. Int Congr Ser 2005;1281:224–229.

38. Prados F, Ashburner J, Blaiotta C, et al. Spinal cord grey matter segmentation challenge. NeuroImage 2017;152(Supplement C):312–329.

39. Feldman HM, Yeatman JD, Lee ES, Barde LHF, Gaman-Bean S. Diffusion tensor imaging: a review for pediatric researchers and clinicians. J Dev Behav Pediatr JDBP 2010;31:346–356.

40. Ellingson BM, Cohen-Adad J. Chapter 3.1. Diffusion-weighted imaging of the spinal cord. In: Quant MRI Spinal Cord. San Diego: Academic Press; 2014:123–145.

Magnetic Resonance Imaging 73 (2020) 23–30

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

Original Contribution

# Voxelwise analysis of diffusion MRI of cervical spinal cord using tract-based spatial statistics

Marek Dostál[a,b], Miloš Keřkovský[a,*], Erik Staffa[b], Josef Bednařík[c], Andrea Šprláková-Puková[a], Marek Mechl[a]

[a] Department of Radiology and Nuclear Medicine, University Hospital Brno and Masaryk University, Czech Republic
[b] Faculty of Medicine, Department of Biophysics, Masaryk University, Brno, Czech Republic
[c] Department of Neurology, University Hospital Brno and Masaryk University, Czech Republic

ARTICLE INFO

*Keywords:*
Diffusion tensor imaging
Spinal cord
Voxelwise analysis

ABSTRACT

Robust voxelwise analysis using tract-based spatial statistics (TBSS) together with permutation statistical method is standardly used in analyzing diffusion tensor imaging (DTI) of brain. A similar analytical method could be useful when studying DTI of cervical spinal cord.

Based on anatomical data of sixty-four healthy volunteers, white (WM) and gray matter (GM) masks were created and subsequently registered into DTI space. Using TBSS, two skeleton types were created (single line and dilated for WM as well as GM). From anatomical data, percentage rates of overlap were calculated for all skeletons in relation to WM and GM masks.

Voxelwise analysis of fractional anisotropy values depending on age and sex was conducted. Correlation of fraction anisotropy values with age of subjects was also evaluated. The two WM skeleton types showed a high overlap rate with WM masks (~94%); GM skeletons showed lower rates (56% and 42%, respectively, for single line and dilated). WM and GM areas where fraction anisotropy values differ between sexes were identified ($p < .05$). Furthermore, using voxelwise analysis such WM voxels were identified where fraction anisotropy values differ depending on age ($p < .05$) and in these voxels linear dependence of fraction anisotropy and age ($r = -0.57$, $p < .001$) was confirmed by regression analysis. This dependence was not proven when using WM anatomical masks ($r = -0.21$, $p = .10$).

The analytical approach presented shown to be useful for group analysis of DTI data for cervical spinal cord.

## 1. Introduction

Diffusion tensor magnetic resonance imaging (DTI) of human brain is commonly used today in many clinical applications [1]. There exist several softwares and approaches for quality control, preprocessing, post-processing, and quantitative analysis of diffusion data [2]. One of the most sophisticated algorithms for automated whole-brain analysis of images from multiple subjects is that of tract-based spatial statistics (TBSS) [3]. TBSS is based on estimating a skeleton representing the most possible voxels of white matter (WM) across all subjects, projecting diffusion parameters like fractional anisotropy (FA) or mean diffusivity (MD) of each subject onto this skeleton, and then running voxelwise cross-subjects statistics such as nonparametric permutation inference [4].

On the other hand, DTI of human spine is not commonly used in clinical practice. It remains mostly in the research realm despite that this anatomical area is important for the diagnosis of many neurological diseases [5]. Data acquisition and all manner of image processing steps are more complicated and challenging when dealing with the spine rather than brain due to artefacts, lower signal-to-noise ratio, heterogeneous anatomy, and the smaller dimensions of the spinal cord. Nonetheless, there exist several software tools for spinal cord detection or segmentation [6,7] and a few for robust analysis [8].

The aim of this study is to create and evaluate the accuracy of a new methodological approach for analyzing DTI data that is based on some of the algorithms used in TBSS but is targeted to human cervical spinal cord with the possibility to apply voxelwise cross-subjects statistics.

* Corresponding author at: Department of Radiology and Nuclear Medicine, The University Hospital Brno and Masaryk University, Jihlavská 340/20, 625 00 Brno, Czech Republic.
*E-mail address:* keriovsky.milos@fnbrno.cz (M. Keřkovský).

https://doi.org/10.1016/j.mri.2020.07.008
Received 15 April 2020; Accepted 14 July 2020

**Table 1**
MRI sequence parameters.

| Sequence | Orientation | TR [ms] | TE [ms] | Acq. voxel size [mm] | Recon voxel size [mm] |
|---|---|---|---|---|---|
| T2 TSE | Sagittal | 3500 | 120 | 0.92 × 1.27 × 3 | 0.48 × 0.48 × 3 |
| $T_2$* FFE | Transverse | 170 | 9.2 | 0.64 × 0.67 × 4 | 0.39 × 0.39 × 4 |
| DTI | Transverse | 3150 | 110 | 1 × 1 × 4 | 0.88 × 0.88 × 4 |

$T_2$-weighted ($T_2$), $T_2$*-weighted ($T_2$*), diffusion tensor imaging (DTI), repetition time (TR), echo time (TE), turbo spin echo (TSE), fast field echo (FFE).

## 2. Material and methods

The study group consisted of 64 volunteers (42 women and 22 men) aged 21–62 years (mean age 34.52 with standard deviation 8.32 years) in whom no pathological spinal cord changes were found by experienced neuroradiologists (MK, MM, AŠP).

All subjects underwent MRI of the cervical spine in a 1.5 T MRI machine (Philips Achieva, Best, Netherlands) using a 16-channel head and neck coil. The scanning protocol included T2-weighted turbo spin echo (TSE) sequence in sagittal plane, two $T_2$*-weighted fast field echo (FFE) sequences in axial plane with the first stack coverage area being from C1 to C3/4 and the second continuously linked to the first stack from C3/4 to C7, and two DTI sequences with exactly the same geometry as for the axial FFE sequences. DTI was acquired by single-shot echo planar imaging technique (SS-EPI) with 15 different directions of diffusion gradient, with b-value = 800 s/mm$^2$, with 6 excitations (i.e., NEX = 6) and 1 b0 image with NEX = 3 (Table 1). Both axial stacks were placed with maximum perpendicularity to the spinal cord while taking into consideration the overlap of areas around the C3/4 disc. This approach partially eliminates the distorsions given by the cord curvature and help to reduce banana effect, so that other registrations or data interpolations, which can become additional source of bias, are not entirely necessary.

This study used Functional Magnetic Resonance Imaging of the Brain Software Library (FSL [9],) for linear (FLIRT) and nonlinear (FNIRT) registration, respectively, [10] TBSS as inspiration, Randomise (permutation-based inference tool for nonparametric statistical thresholding) as a tool for voxelwise nonparametric permutation inference, and FMRIB's Diffusion Toolbox (FDT) for estimating diffusion tensor parameters [11]. Preprocessing and segmentation of spinal cord tissue was done by MD using a semiautomated method (CLASS) based on machine learning with teacher implemented in ITK-SNAP software [12]. This method is not atlas based and as such does not need any large training dataset; it uses random forest and geodesic active contour method. CLASS method works with each subject individually without any interpolation or registration and it can reach better agreement with manual segmentation than atlas based method, as evaluated in previous work [6]. All statistical tests were computed using Statistica 12 software (StatSoft, Tulsa, OK).

### 2.1. Anatomical data preprocessing

In the first step, the anatomical data of axial $T_2$*-weighted images were aligned on the axis whereby one image of the entire cervical spinal cord was created. This was carried out using rigid registration (3 degrees of freedom [DoF]) whereby coordinates of the central canal of overlapping areas were used as reference points for alignment. These coordinates were determined manually by a radiologist (MK). Segmentation of the entire spinal cord, white (WM) and gray (GM) matter was carried out semi-automatically using the CLASS method. The entire spinal cord was straightened slice by slice based upon coordinates for the center of gravity of the binary mask of the entire spinal cord and centered into the middle of the image using 2D registration with 3 DoF.

### 2.2. Diffusion data preprocessing

Diffusion tensor parameters were calculated for both DTI stacks using the FDT tool. Next, the two stacks were combined in the direction of the axis in the same manner as in the case of $T_2$* images. Based on the FA image, the entire spinal cord was segmented using the CLASS method, straightened, then centered in the same manner as in the case of anatomical data. Finally, GM and WM segmentation masks were registered slice after slice from anatomical space into the DTI space in a manner described previously [6]. The position of axial slices corresponding to intervertebral discs C1/2 and C6/7 were determined manually based on sagittal T2-weighted images for all subjects. Axial images outside of this area were removed.

### 2.3. Atlas creation

We calculated the mean number of slices from discs C1/2 to C6/7 for all 64 subjects (20.84 slices), rounded that number, then set it as the optimal number of slices of the $T_2$* and FA atlases. We chose a similar iterative registration approach as when creating a generally used MNI 152 brain atlas [13].

In the first step, a segmented, straightened, and centered FA image of the subject's entire spinal cord with the mean number of slices and the best image quality was chosen as the reference image. The voxel size was rescaled by a factor of 2 in all dimensions using trilinear interpolation implemented in FLIRT. That resulted in the dimensions 0.44 × 0.44 × 2 mm and field of view was cropped to size 24.8 × 18.1 mm$^2$. Images of the remaining subjects were either expanded or contracted to the number of 21 slices depending on the length of the spinal cord, rescaled, cropped, and the corresponding images were registered (3 DoF) based on the coordinates of the center of gravity. The mean image across all objects registered in this manner was set as the new reference image for the next step (a so-called alpha template).

In the second iteration step, rigid registration of the corresponding slices for all 64 subjects was again carried out whereby the alpha template was the reference image and by means of averaging a beta template was created. In the final step, rigid registration was used for the third time whereby the beta template was the reference image and after that a nonlinear image registration (FNIRT) was used. Nonlinear registration was carried out in the multiresolution mode consisting of four substeps, each with five iterations. Each substep used a different subsampling factor value (4, 2, 1, 1) and Gaussian filter with various values for full width at half maximum (FWHM = 4; 2; 1; 0). The final template of the FA image of the cervical spinal cord was created by averaging across all subjects. All registration matrices and maps were saved and subsequently used to create an atlas of WM and GM based on the registered masks from anatomical space. The same method was used to create templates of anatomical images of cervical spinal cord for individual tissues based on $T_2$* images. By manual labeling of individual discs, a probabilistic map of intervertebral discs was created (Fig. 5G). Outputs of the key steps are displayed in Fig. 1.

### 2.4. WM skeleton calculation

The template of the whole spinal cord created in the previous step

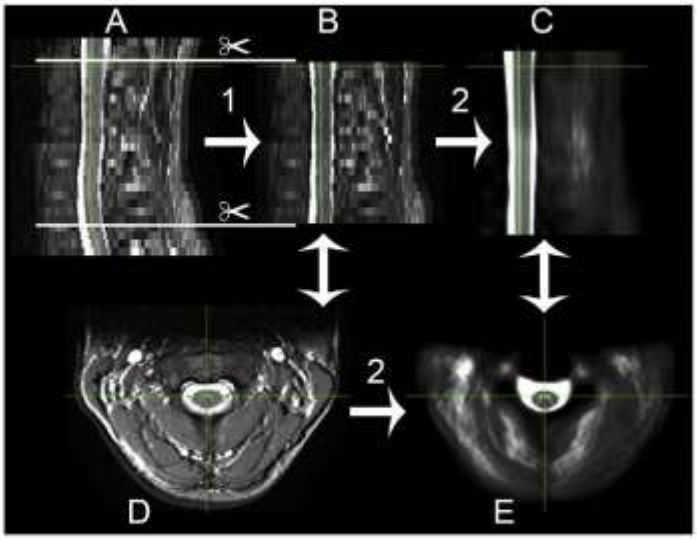



**Fig. 1.** Illustration of atlas creation. Initially, in the original axial $T_2$*-weighted image (A – sagittal, D – axial view) only slices from the C1/2 to C6/7 region were chosen and spinal cord was straightened (1, B). This was followed by registration of all subjects, interpolation, and averaging (2). The final image of the atlas is depicted in sagittal (C) and axial (E) view.

was chosen as the reference image for the one-step linear and nonlinear registration of FA images of individual subjects. A mean image from FA data was used to project an alignment-invariant tract representation (single line skeleton) which identified voxels with the highest FA value and created a single line tract shared with the largest-possible number of subjects [3]. This binary curve should most probably correspond to the WM voxels across all subjects. In the vicinity of the central canal surrounded by WM, however, the algorithm tends to identify the voxels in this area incorrectly as WM. Therefore, the atlas GM binary mask (threshold 0.9), which very probably represents the GM voxels and the central canal area, was subtracted from the binary mask of the single line skeleton. From a statistical power viewpoint, it is generally advantageous when more data are entering into subsequent analyses. Therefore, the single line skeleton can alternatively be extended by binary operation of dilation (matrix size 3 × 3 × 1) whereby a dilated WM skeleton is created from a greater number of voxels, albeit probably with greater risk of cerebrospinal fluid (CSF) contamination.

### 2.5. GM skeleton calculation

By inverting and increasing contrast of the FA spinal cord image (Eq. 1), the GM voxels will have greater intensity than do WM voxels, and therefore we can use the same procedure to create the GM skeleton.

$$I_{new} = |FA| - 1, \tag{1}$$

where $I_{new}$ is the new voxel value and $FA$ is the fractional anisotropy value of the voxel. Dilation using the same matrix was applied also to the GM single line skeleton.

### 2.6. Skeleton agreement evaluation

To evaluate the accuracy of skeleton on an individual level, we calculated the ratio of the number of voxels of the WM skeleton intersection and of the registered WM, GM, or CSF mask to determine the inaccuracy rate according to Eq. (2).

$$s\,(\%) = \frac{|M_{skel} \cap M_{tiz}|}{|M_{skel}|} * 100, \tag{2}$$

where the numerator represents the number of voxels in the common intersection of the skeleton ($M_{skel}$) and the tissue mask ($M_{tiz}$). The denominator is the number of skeleton voxels. In a case of all skeleton voxels corresponding to the voxels of registered tissue mask $M_{tiz}$, the value will be 100%.

To imagine what percentage of voxels of the given tissue *tis* the skeleton represents, the ratio was calculated according to Eq. (3), which theoretically reaches 100% if the skeleton covers all voxels of the given tissue.

$$e\,(\%) = \frac{|M_{skel} \cap M_{tiz}|}{|M_{tiz}|} * 100 \tag{3}$$

The threshold for registered binary masks was set by an experienced radiologist at 0.3 for WM and 0.7 for GM. Voxels outside of the mask of the entire spinal cord were assigned to CSF. A small percentage of voxels was assigned to none of the tissues (WM, GM, CSF), which was caused by the threshold of segmentation masks and their interpolation during registration into space of the FA images. Both single as well as dilated skeletons for WM and GM were evaluated in this manner.

Median FA values obtained by means of registered segmentation masks and skeletons of individual subjects were mutually compared using paired Student's *t*-test.

By means of voxelwise nonparametric permutation statistics with multiple testing correction of the entire cohort based upon 5000 permutations, the dependence of FA on sex for all skeletons types was verified where age was the independent covariate. The same dependence was tested using Student's *t*-test on median data of registered segmentation masks of individual tissues.

Using the same statistics, the dependence of FA on age for all skeleton types was verified where sex was the independent covariate. Median values of statistically significant voxels identified by the previous voxelwise analysis were subjected to regression analysis. The same regression analysis was applied to data for registered segmentation masks of individual tissues. The significance level for all statistical tests was set at $\alpha = 0.05$.

All volunteers signed informed consent agreements to participate in the study, which was approved by the University Hospital's Ethics Committee.

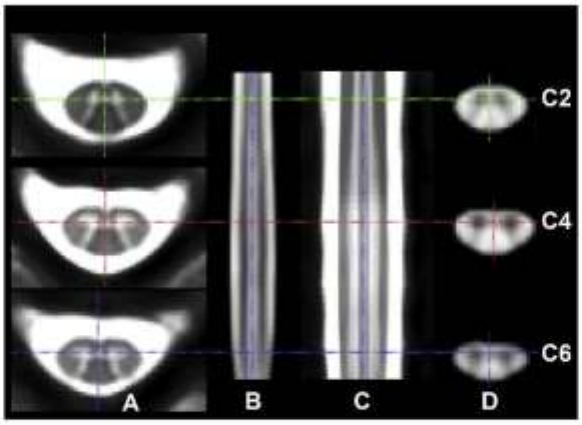



**Fig. 2.** Illustration of the final template $T_2^*$ in axial (A), sagittal (B), and coronal (C) plane and FA template in axial plane (D). Axial slices of both templates correspond to vertebral levels C2, C4, and C6.

## 3. Results

Atlases of cervical spinal cord were created in the region from disc C1/2 to disc C6/7 for anatomical $T_2^*$ data as well as for FA data in common space. Three axial slices of both atlases together with corresponding sagittal and coronal plane are depicted in Fig. 2.

Based on the mean FA image of all subjects, WM (Fig. 3, upper line) and GM (Fig. 3, bottom line) skeletons were calculated together with their dilated version (Fig. 3, right column). To suppress voxels in the area of the central canal in the WM skeleton, the GM atlas mask with threshold 0.9 was subtracted (Fig. 3B).

The calculated skeletons correspond well visually to voxels of WM and GM in the mean image, where the gray matter is visible as an area with reduced FA value (Fig. 3). Quantitative verification of skeleton accuracy is captured in Table 2, which describes percentage correspondence of common voxels of skeletons and masks in individual tissues (WM, GM, CSF) for parameters *s* and *e*. Statistically significant differences ($p < .05$) were proven by means of paired *t*-test between single line and dilated skeleton for all parameters shown in Table 2. It is nevertheless important also to look at the projections of skeleton into the images of individual subjects (Fig. 4), where it is obvious that, despite the fact that nonlinear registration was used, the position of gray matter for individual subjects varies a great deal in relation to skeleton GM. This, together with the low number of GM voxels, explains the lower values for parameter *e* in the case of determination of the GM skeleton as GM.

Voxelwise analysis of single line as well as dilated WM skeletons identified as having statistically significantly different voxels ($p < .05$) in the area of anterior and lateral columns of WM depending on sex of subjects with age as covariate and with multiple testing correction (Fig. 5A, B, C) where the mean value (SD) of WM FA of statistically significant voxels of the single line skeleton was 0.653 (0.056) in women and 0.730 (0.060) in men. Several voxels where a significant dependence of FA on sex was proven were identified also within GM in segment C2/C3. In women mean (SD) values of 0.576 (0.050) were measured in these areas. In men, they were 0.640 (0.051). Student's *t*-

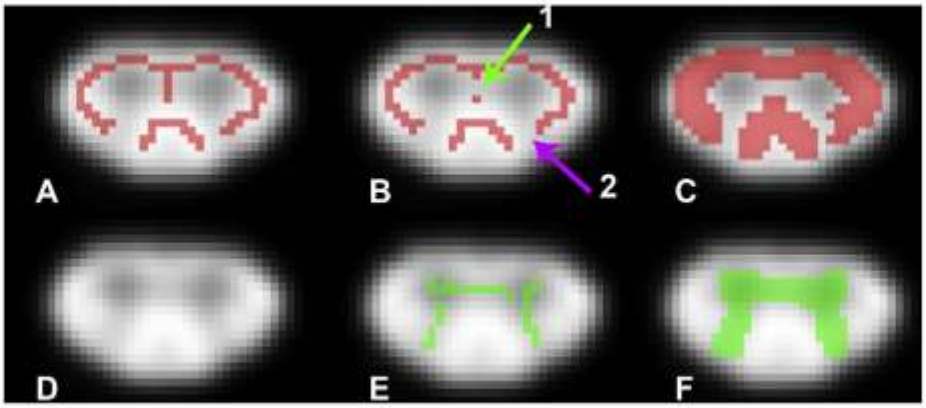



**Fig. 3.** Illustration of WM skeleton before (A) and after (B) subtracting GM mask (there are fewer voxels in the area of the central canal (1) and left posterior horn (2)) and after expansion by dilation (C). Original FA image is captured in image D and the GM skeleton together with its dilated version are in images E and F, respectively.

**Table 2**
Verification of percentage correspondence between WM and GM skeletons before and after dilation and between masks of individual tissues (i.e., percentage of skeleton voxels corresponding with WM, GM, CSF mask or with no mask.) Statistically significant differences ($p < .05$) were proven between single line and dilated skeleton for all parameters (paired t-test).

| | Single line WM skeleton | | | | | Dilated WM skeleton [%] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | s [%] | | | | e [%] | s [%] | | | | e [%] |
| Tissue | CSF | GM | WM | None | WM | CSF | GM | WM | None | WM |
| Mean | 0.02 | 4.70 | 94.95 | 0.35 | 16.04 | 0.12 | 5.87 | 93.28 | 0.79 | 49.15 |
| SD | 0.08 | 2.45 | 2.59 | 0.39 | 1.08 | 0.22 | 2.39 | 2.65 | 0.67 | 3.03 |
| Min | 0.00 | 1.12 | 86.29 | 0.00 | 12.93 | 0.00 | 1.44 | 85.38 | 0.01 | 39.94 |
| Max | 0.56 | 13.58 | 98.88 | 1.64 | 18.67 | 1.29 | 14.41 | 98.41 | 2.61 | 57.11 |
| #Voxels | 2212 | | | | | 7215 | | | | |

| | Single line GM skeleton | | | | | Dilated GM skeleton [%] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | s [%] | | | | e [%] | s [%] | | | | e [%] |
| Tissue | CSF | GM | WM | None | GM | CSF | GM | WM | None | GM |
| Mean | 0.00 | 56.07 | 43.74 | 0.03 | 34.18 | 0.02 | 41.82 | 57.71 | 0.24 | 83.54 |
| SD | 0.00 | 9.01 | 9.01 | 0.05 | 3.59 | 0.06 | 6.81 | 6.97 | 0.28 | 4.79 |
| Min | 0.00 | 32.71 | 20.44 | 0.00 | 23.87 | 0.00 | 28.80 | 36.96 | 0.00 | 69.11 |
| Max | 0.00 | 79.56 | 67.21 | 0.21 | 44.22 | 0.36 | 62.60 | 71.04 | 1.16 | 95.76 |
| #Voxels | 1174 | | | | | 3861 | | | | |

White matter (WM), gray matter (GM), voxels overlap (s), percentage of voxels of the given tissue the skeleton represents (e), standard deviation (SD).

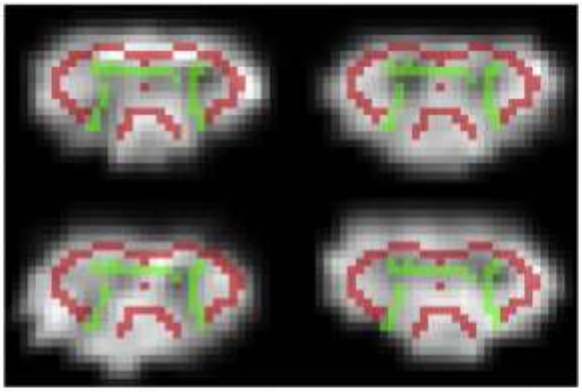


**Fig. 4.** Illustration of projection of WM and GM skeletons for four different subjects (1, 15, 37, 57) of corresponding axial slices.

test proved a statistically significant difference of FA values depending on sex ($p < .05$) in the case of registered GM masks, where the mean (SD) FA value in women was 0.616 (0.036) and in men it was 0.638 (0.038). Similar dependence was not proven in WM with registered data ($p = .12$).

In analyzing the influence of age on FA values using the single line skeleton, no statistically significant areas were identified ($p > .05$). In the area of dorsal columns of spinal cord at level C4/C5 when analyzing the dilated WM skeleton, a dependence of FA values on age of the subjects was discovered where sex was a covariate and after multiple testing correction (Fig. 5D, E, F). To verify the results of voxelwise analysis and for the sake of better comparison with registered masks, the dependence of FA median values measured within voxels identified by voxelwise analysis as significantly dependent on age of subjects was verified also by means of linear regression. As expected, this method confirmed the dependence ($r = -0.56$, $p < .001$) (Fig. 6A). Linear regression based on registered WM masks indicates decrease of FA with aging ($r_{WM} = -0.21$), but statistical significance was not proven for any of the tissues ($p_{WM} = 0.10$) (Fig. 6B).

Paired Student's t-test proved statistically significant difference between the methods used ($p < .001$) in FA values of individual subjects for WM as well as for GM. WM single line skeleton reached higher values than did other methods and lower values in the case of GM. All methods demonstrated significantly ($p < .001$) lower FA values of GM than WM (Table 3).

## 4. Discussion

This study presents a robust approach to large-scale voxelwise DTI data analysis of cervical spinal cord. This methodology does not require the segmentation of individual tissues or manual region-of-interest positioning, which usually is a source of large inaccuracies. For voxelwise analysis, we used a method which is very well proven on brain DTI data for healthy as well as sick subjects, with a minimal number of input parameters that increase its user independence. The possibilities of finding clusters in data without the necessity for binary masks (threshold-free cluster enhancement) and for multiple testing correction using the family-wise error rate makes this method easier to use.

The percentage correspondence between the two skeletons with segmented WM mask is around 94%, which may be regarded as a good

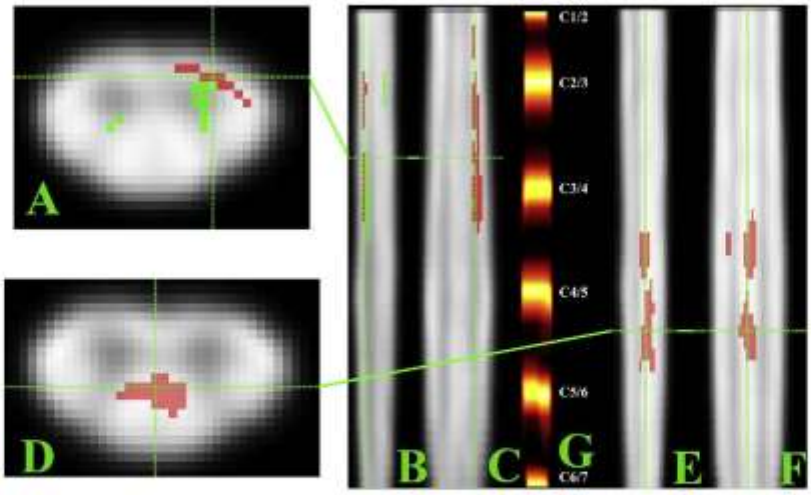



**Fig. 5.** Illustration of statistically different voxels in WM and GM single line skeleton ($p < .05$) with multiple testing correction in which sex significantly influences the FA values in voxels depicted in axial (A), sagittal (B), and coronal (C) planes. Illustration of statistically different voxels of dilated WM skeleton ($p < .05$) with multiple testing correction wherein age significantly influences the FA values in voxels depicted in axial (D), sagittal (E), and coronal (F) planes. Position of individual segments is depicted in the probability map of discs created according to template (G). Position of axial slices corresponds to green guiding lines in corresponding sagittal and coronal planes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

result. Due to skeleton dilation, there was only a minimal increase in incorrectly specified voxels but there was a more than three-fold increase in the number of voxels entering the analysis and the percentage coverage of WM reached nearly 50%. That brings greater statistical power to the analysis and may play a key role. Very low contamination of WM skeletons by CSF voxels may be crucial in the final statistics because FA values of CSF are significantly lower than those for WM. In contrast to this finding, the percentage correspondence between GM skeletons was around 56% in the case of the single line and 41% for the dilated version. Such low numbers may be caused by lower number of voxels entering the TBSS algorithm so that the same number of incorrectly identified voxels means greater relative error and imperfect registration of segmentation masks onto the template when an error could occur during interpolation or data thresholding. Likewise, an error in the segmentation of GM on $T_2$* data itself can play a significant role because this is a small anatomical structure and therefore the error rate of segmentation methods is higher than in the case of WM [14–19]. On the other hand, percentage coverage of GM by means of dilated

**Table 3**
Mean value of median FA values for individual subjects ± SD.

| | $FA_{WM}$ (SD) | $FA_{GM}$ (SD) |
|---|---|---|
| Mask | 0.674 (0.029) | 0.624 (0.038) |
| Dilated skeleton | 0.709 (0.029) | 0.631 (0.037) |
| Single line skeleton | 0.727 (0.031) | 0.607 (0.041) |

Fraction anisotropy (FA), white matter (WM), gray matter (GM), standard deviation (SD).

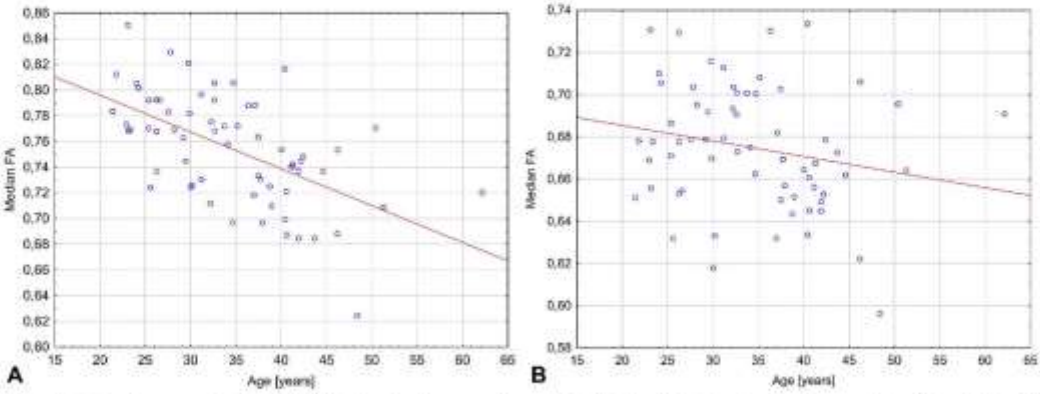



**Fig. 6.** Illustration of linear regression of median FA values depending on age in voxels of statistically significant regions based on voxelwise analysis of dilated WM skeleton (A) and median FA values of registered WM masks (B).

skeleton reached a mean 83% in comparison to CLASS segmentation.

Mean FA values for single line skeleton are lower than in the case of registered GM masks. We can conclude from this, considering the zero contamination by CSF voxels, that this mask is less strained by partial volume effect (PVE) in relation to WM, where the presence of WM would increase the mean value. This conclusion cannot be stated, however, for dilated GM skeleton, which reaches higher mean FA value than do registered masks, and therefore a greater representation of WM voxels can be anticipated. This can be caused also by the use of square dilation mask 3 × 3, which may not be optimal for the case of a specific shape of GM with small volume. In cases of WM single line skeleton, the minimization of PVE is even more distinct than in cases of GM, where, despite moderate contamination by CSF and GM voxels, the FA values are noticeably higher than in cases of registered WM masks. Although skeleton dilation increases contamination by foreign tissue, the PVE is nevertheless statistically significantly smaller than in cases of registered masks, and, due to the higher number of voxels, the statistical power of the tests is greater than in cases of using single line skeleton. On average, moreover, it in itself contains almost 50% of the volume of the entire white matter identified by segmentations.

In a study from 2008 [20], a weak but statistically significant dependence of FA values of the entire spinal cord on age was proven. Chan et al. proved the strongest dependence of FA values on age in dorsal WM columns using manual segmentation [21]. Accordingly, in our study a significant dependence of FA values on age was proven at some levels of dorsal columns using voxelwise analysis of dilated skeleton voxels. Vedantam et al. proved a strong dependence of FA values on age in healthy subjects over 55 years of age in all levels of cervical spinal cord for cross-section of the entire spinal cord [22]. Within a graph from the same study showing dependence of FA values on age for the middle segment (C4–C6), a weak negative correlation between FA values and age is noticeable also in subjects under 55 years of age. Considering that the age composition of our cohort corresponds rather to the subgroup up to 55 years of age in the study just cited, in this respect a corresponding correlation can be seen in our finding of age dependence of FA in voxels of middle segments of spinal cord (C4–C5). Differences between segments in age dependence of FA values could be explained in part by the study of Valsasina et al., who in a group of healthy volunteers proved within the same segment (C4–C5) a correlation between cervical cord atrophy and age of subjects based on 3D T1-weighted images [23]. This reality could to a certain extent influence the diffusivity of spinal cord in the mentioned region, but mutual correlation of spinal cord volume and diffusion parameters would need to be further verified.

In contrast to the voxelwise analysis, we did not prove significant correlation between FA values and age when using linear regression of WM and GM mean values from CLASS segmentation. This further documents advantages of the voxelwise method, which enables the detection of smaller regional changes of spinal cord diffusivity that can disappear when evaluating parameters measured within the entire WM or GM. At the same time, in this case a higher statistical power of the voxelwise analysis was confirmed when a greater number of voxels was included. On the other hand, significant correlation with age was not proven when using single line skeleton.

The authors of some of those studies cited above also evaluated the dependence of FA values on sex. Chan et al. compared FA values between sexes in different segments for entire spinal cord cross-section and found no significant difference [21]. Nor did Vedantam et al. in evaluating FA values of the entire spinal cord, as well as in white and gray matter, prove a statistically significant difference between sexes [22]. The results of the work presented here differ from those cited mainly by the fact that, unlike that of Chan et al., this work employed voxel-based statistical analysis, which leads to accentuation of regional differences. In their work, Vedantam et al. compared a smaller number of subjects (12 men and 13 women) across a large age span (20–89 years). Therefore, the difference in FA values caused by age most likely prevailed over the differences caused by different sexes. Dependence of FA values on sex was proven in both the dilation and single line skeletons by lower numbers of voxels, and that supports the significance of this finding. Regional dependence of diffusivity on sex had been proven earlier on brain data, and therefore it cannot be ruled out that regional changes across sexes can occur also in spinal cord [24]. Kerkovsky et al. proved changes in the apparent diffusion coefficient of cervical spinal cord depending on sex in patients with degenerative compression of the cervical spinal cord as well as in healthy controls.

The authors were able to find three studies that had taken similar approaches to analyzing cervical spinal cord DTI data. The first study, from 2010 [25], identified by means of voxelwise analysis with multiple testing correction regions with different diffusivity among a small group of healthy volunteers (15) and patients (14) with amyotrophic lateral sclerosis (ALS), albeit at level of significance $\alpha = 0.1$. Another, similar study [26] dealt with voxelwise analysis of a narrow region of cervical spinal cord (C1–C3) in 14 patients with multiple sclerosis (MS) and having at least one spinal cord lesion in this region. Those authors proved a statistically significant decline of FA values in the C2–C3 region in patients with MS compared to healthy volunteers. The decline was mostly in GM and at level of significance $\alpha = 0.01$, but without multiple testing correction. The newest study [27] used voxelwise analysis of atlas segmented WM tracts in 20 patients with cervical spondylotic myelopathy for the level of disc C2/C3 and identified regions with different diffusivity in comparison to a group of healthy volunteers (18). In the cervical spinal cord region, the authors found only one study using the TBSS like algorithm directly to detect WM based on coronary DTI data [25], but that was without any quantitative testing for accuracy.

This study has several limitations. The first relates to its use of a 1.5 T machine, which can result in a poorer signal-to-noise ratio in lower segments of cervical spinal cord and therefore also in lower resolution of anatomical images compared to images from machines with higher field. For that reason, segmentation masks in this region can be degraded or the estimate of diffusion tensor can be less accurate [22]. A second limitation can be in the absence of DTI data with reverse phase encoding direction or use of sequences with reduced phase field of view (ZOOM, FOCUS) to better correct for distorted artefacts. This shortcoming is partially compensated, however, by using a machine with lower field that by its physical nature produces smaller numbers of distorted artefacts compared to machines with higher induction. The next limitation could be the relatively low mean age and narrower age span of the subjects. As a result, direct comparison with other studies dealing primarily with the influence of age on spinal cord diffusivity is limited and the results obtained less robust. Despite this limitation, however, it was possible to prove the correlation of FA values of spinal cord with age by the method used. Furthermore, it is necessary to think about using fixation means or shot triggering of DTI data with ECG to eliminate movement artefacts that can in the end lead to acquiring more robust results. These methods were not utilized in an effort to move closer to an MR image protocol of spinal cord diffusion that could be utilized in real situations where such tools usually are not used due to time and technical constraints.

In closing, we can state that this study quantitatively verified the applicability of voxelwise analysis by a tract-based spatial statistics approach on DTI data of cervical spinal cord in healthy volunteers not only for white but also for gray matter. High accuracy of this method for voxel detection of individual tissues in correlation with semiautomatic segmentation of WM and GM while minimizing the partial volume effect was proven. Using this method, a significant dependence of fractional anisotropy on age and sex was proven and which more or less corresponds with the assumptions and previous studies dealing with spinal cord or brain DTI.

## Declaration of competing interest

None.

## Acknowledgments

Contract grant sponsor: Czech Health Research Council; contract grant number: AZV-15-32133A.

## References

[1] Tae W-S, Ham B-J, Pyun S-B, Kang S-H, Kim B-J. Current clinical applications of diffusion-tensor imaging in neurological disorders. J Clin Neurol 2018;14:129–40. https://doi.org/10.3988/jcn.2018.14.2.129.
[2] Soares JM, Marques P, Alves V, Sousa N. A hitchhiker's guide to diffusion tensor imaging. Front Neurosci 2013;7. https://doi.org/10.3389/fnins.2013.00031.
[3] Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage 2006;31:1487–505. https://doi.org/10.1016/j.neuroimage.2006.02.024. Amsterdam.
[4] Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. NeuroImage 2014;92:381–97. https://doi.org/10.1016/j.neuroimage.2014.01.060.
[5] Borsook D. Neurological diseases and pain. Brain 2012;135:320–44. https://doi.org/10.1093/brain/aww271.
[6] Dostál M, Keřkovský M, Kerčkiová E, Němcová E, Stulík J, Staňková M, et al. Analysis of diffusion tensor measurements of the human cervical spinal cord based on semiautomatic segmentation of the white and gray matter. J Magn Reson Imaging 2018. https://doi.org/10.1002/jmri.26166.
[7] Leener BD, Taso M, Cohen-Adad J, Callot V. Segmentation of the human spinal cord. Magn Reson Mater Phys Biol Med 2016;29:125–53. https://doi.org/10.1007/s10334-015-0507-2.
[8] De Leener B, Lévy S, Dupont SM, Fonov VS, Stikov N, Louis Collins D, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. NeuroImage 2017;145:24–43. https://doi.org/10.1016/j.neuroimage.2016.10.009.
[9] Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. NeuroImage 2012;62:782–90. https://doi.org/10.1016/j.neuroimage.2011.09.015.
[10] Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal 2001;5:143–56. https://doi.org/10.1016/S1361-8415(01)00036-6.
[11] Behrens TEJ, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, et al. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 2003;50:1077–88. https://doi.org/10.1002/mrm.10609.
[12] Yushkevich PA, Gerig G. ITK-SNAP: an interactive medical image segmentation tool to meet the need for expert-guided segmentation of complex medical images. IEEE Pulse 2017;8:54–7. https://doi.org/10.1109/MPUL.2017.2701493.
[13] Grabner G, Janke AL, Budge MM, Smith D, Pruessner J, Symmetric Atlasing Collins DL. Model based segmentation: An application to the Hippocampus in older adults. In: Larsen R, Nielsen M, Sporring J, editors. Medical image computing and computer-assisted intervention – MICCAI 2006. Springer Berlin Heidelberg; 2006. p. 58–66.
[14] Asman AJ, Bryan FW, Smith SA, Reich DS, Landman BA. Groupwise multi-atlas segmentation of the spinal cord's internal structure. Med Image Anal 2014;18:460–71. https://doi.org/10.1016/j.media.2014.01.003.
[15] Chen M, Carass A, Oh J, Nair G, Pham DL, Reich DS, et al. Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. NeuroImage 2013;83:1051–62. https://doi.org/10.1016/j.neuroimage.2013.07.060.
[16] El Mendili M-M, Chen R, Tiret B, Villard N, Trunet S, Pélégrini-Issac M, et al. Fast and accurate semi-automated segmentation method of spinal cord MR images at 3T applied to the construction of a cervical spinal cord template. PLoS One 2015;10:e0122224https://doi.org/10.1371/journal.pone.0122224.
[17] Prados F, Ashburner J, Blaiotta C, Brosch T, Carballido-Gamio J, Cardoso MJ, et al. Spinal cord grey matter segmentation challenge. NeuroImage 2017;152:312–29. https://doi.org/10.1016/j.neuroimage.2017.03.010.
[18] Taso M, Le Troter A, Sdika M, Cohen-Adad J, Arnoux P-J, Guye M, et al. A reliable spatially normalized template of the human spinal cord — applications to automated white matter/gray matter segmentation and tensor-based morphometry (TBM) mapping of gray matter alterations occurring with age. NeuroImage 2015;117:20–8. https://doi.org/10.1016/j.neuroimage.2015.05.034.
[19] Yiannakas MC, Kearney H, Samson RS, Chard DT, Ciccarelli O, Miller DH, et al. Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: a pilot study with application to magnetization transfer measurements. NeuroImage 2012;63:1054–9. https://doi.org/10.1016/j.neuroimage.2012.07.048.
[20] Henke WV, Leemans A, Sijbers J, Vandervliet E, Goethem JV, Parizel PM. A tracking-based diffusion tensor imaging segmentation method for the detection of diffusion-related changes of the cervical spinal cord with aging. J Magn Reson Imaging 2008;27:978–91. https://doi.org/10.1002/jmri.21338.
[21] Chen T, Li X, Mak K, Cheung JP, Luk KD, Hu Y. Normal values of cervical spinal cord diffusion tensor in young and middle-aged healthy Chinese. Eur Spine J 2015;24:2991–8. https://doi.org/10.1007/s00586-015-4144-2. Heidelberg.
[22] Vedantam A, Jirsa MB, Schmit BD, Wang MC, Ulmer JL, Kurpad SN. Characterization and limitations of diffusion tensor imaging metrics in the cervical spinal cord in neurologically intact subjects. J Magn Reson Imaging 2013;38:861–7. https://doi.org/10.1002/jmri.24039.
[23] Valsasina P, Horsfield MA, Rocca MA, Absinta M, Comi G, Filippi M. Spatial normalization and regional assessment of cord atrophy: voxel-based analysis of cervical cord 3D T1-weighted images. Am J Neuroradiol 2012;33:2195–200. https://doi.org/10.3174/ajnr.A3139.
[24] Inano S, Takao H, Hayashi N, Abe O, Ohtomo K. Effects of age and gender on white matter integrity. Am J Neuroradiol 2011;32:2103–9. https://doi.org/10.3174/ajnr.A2785.
[25] Nair G, Carew JD, Usher S, Lu D, Hu XP, Benatar M. Diffusion tensor imaging reveals regional differences in the cervical spinal cord in amyotrophic lateral sclerosis. NeuroImage 2010;53:576–83. https://doi.org/10.1016/j.neuroimage.2010.06.060.
[26] Toosy AT, Ku N, Altmann D, Wheeler-Kingshott CAM, Thompson AJ, Ciccarelli O. Voxel-based cervical spinal cord mapping of diffusion abnormalities in MS-related myelitis. Neurology 2014;83:1321–5. https://doi.org/10.1212/WNL.0000000000000857.
[27] Grabher P, Mohammadi S, Trachsler A, Friedl S, David G, Sutter R, et al. Voxel-based analysis of grey and white matter degeneration in cervical spondylotic myelopathy. Sci Rep 2016;6:24636https://doi.org/10.1038/srep24636.

nature communications

Article

https://doi.org/10.1038/s41467-022-33407-5

# Federated learning enables big data for rare cancer boundary detection

Received: 7 April 2022

Accepted: 16 September 2022

Published online: 05 December 2022

Check for updates

A list of authors and their affiliations appears at the end of the paper

Although machine learning (ML) has shown promise across disciplines, out-of-sample generalizability is concerning. This is currently addressed by sharing multi-site data, but such centralization is challenging/infeasible to scale due to various limitations. Federated ML (FL) provides an alternative paradigm for accurate and generalizable ML, by only sharing numerical model updates. Here we present the largest FL study to-date, involving data from 71 sites across 6 continents, to generate an automatic tumor boundary detector for the rare disease of glioblastoma, reporting the largest such dataset in the literature ($n = 6,314$). We demonstrate a 33% delineation improvement for the surgically targetable tumor, and 23% for the complete tumor extent, over a publicly trained model. We anticipate our study to: 1) enable more healthcare studies informed by large diverse data, ensuring meaningful results for rare diseases and underrepresented populations, 2) facilitate further analyses for glioblastoma by releasing our consensus model, and 3) demonstrate the FL effectiveness at such scale and task-complexity as a paradigm shift for multi-site collaborations, alleviating the need for data-sharing.

Recent technological advancements in healthcare, coupled with patients' culture shifting from reactive to proactive, have resulted in a radical growth of primary observations generated by health systems. This contributes to the burnout of clinical experts, as such observations require thorough assessment. To alleviate this situation, there have been numerous efforts for the development, evaluation, and eventual clinical translation of machine learning (ML) methods to identify relevant relationships among these observations, thereby reducing the burden on clinical experts. Advances in ML, and particularly deep learning (DL), have shown promise in addressing these complex healthcare problems. However, there are concerns about their generalizability on data from sources that did not participate in model training, i.e., "out-of-sample" data[1,2]. Literature indicates that training robust and accurate models requires large amounts of data[3-5], the diversity of which affects model generalizability to "out-of-sample" cases[6]. To address these concerns, models need to be trained on data originating from numerous sites representing diverse population samples. The current paradigm for such multi-site collaborations is "centralized learning" (CL), in which data from different sites are shared to a centralized location following inter-site agreements[7-9]. However, such data centralization is difficult to scale (and might not even be feasible), especially at a global scale, due to concerns[10,11] relating to privacy, data ownership, intellectual property, technical challenges (e.g., network and storage limitations), as well as compliance with varying regulatory policies (e.g., Health Insurance Portability and Accountability Act (HIPAA) of the United States[12] and the General Data Protection Regulation (GDPR) of the European Union[13]). In contrast to this centralized paradigm, "federated learning" (FL) describes a paradigm where models are trained by only sharing model parameter updates from decentralized data (i.e., each site retains its data locally)[10,11,14-16], without sacrificing performance when compared to CL-trained models[6,15,17-21]. Thus, FL can offer an alternative to CL, potentially creating a paradigm shift that alleviates the need for data sharing, and hence increase access to geographically distinct collaborators, thereby increasing the size and diversity of data used to train ML models.

FL has tremendous potential in healthcare[22,23], particularly towards addressing health disparities, under-served populations, and "rare" diseases[24], by enabling ML models to gain knowledge from ample and diverse data that would otherwise not be available. With

e-mail: speakes@upenn.edu

that in mind, here we focus on the "rare" disease of glioblastoma, and particularly on the detection of its extent using multi-parametric magnetic resonance imaging (mpMRI) scans[18]. While glioblastoma is the most common malignant primary brain tumor[19–23], it is still classified as a "rare" disease, as its incidence rate (i.e., 3/100,000 people) is substantially lower than the rare disease definition rate (i.e., <10/100,000 people)[24]. This means that single sites cannot collect large and diverse datasets to train robust and generalizable ML models, necessitating collaboration between geographically distinct sites. Despite extensive efforts to improve the prognosis of glioblastoma patients with intense multimodal therapy, their median overall survival is only 14.6 months after standard-of-care treatment, and 4 months without treatment[25]. Although the subtyping of glioblastoma has been improved[10] and the standard-of-care treatment options have expanded during the last 20 years, there have been no substantial improvements in overall survival[21]. This reflects the major obstacle in treating these tumors which is their intrinsic heterogeneity[26,27], and the need for analyses of larger and more diverse data toward a better understanding of the disease. In terms of radiologic appearance, glioblastomas comprise of three main sub-compartments, defined as (i) the "enhancing tumor" (ET), representing the vascular blood-brain barrier breakdown within the tumor, (ii) the "tumor core" (TC), which includes the ET and the necrotic (NCR) part, and represents the surgically relevant part of the tumor, and (iii) the "whole tumor" (WT), which is defined by the union of the TC and the peritumoral edematous/infiltrated tissue (ED) and represents the complete tumor extent relevant to radiotherapy (Fig. 1b). Detecting these sub-compartment boundaries, therefore, defines a multi-parametric multi-class learning problem and is a critical first step towards further quantifying and assessing this heterogeneous rare disease and ultimately influencing clinical decision-making.

Co-authors in this study have previously introduced FL in healthcare in a simulated setting[11] and further conducted a thorough quantitative performance evaluation of different FL workflows[11] (refer to supplementary figures for illustration) for the same use-case as the present study, i.e., detecting the boundaries of glioblastoma sub-compartments. Findings from these studies supported the superiority of the FL workflow used in the present study (i.e., based on an aggregation server[13,14]), which had almost identical performance to CL, for this use-case. Another study[32] has explored the first real-world federation for a breast cancer classification task using 5 sites, and another[33] used electronic medical records along with x-ray images from 20 sites to train a classifier to output a label corresponding to future oxygen requirement for COVID-19 patients.

This study describes the largest to-date global FL effort to develop an accurate and generalizable ML model for detecting glioblastoma sub-compartment boundaries, based on data from 6314 glioblastoma patients from 71 geographically distinct sites, across six continents (Fig. 1a). Notably, this describes the largest and most diverse dataset of glioblastoma patients ever considered in the literature. It was the use of FL that successfully enabled our ML model to gain knowledge from such an unprecedented dataset. The extended global footprint and the task complexity are what sets this study apart from current literature, since it dealt with a multi-parametric multi-class problem with reference standards that require expert clinicians following an involved manual annotation protocol, rather than simply recording a categorical entry from medical records[18,32]. Moreover, varying characteristics of the mpMRI data due to scanner hardware and acquisition protocol differences[34,34] were handled at each collaborating site via established harmonized preprocessing pipelines[35–38].

The scientific contributions of this manuscript can be summarized by (i) the insights garnered during this work that can pave the way for more successful FL studies of increased scale and task complexity, (ii) making a potential impact for the treatment of the rare disease of glioblastoma by publicly releasing clinically deployable trained consensus models, and most importantly, iii) demonstrating the effectiveness of FL at such scale and task complexity as a paradigm shift redefining multi-site collaborations, while alleviating the need for data sharing.

## Results

The complete federation followed a staged approach, starting from a "public initial model" (trained on data of 231 cases from 16 sites), followed by a "preliminary consensus model" (involving data of 2471 cases from 35 sites), to conclude on the "final consensus model" (developed on data of 6314 cases from 71 sites). To quantitatively evaluate the performance of the trained models, 20% of the total cases contributed by each participating site were excluded from the model training process and used as "local validation data". To further evaluate the generalizability of the models in unseen data, 6 sites were not involved in any of the training stages to represent an unseen "out-of-sample" data population of 590 cases. To facilitate further evaluation without burdening the collaborating sites, a subset ($n = 332$) of these cases was aggregated to serve as a "centralized out-of-sample" dataset. The training was initiated from a pre-trained model (i.e., our public initial model) rather than a random initialization point, in order to have faster convergence of the model performance[40,41]. Model performance was quantitatively evaluated here using the Dice similarity coefficient (DSC), which assesses the spatial agreement between the model's prediction and the reference standard for each of the three tumor sub-compartments (ET, TC, WT).

### Increased data can improve performance

When the federation began, the public initial model was evaluated against the local validation data of all sites, resulting in an average (across all cases of all sites) DSC per sub-compartment, of $DSC_{ET} = 0.63$, $DSC_{TC} = 0.62$, $DSC_{WT} = 0.75$. To summarize the model performance with a single collective score, we then calculate the average DSC (across all 3 tumor sub-compartments per case, and then across all cases of all sites) as equal to 0.66. Following model training across all sites, the final consensus model garnered significant performance improvements against the collaborators' local validation data of 27% ($p_{ET} < 1 \times 10^{-36}$), 33% ($p_{TC} < 1 \times 10^{-57}$), and 16% ($p_{WT} < 1 \times 10^{-21}$), for ET, TC, and WT, respectively (Fig. 1c). To further evaluate the potential generalizability improvements of the final consensus model on unseen data, we compared it with the public initial model against the complete out-of-sample data and noted significant performance improvements of 15% ($p_{ET} < 1 \times 10^{-5}$), 27% ($p_{TC} < 1 \times 10^{-10}$), and 16% ($p_{WT} < 1 \times 10^{-7}$), for ET, TC, and WT, respectively (Fig. 1d). Notably, the only difference between the public initial model and the final consensus model, was that the latter gained knowledge during training from increased datasets contributed by the complete set of collaborators. The conclusion of this finding reinforces the importance of using large and diverse data for generalizable models to ultimately drive patient care.

### Data size alone may not predict success

This is initially observed in our federated setting, where the comparative evaluation of the public initial model, the preliminary consensus model, and the final consensus model, against the centralized out-of-sample data, indicated performance improvements not directly related to the amount of data used for training. Specifically, we noted major significant ($p < 7 \times 10^{-19}$, Wilcoxon signed-rank test) performance improvements between the public initial model and the preliminary consensus model, as opposed to the insignificant ($p > 0.067$, Wilcoxon signed-rank test) ones between the preliminary and the final consensus model, as quantified in the centralized out-of-sample data for all sub-compartments and their average (Fig. 2).

We further expanded this analysis to assess this observation in a non-federated configuration, where we selected the largest collaborating sites (comprehensive cancer centers contributing > 200 cases,

https://doi.org/10.1038/s41467-022-33407-5

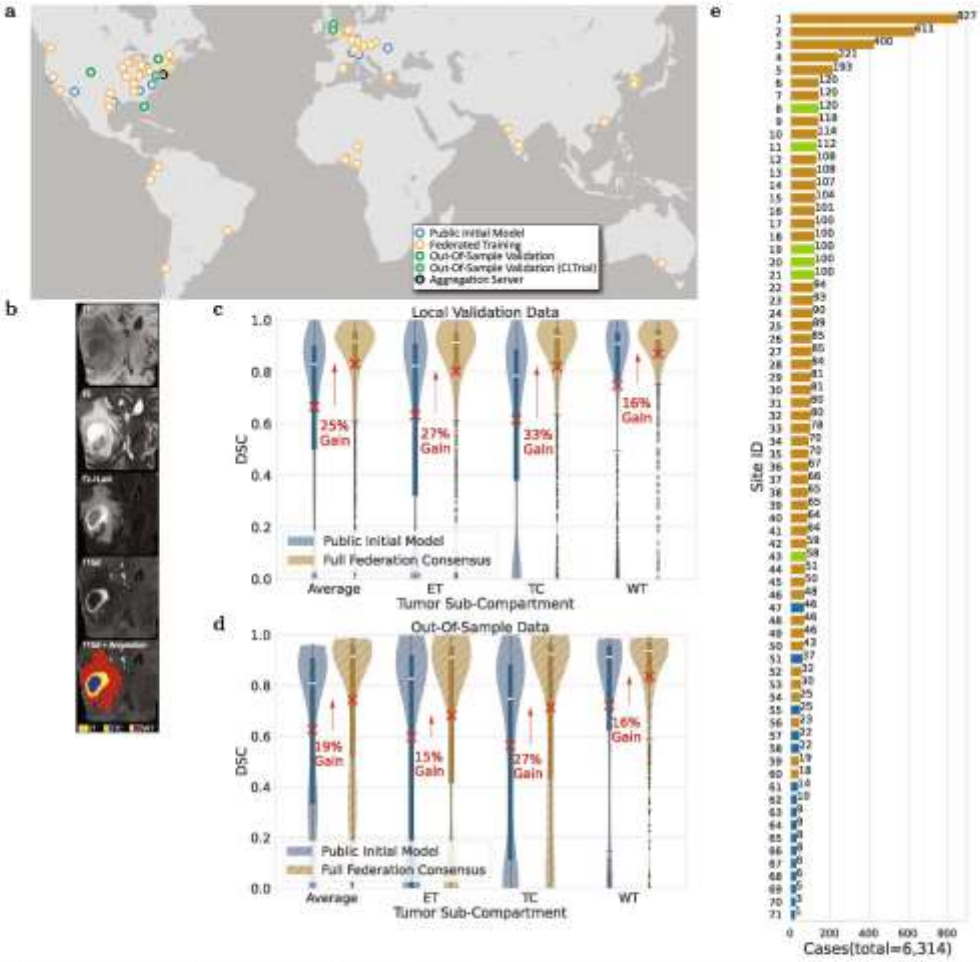



**Fig. 1 | Representation of the study's global scale, diversity, and complexity.** **a** The map of all sites involved in the development of FL consensus model. **b** Example of a glioblastoma mpMRI scan with corresponding reference annotations of the tumor sub-compartments (ET enhancing tumor, TC tumor core, WT whole tumor). **c**, **d** Comparative Dice similarity coefficient (DSC) performance evaluation of the final consensus model with the public initial model on the collaborators' local validation data (in **c** with $n = 1043$ biologically independent cases) and on the complete out-of-sample data (in **d** with $n = 518$ biologically independent cases), per tumor sub-compartment (ET enhancing tumor, TC tumor core, WT whole tumor). Note the box and whiskers inside each violin plot represent the true min and max values. The top and bottom of each "box" depict the 3rd and 1st quartile of each measure. The white line and the red '×', within each box, indicate the median and mean values, respectively. The fact that these are not necessarily at the center of each box indicates the skewness of the distribution over different cases. The "whiskers" drawn above and below each box depict the extremal observations still within 1.5 times the interquartile range, above the 3rd or below the 1st quartile. Equivalent plots for the Jaccard similarity coefficient (JSC) can be observed in supplementary figures. **e** Number of contributed cases per collaborating site.

and familiar with computational analyses), and coordinated independent model training for each, starting from the public initial model and using only their local training data. The findings of this evaluation indicate that the final consensus model performance is always superior or insignificantly different ($p_{Average} = 0.1$, $p_{ET} = 0.5$, $p_{TC} = 0.2$, $p_{WT} = 0.06$, Wilcoxon signed-rank test) to the ensemble of the local models of these four largest contributing collaborators, for all tumor sub-compartments (Fig. 2). This finding highlights that even large sites can benefit from collaboration.

### FL is robust to data quality issues

Data quality issues relating to erroneous reference annotations (with potential negative downstream effects on output predictions) were identified by monitoring the global consensus model performance during training. However, only data quality issues that largely affected the global validation score could be identified and corrected during training. Those with more subtle effects in the global validation score were only identified after the completion of the model training by looking for relatively low local validation scores of the consensus

https://doi.org/10.1038/s41467-022-33407-5

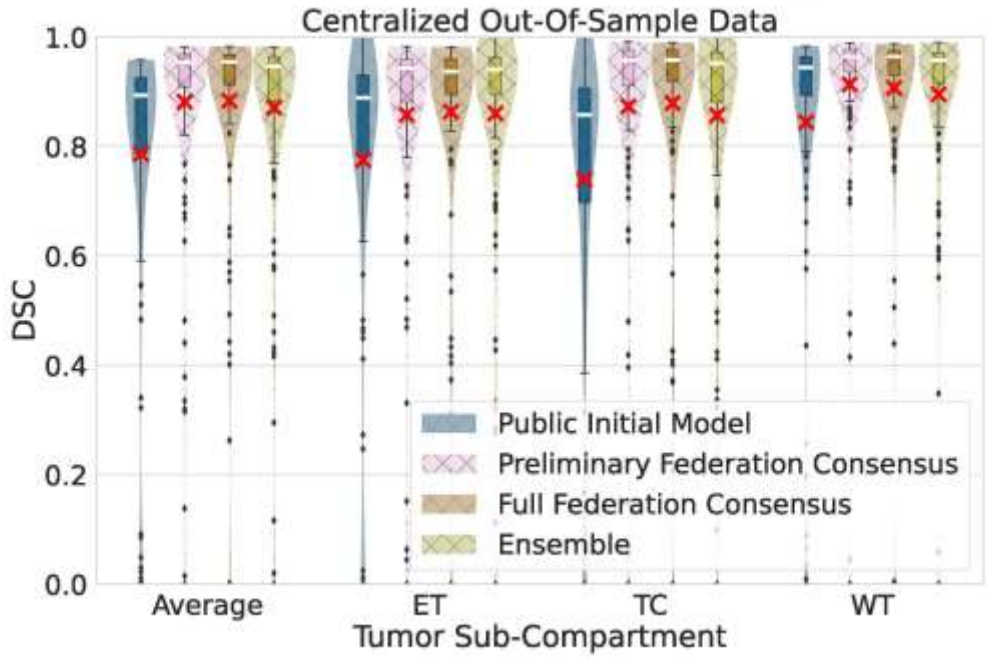



**Fig. 2 | Generalizable Dice similarity coefficient (DSC) evaluation on 'centralized' out-of-sample data ($n = 154$ biologically independent cases), per tumor sub-compartment (ET enhancing tumor, TC tumor core, WT whole tumor) and averaged across cases.** Comparative performance evaluation across the public initial model, the preliminary consensus model, the final consensus model, and an ensemble of single site models from collaborators holding > 200 cases. Note the box and whiskers inside each violin plot, represent the true min and max values. The top and bottom of each "box" depict the 3rd and 1st quartile of each measure. The white line and the red '×', within each box, indicate the median and mean values, respectively. The fact that these are not necessarily at the center of each box indicates the skewness of the distribution over different cases. The "whiskers" drawn above and below each box depict the extremal observations still within 1.5 times the interquartile range, above the 3rd or below the 1st quartile. Equivalent plots for Jaccard similarity coefficient (JSC) can be observed in supplementary figures.

model across collaborating sites. An example of such a quality issue with erroneous reference labels (from Site 48) is shown in Fig. 3c. Looking closer, local validation scores at Site 48 (Fig. 3b) are significantly different ($p_{ET} < 3 \times 10^{-12}$, $p_{TC} < 3 \times 10^{-12}$, $p_{WT} < 3 \times 10^{-12}$, Wilcoxon signed-rank test) than the average scores across the federation (Fig. 3a). Significant differences were calculated by sample pairs for each federated round, where a sample pair consists of the mean validation score over samples for Site 48 paired with those across all sites. These local validation scores (Fig. 3b) indicate that the model is not gaining knowledge from these local data, and their comparison with the average scores across the federation (Fig. 3a) indicates that the global consensus model performance is not adversely affected. This finding supports the importance of robustness at a global scale.

### FL benefits the more challenging tasks

The complexity of boundary detection drops when moving from smaller to larger sub-compartments, i.e., from ET to TC, and then to WT[33–35]. This is further confirmed here, as evidenced by the model's relative performance indicated by the local validation curves and their underlying associated areas in Fig. 3.a. Since the current clinically actionable sub-compartments are TC (i.e., considered for surgery) and WT (i.e., considered for radiotherapy)[42], performance improvements of their boundary detection may contribute to the model's clinical impact and relevance.

Our findings indicate that the benefits of FL are more pronounced for the more challenging sub-compartments, i.e., larger performance improvements for ET and TC compared to WT (Fig. 1c). Notably, the largest and most significant improvement (33%, $p < 7 \times 10^{-60}$) is noted for the TC sub-compartment, which is surgically actionable and not a trivial sub-compartment to delineate accurately[43,44]. This finding of FL benefiting the more challenging tasks rather than boosting performance on the relatively easier task (e.g., thresholding the abnormal T2-FLAIR signal for the WT sub-compartment) by gaining access to larger amounts of good quality data holds a lot of promise for FL in healthcare.

### Optimal model selection is non-trivial

Using the performance of the global consensus model during training across all local validation cases, two distinct model configurations were explored for selecting the final consensus model. Analyzing the sequence of consensus models produced during each federated round, we selected four different models: the *singlet*, for which the average DSC across all sub-compartments scored high, and three independent models, each of which yielded high *DSC* scores for each tumor sub-compartment, i.e., ET, TC, WT. We defined the collection of these three independent consensus models as a *triplet*.

To identify the best model, 5 *singlets* and 5 *triplets* were selected based on their relative performance on all local validation cases and

https://doi.org/10.1038/s41467-022-33407-5

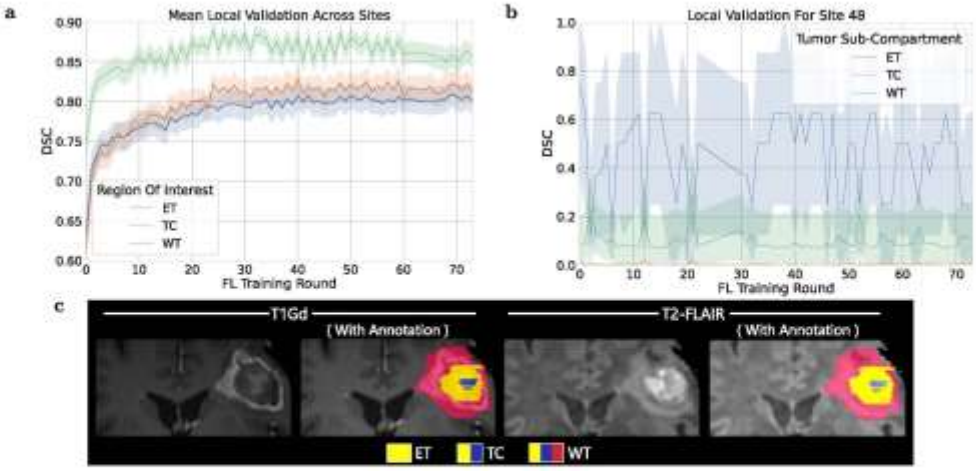



**Fig. 3 | Per-tumor region (ET enhancing tumor, TC tumor core, WT whole tumor) mean Dice similarity coefficient (*DSC*) over validation samples (with shading indicating 95% confidence intervals again over samples).** **a** At all participating sites across training rounds showing that the score is greater for sub-compartments with larger volumes. **b** For a site with problematic annotations (Site 48). The instability in these curves could be caused by errors in annotation for the local validation data (similar to errors that were observed for a small shared sample of data from this site). **c** Provides an example of a case with erroneous annotations in the data used by Site 48. Equivalent plots for Jaccard similarity coefficient (JSC) can be observed in supplementary figures.

evaluated against the centralized out-of-sample data. Only small differences are observed between the *singlet* and *triplet* models, and these differences diminish as the sub-compartment size increases. Comparing the means of *singlet* and *triplet*, the larger (and only significant) performance improvement difference compared to the public initial model is noted for the ET sub-compartment (improved by < 3%, $p_{ET} = 0.02$), followed by TC (improved by < 1.4%, $p_{TC} = 0.09$), and then lastly WT (improved by < 1.1%, $p_{WT} = 0.2$) (Tables S1 and S2). However, the decision of using a *singlet* or a *triplet* model should also rely on computational cost considerations, as *triplets* will be three times more expensive than *singlets* during model inference.

## Discussion

In this study, we have described the largest real-world FL effort to-date utilizing data of 6314 glioblastoma patients from 71 geographically unique sites spread across 6 continents, to develop an accurate and generalizable ML model for detecting glioblastoma sub-compartment boundaries. Notably, this extensive global footprint of the collaborating sites in this study also yields the largest dataset ever reported in the literature assessing this rare disease. It is the use of FL that successfully enabled (i) access to such an unprecedented dataset of the most common and fatal adult brain tumor, and (ii) meaningful ML training to ensure the generalizability of models across out-of-sample data. In comparison with the limited existing real-world FL studies[26,30], our use-case is larger in scale and substantially more complex, since it (1) addresses a multi-parametric multi-class problem, with reference standards that require expert collaborating clinicians to follow an involved manual annotation protocol, rather than simply recording a categorical entry from medical records, and (2) requires the data to be preprocessed in a harmonized manner to account for differences in MRI acquisition. Since glioblastoma boundary detection is critical for treatment planning and the requisite first step for further quantitative analyses, the models generated during this study have the potential to make a far-reaching clinical impact.

The large and diverse data that FL enabled, led to the final consensus model garnering significant performance improvements over the public initial model against both the collaborators' local validation data and the complete out-of-sample data. The improved result is a clear indication of the benefit that can be afforded through access to more data. However, increasing the data size for model training without considerations relating to data quality, reference labels, and potential site bias (e.g., scanner acquisition protocols, demographics, or sociocultural considerations, such as more advanced presentation of disease at diagnosis in low-income regions[45]) might not always improve results. Literature also indicates an ML performance stagnation effect, where each added case contributes less to the model performance as the number of cases increase[46]. This is in line with our finding in the federated setting (Fig. 2), where performance improvements across the public initial model, the preliminary consensus model, and the final consensus model, were not directly/linearly related to the amount of data used for training. This happened even though the final consensus model was trained on over twice the number of cases (and included 2 of the largest contributing sites – Sites 1 and 4) when compared to the preliminary consensus model. Further noting that the preliminary federation model was already within the intra- and inter-rater variability range for this use-case (20% and 28%, respectively)[47], any further improvements for the full federation consensus model would be expected to be minimal[11–15].

To further assess these considerations, we coordinated independent model training for the four largest collaborating sites (i.e., >200 cases) by starting from the same public initial model and using only their local training data. The ensemble of these four largest site local models did not show significant performance differences to the final consensus model for any tumor sub-compartment, yet the final consensus model showed superior performance indicating that even sites with large datasets can benefit from collaboration. The underlying assumption for these results is that since each of these collaborators initiated their training from the public initial model (which included diverse data from 16 sites), their independent models and their ensemble could have inherited some of the initial model's data diversity, which could justify the observed insignificant differences (Fig. 2 and Supplementary Fig. 3). Though these findings are an indication

that the inclusion of more data alone may not lead to better performance, it is worth noting that these four largest sites used for the independent model training represent comprehensive cancer centers (compared to hospitals in community settings) with affiliated sophisticated labs focusing on brain tumor research, and hence were familiar with the intricacies of computational analyses. Further considering the aforementioned ML performance stagnation effect, we note the need for generalizable solutions to quantify the contribution of collaborating sites to the final consensus model performance, such that future FL studies are able to formally assess both the quantity and the quality of the contributed data needed by the collaborating sites and decide on their potential inclusion on use-inspired studies.

As noted in our results, due to the lack of such generalizable solutions, we were only able to identify quality issues after the model training. Specifically, we hypothesize that although Site 48 had data quality issues, its effect on the consensus model performance was not significant due to its relatively small dataset ($n = 46$) when compared to the other collaborating sites. The curves of Fig. 3a indicate that the global consensus model continues to consistently gain knowledge from the federation as a whole during training, highlighting robustness to such data quality issues. It remains unknown, however, how much better the consensus model would have performed if sites with problematic data were excluded or if these specific problematic data at Site 48 were excluded or corrected. These findings are aligned with literature observations (on the same use-case)[48], where a DL model[49] trained on 641 glioblastoma cases from 8 sites produced higher quality predictions on average than those created as reference standard labels by radiology expert operators. Quality was judged by 20 board-certified neuroradiologists, in a blinded side-by-side comparison of 100 sequestered unseen cases, and concluded that perfect or near-perfect reference labels may not be required to produce high-quality prediction systems. In other words, DL models may learn to see past imperfect reference training labels. These findings provide the impetus for further experimentation as they have implications for future FL studies. Future research is needed to automatically detect anomalies in the consensus model performance during training, particularly associated with contributions from individual sites.

There are a number of practical considerations that need to be taken into account to set up a multi-national real-world federation, starting with a substantial amount of coordination between each participating site. As this study is the first at this scale and task complexity, we have compiled a set of governance insights from our experience that can serve as considerations for future successful FL studies. These insights differ from previous literature that describes studies that were smaller in scale and involved simpler tasks[50,52]. By "governance" of the federation we refer both to the accurate definition of the problem statement (including reference labels and harmonization considerations accounting for inter-site variability), and the coordination with the collaborating sites for eligibility and compliance with the problem statement definition, as well as security and technical considerations. For future efforts aiming to conduct studies of a similar global scale, it would be beneficial to identify a solution for governance prior to initiating the study itself.

The coordination began with engaging the security teams of collaborating sites and providing them access to the source code of the platform developed to facilitate this study. These security discussions highlighted the benefit of the platform being open-source, making security code reviews easier. Resource gathering was then carried out by identifying technical leads and assessing computational resources at each site. With the technical leads, we then proceeded to test the complete workflow to further identify gaps in the requirements, such as network configurations and hardware requirements. We then proceeded with data curation and preprocessing, and finally connected individual sites to the aggregation server to initiate their participation.

Following the precise definition of our problem statement[31–33], ensuring strict compliance with the preprocessing and annotation protocol for the generation of reference standards was vital for the model to learn correct information during training. To this end, we instituted an extensively and comprehensively documented annotation protocol with visual example representations and common expected errors (as observed in the literature[38,50]) to all collaborators. We have further circulated an end-to-end platform[50] developed to facilitate this federation, providing to each collaborating site all the necessary functionalities to (i) uniformly curate their data and account for inter-site acquisition variability, (ii) generate the reference standard labels, and (iii) participate in the federated training process. Finally, we held interactive sessions to complement the theoretical definition of the reference standards, and further guide collaborating sites. Particular pain points regarding these administrative tasks included managing the large volume of communication (i.e., emails and conference calls) needed to address questions and issues that arose, as well as the downtime incurred in FL training due to issues that had not yet been identified and were adversely affecting the global model. Though we developed many ad-hoc tools for this workflow ourselves (particularly for the data processing and orchestration steps), many issues we encountered were common enough in retrospect (for example common Transport Layer Security (TLS) errors) that mature automated solutions will address them. Many of these automations will be use-case dependent, such as the MRI data corruption checks we used from the FeTS tool[50]. For these use-case-dependent automation, more associated tools are expected to become available as various domain experts enter into the FL community, while some will be more general purpose. As our inspection of both local and global model validation scores was manual during our deployment, we in retrospect see great value in automated notifications (performed at the collaborator infrastructure to help minimize data information leakage) to alert a collaborator (or the governor) when their local or global model validation is significantly low. Such an alert can indicate the potential need to visually inspect example failure cases in their data for potential issues. With continued efforts towards developing automated administration tools around FL deployments, we expect the coordination for large FL deployments to become easier.

In general, debugging issues with the inputted local data and annotations is more difficult during FL due to the level of coordination and/or privacy issues involved, since the data are always retained at the collaborating site. We gained substantial experience during this effort that went into further development of use-inspired but generalizable data sanity-checking functionality in the tools we developed, towards facilitating further multi-site collaborations.

Upon conclusion of the study, sites participating in the model training process were given a survey to fill in regarding various aspects of their experience. According to the provided feedback, 96% of the sites found the comprehensive documentation on preprocessing and data curation essential and thought that lack of such documentation could have resulted in inconsistent annotations. Additionally, 92% found the documentation relating to establishing secure connectivity to the aggregation server easy to follow and essential to expedite reviews by the related groups. Furthermore, 84% of the sites appreciated the user-friendly interface of the provided tool and its associated complete functionality (beyond its FL backend), and indicated their intention to use it and recommend it for projects and data analysis pipelines beyond the scope of this study. To generate the reference standard labels for their local data, 86% of the collaborating sites indicated that they used either the FeTS Tool[18] (i.e., the tool developed for this study), CaPTk[51], or ITK-SNAP[52], whereas the remaining 14% used either 3D-Slicer[53], the BraTS toolkit[54], or something else. In terms of hardware requirements at each site, 88% used a dedicated workstation for their local workload, and the remaining 12% used either a containerized form of the FeTS tool or a virtual machine.

Although data are always retained within the acquiring site during FL (and hence FL is defined as private-by-design), different security and privacy threats remain[55–57]. These threats include attempted extraction of training data information from intermediate and final models, model theft, and submission of poison model updates with the goal of introducing unwanted model behavior (including incentivizing the model to memorize more information about the training data in support of subsequent extraction, i.e., leakage). A number of technologies can be used to mitigate security and privacy concerns during FL[55–57]. Homomorphic encryption[59], secure multiparty compute[58], and trusted execution environments (TEEs)[60,61] allow for collaborative computations to be performed with untrusted parties while maintaining confidentiality of the inputs to the computation. Differentially private training algorithms[62–64] allow for mitigation of information leakage from both the collaborator model updates and the global consensus aggregated models. Finally, assurance that remote computations are executed with integrity can be designed for with the use of hardware-based trust provided by TEEs, as well as with some software-based integrity checking[65]. Each of these technologies comes with its own benefits in terms of security and/or privacy, as well as costs and limitations, such as increased computational complexity, associated hardware requirements and/or reduced quality of computational output (such as the reduction of model utility that can be associated with differentially private model training). Further experimentation needs to be done in order to best inform prospective federations as to which technologies to use towards addressing their specific concerns within the context of the collaborator infrastructure and trust levels, depending on the use-case, the extent of the collaborating network, and the level of trust within the involved parties. Our study was based on a collaborative network of trusted sites, where authentication was based on personal communication across collaborating sites and the combination of TLS and TEEs were considered sufficient.

Although our study has the potential to become the baseline upon which future ML research studies will be done, there is no automated mechanism to assess inputted data quality from collaborators, which could result in models trained using sub-optimal data. Additionally, we used a single off-the-shelf neural network architecture for training, but it has been shown that model ensembles perform better for the task at hand[35–38], and it remains to be explored how such a strategy could be explored in a federated study. Moreover, the instantiation of the federation involved a significant amount of coordination between each site and considering the limited real-world FL studies at the time, there were no tools available to automate such coordination and orchestration. These involved (i) getting interviewed by information security officers of collaborating sites, (ii) ensuring that the harmonized preprocessing pipeline was used effectively, (iii) clear communication of the annotation protocol, and iv) testing the network communication between the aggregator and each site. This amount of effort, if not aided by automated tools, will continue to be a huge roadblock for FL studies, and dedicated coordination and orchestration resources are required to conduct this in a reproducible and scalable manner.

We have demonstrated the utility of an FL workflow to develop an accurate and generalizable ML model for detecting glioblastoma subcompartment boundaries, a finding which is of particular relevance for neurosurgical and radiotherapy planning in patients with this disease. This study is meant to be used as an example for future FL studies between collaborators with an inherent amount of trust that can result in clinically deployable ML models. Further research is required to assess privacy concerns in a detailed manner[55,64] and to apply FL to different tasks and data types[66–69]. Building on this study, a continuous FL consortium would enable downstream quantitative analyses with implications for both routine practice and clinical trials, and most importantly, increase access to high-quality precision care worldwide. Furthermore, the lessons learned from this study with such a global footprint are invaluable and can be applied to a broad array of clinical scenarios with the potential for great impact on rare diseases and underrepresented populations.

## Methods

The study and results presented in this manuscript comply with all relevant ethical regulations and follow appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding the treatment of human subjects. Use of the private retrospective data collection of each collaborating site has been approved by their respective institutional review board, where informed consent from all participants was also obtained and stored.

### Data

The data considered in this study described patient populations with adult-type diffuse glioma[70], and specifically displaying the radiological features of glioblastoma, scanned with mpMRI to characterize the anatomical tissue structure[71]. Each case is specifically described by (i) native T1-weighted (T1), (ii) Gadolinium-enhanced T1-weighted (T1Gd), (iii) T2-weighted (T2), and (iv) T2-weighted-Fluid-Attenuated-Inversion-Recovery (T2-FLAIR) MRI scans. Cases with any of these sequences missing were not included in the study. Note that no inclusion/exclusion criterion applied relating to the type of acquisition (i.e., both 2D axial and 3D acquisitions were included, with a preference for 3D if available), or the exact type of sequence (e.g., MP-RAGE vs. SPGR). The only exclusion criterion was for T1-FLAIR scans that were intentionally excluded to avoid mixing varying tissue appearance due to the type of sequence, across native T1-weighted scans.

The publicly available data from the International Brain Tumor Segmentation (BraTS) 2020 challenge[34–37], was used to train the public initial model of this study. The BraTS challenge[35–38], seeking methodological advancements in the domain of neuro-oncology, has been providing the community with (i) the largest publicly available and manually-curated mpMRI dataset of diffuse glioma patients (an example of which is illustrated in Fig. 1b), and (ii) a harmonized preprocessing pipeline[34,36,38] to handle differences in inter-site acquisition protocols. The public initial model was used to initialize the FL training, instead of a randomly generated initialization, as starting from a pre-trained model leads to faster convergence[41]. The complete BraTS 2020 dataset originally included cases from sites that also participated in this study as independent collaborators. To avoid any potential data leakage, we reduced the size of the complete BraTS dataset by removing cases acquired by these specific sites, resulting in a dataset of 231 cases from 16 international sites, with varying contributing cases across sites (Fig. 1e). The exact site IDs that construct the data of the public initial model are: 47, 51, 55, 57, 58, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, and 71. Subsequently, the resulting dataset was split at a 4:1 ratio between cases for training ($n = 185$) and validation ($n = 46$).

The eligibility of collaborating sites to participate in the federation was determined based on data availability, and approval by their respective institutional review board. 55 sites participated as independent collaborators in the study defining a dataset of 6083 cases. The MRI scanners used for data acquisition were from multiple vendors (i.e., Siemens, GE, Philips, Hitachi, Toshiba), with magnetic field strength ranging from 1T to 3T. The data from all 55 collaborating sites followed a male:female ratio of 1.47:1 with ages ranging between 7 and 94 years.

From all 55 collaborating sites, 49 were chosen to be part of the training phase, and 6 sites were categorized as *"out-of-sample"*, i.e., none of these were part of the training stage. These specific 6 out-of-sample sites (Site IDs: 8, 11, 19, 20, 21, 43) were allocated based on their availability, i.e., they have indicated expected delayed participation rendering them optimal for model generalizability validation. One of these 6 out-of-sample sites (Site 11) contributed aggregated a priori data from a multi-site randomized clinical trial for newly diagnosed

glioblastoma (ClinicalTrials.gov Identifier: NCT00884741, RTOG0825[73,75], ACRIN6686[74,76]), with inherent diversity benefiting the intended generalizability validation purpose. The American College of Radiology (ACR - Site 11) serves as the custodian of this trial's imaging data on behalf of ECOG-ACRIN, which made the data available for this study. Following screening for the availability of the four required mpMRI scans with sufficient signal-to-noise ratio judged by visual observation, a subset of 362 cases from the original trial data were included in this study. The out-of-sample data totaled 590 cases intentionally held out of the federation, with the intention of validating the consensus model in completely unseen cases. To facilitate further such generalizability evaluation without burdening the collaborating sites, a subset consisting of 332 cases (including the multi-site clinical data provided by ACR) from this out-of-sample data was aggregated, to serve as the "centralized out-of-sample" dataset. Furthermore, the 49 sites participating in the training phase define a collective dataset of 5493 cases. The exact 49 site IDs are: 1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15, 16, 17, 18, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 48, 49, 50, 52, 53, 54, 56, 59, 60. These cases were automatically split at each site following a 4:1 ratio between cases for training and local validation. During the federated training phase, the data used for the public initial model were also included as a dataset from a separate node, such that the contribution of sites providing the publicly available data is not forgotten within the global consensus model. This results in the final consensus model being developed based on data from 71 sites over a total dataset of 6314 cases. Collective demographic information of the included population is provided in Table S3.

### Harmonized data preprocessing

Once each collaborating site identified its local data, they were asked to use the preprocessing functionality of the software platform we provided. This functionality follows the harmonized data preprocessing protocol defined by the BraTS challenge[18–19], as described below. This would allow accounting for inter-site acquisition protocol variations, e.g., 3D vs. 2D axial plane acquisitions.

**File-type conversion/patient de-identification.** The respective mpMRI scans (i.e., T1, T1Gd, T2, T2-FLAIR) of every case are downloaded onto a local machine in the Digital Imaging and Communications in Medicine (DICOM) format[77–78] and converted to the Neuroimaging Informatics Technology Initiative (NIfTI) file format[79] to ensure easier parsing of the volumetric scans during the computational process. The conversion of DICOM to NIfTI files has the benefit of eliminating all patient-identifiable metadata from the header portion of the DICOM format[80,81].

**Rigid registration.** Once the scans are converted to the NIfTI format, each volume is registered to a common anatomical space, namely the SRI24 atlas[82], to ensure a cohesive data shape ([240, 240, 155]) and an isotropic voxel resolution (1 $mm^3$), thereby facilitating in the tandem analysis of the mpMRI scans. One of the most common types of MRI noise is based on the inhomogeneity of the magnetic field[83]. It has been previously[34] shown that the use of non-parametric, non-uniform intensity normalization to correct for these bias fields[84,85] obliterates the MRI signal relating to the regions of abnormal T2-FLAIR signal. Here, we have taken advantage of this adverse effect and used the bias field-corrected scans to generate a more optimal rigid registration solution across the mpMRI sequences. The bias field-corrected images are registered to the T1Gd image, and the T1Gd image is rigidly registered to the SRI24 atlas, resulting in two sets of transformation matrices per MRI sequence. These matrices are then aggregated into a single matrix defining the transformation of each MRI sequence from its original space to the atlas. We then apply this single aggregated matrix to the NIfTI scans prior to the application of the bias field correction to maximize the fidelity of the finally registered images.

**Brain extraction.** This process focuses on generating a brain mask to remove all non-brain tissue from the image (including neck, fat, eyeballs, and skull), to enable further computational analyses while avoiding any potential face reconstruction/recognition[86]. For this step we utilized the Brain Mask Generator (BrainMaGe)[87], which has been explicitly developed to address brain scans in presence of diffuse glioma and considers brain shape as a prior, hence being agnostic to the sequence/modality input.

**Generation of automated baseline delineations of tumor sub-compartment boundaries.** We provided the ability to the collaborating sites to generate automated delineations of the tumor sub-compartments from three popular methods from the BraTS challenge, using models trained using the challenge's training data: (i) DeepMedic[88], (ii) DeepScan[89], and (iii) nnU-Net[90]. Along with segmentations from each method, label fusion strategies were also employed to provide a reasonable approximation to the reference labels that should be manually refined and approved by expert neuroradiologists to create the final reference labels. The label fusion approaches considered were i) standard voting[91], (ii) Simultaneous Truth And Performance Level Estimation (STAPLE)[91,92], iii) majority voting[91], and iv) Selective and Iterative Method for Performance Level Estimation (SIMPLE)[24].

**Manual refinements towards reference standard labels.** It was communicated to all participating sites to leverage the annotations generated using the automated mechanism as a baseline on which manual refinements were needed by neuroradiology experts, following a consistently communicated annotation protocol. The reference annotations comprised the Gd-enhancing tumor (ET—label '4'), the peritumoral edematous/invaded tissue (ED—label '2'), and the necrotic tumor core (NCR—label '1'). ET is generally considered the most active portion of the tumor, described by areas with both visually avid, as well as faintly avid, enhancement on the T1Gd scan. NCR is the necrotic part of the tumor, the appearance of which is hypointense on the T1Gd scan. ED is the peritumoral edematous and infiltrated tissue, defined by the abnormal hyperintense signal envelope on the T2-FLAIR scans, which includes the infiltrative non-enhancing tumor, as well as vasogenic edema in the peritumoral region[35–38] (an illustration can be seen in Fig. 1b).

**Data splits.** Once the data were preprocessed, training and validation cohorts were created randomly in a 4:1 ratio, and the splits were preserved during the entire duration of the FL training to prevent data leakage. The performance of every model was compared against the local validation data cohort on every federated round.

### Data loading and processing

We leveraged the data loading and processing pipeline from the Generally Nuanced Deep Learning Framework (GaNDLF)[93], to enable experimentation with various data augmentation techniques. Immediately after data loading, we removed the all-zero axial, coronal, and sagittal planes from the image, and performed a z-score normalization of the non-zero image intensities[94]. Each tumor sub-compartment of the reference label is first split into an individual channel and then passed to the neural network for processing. We extracted a single random patch per mpMRI volume set during every federated round. The patch size was kept constant at [128, 128, 128] to ensure that the trained model can fit the memory of the baseline hardware requirement of each collaborator, i.e., a discrete graphics processing unit with a minimum of 11 GB dedicated memory. For data augmentation, we added random noise augmentation ($\mu = 0.0$, $\sigma = 0.1$) with a probability

of $p = 0.2$, random rotations (90° and 180°, with the axis of rotation being uniformly selected in each case from the set of coronal, sagittal, and axial planes) each with a probability of $p = 0.5$, and a random flip augmentation with a probability of $p = 1.0$ with equal likelihood of flips across the sagittal, coronal, and axial planes.

### The neural network architecture

The trained model to delineate the different tumor sub-compartments was based on the popular 3D U-Net with residual connections (3D-ResUNet)[97–101], an illustration of which can be seen in the Supplementary Fig. 1. The network had 30 base filters, with a learning rate of $lr = 5 \times 10^{-5}$ optimized using the Adam optimizer[102]. For the loss function used in training, we used the generalized DSC score[103,104] (represented mathematically in Eq. (1)) on the absolute complement of each tumor sub-compartment independently. Such mirrored DSC loss has been shown to capture variations in smaller regions better[105]. No penalties were used in the loss function, due to our use of 'mirrored' DSC loss[106–107]. The final layer of the model was a sigmoid layer, providing three channel outputs for each voxel in the input volume, one output channel per tumor sub-compartment. While the generalized *DSC* score was calculated using a binarized version of the output (check sigmoid value against the threshold 0.5) for the final prediction, we used the floating point *DSC*[108] during the training process.

$$\mathrm{DSC} = \frac{2|\mathrm{RL} \odot \mathrm{PM}|_1}{|\mathrm{RL}|_1 + |\mathrm{PM}|_1} \quad (1)$$

where RL serves as the reference label, PM is the predicted mask, $\odot$ is the Hadamard product[109] (i.e., component-wise multiplication), and $|x|_1$ is the L1-norm[110], i.e., the sum of the absolute values of all components).

### The Federation

The collaborative network of the present study spans 6 continents (Fig. 1), with data from 71 geographically distinct sites. The training process was initiated when each collaborator securely connected to a central aggregation server, which resided behind a firewall at the University of Pennsylvania. We have identified this FL workflow (based on a central aggregation server) as the optimal for this use-case, following a performance evaluation[11] for this very same task, i.e., detecting glioblastoma sub-compartment boundaries. As soon as the secure connection was established, the public initial model was passed to the collaborating site. Using FL based on an aggregation server (refer to supplementary figures for illustration), collaborating sites then trained the same network architecture on their local data for one epoch, and shared model updates with the central aggregation server. The central aggregation server received model updates from all collaborators, combined them (by averaging model parameters) and sent the consensus model back to each collaborator to continue their local training. Each such iteration is called a "federated round". Based on our previously conducted performance evaluation for this use-case[11], we chose to perform aggregation of all collaborator updates in the present study, using the federated averaging (FedAvg) approach[14], i.e., average of collaborator's model updates weighted according to collaborator's contributing data. We expect these aggregation strategy choices to be use-case dependent, by providing due consideration to the collaborators' associated compute and network infrastructure. In this study, all the network communications during the FL model training process were based on TLS[111], to mitigate potential exposure of information during transit. Additionally, we demonstrated the feasibility of TEEs[112,113] for federated training by running the aggregator workload on the secure enclaves of Intel's Secure Guard Extensions (SGX) hardware (Intel® Xeon® E-2286M vPro 8-Core 2.4-5.0GHz Turbo), which ensured the confidentiality of the updates being aggregated and the integrity of the consensus model. TLS and TEEs can help mitigate some of the security and privacy concerns that remain for FL[11]. After not observing any meaningful changes since round 42, we stopped the training after a total of 73 federated rounds. Additionally, we performed all operations on the aggregator on secure hardware (TEE[112]), in order to increase the trust by all parties in the confidentiality of the model updates being computed and shared, as well as to increase the confidence in the integrity of the computations being performed[115].

We followed a staged approach for the training of the global consensus model, starting from a preliminary smaller federation across a subset ($n = 35$) of the participating sites to evaluate the complete process and resolve any initial network issues. Note that 16 of these 35 sites were used to train the public initial model, and used in the preliminary federation as an aggregated dataset. The exact 19 site IDs that participated in the training phase of the preliminary federation, as independent sites are: 2, 3, 9, 14, 22, 23, 24, 27, 28, 29, 31, 33, 36, 37, 41, 46, 53, 54, and 59. The total data held by this smaller federation represented approximately 42% ($n = 2471$) of the data used in the full federation. We also trained individual models (initialized using the public initial model) using centralized training at all sites holding >200 training cases, and performed a comparative evaluation of the consensus model with an ensemble of these "single site models". The per voxel sigmoid outputs of the ensemble were computed as the average of such outputs over the individual single-site models. As with all other models in this study, binary predictions were computed by comparing these sigmoid outputs to a threshold value of 0.5. The single-site model ensemble utilized (via the data at the single site) approximately 33% of the total data across the federation.

### Model runtime in low-resource settings

Clinical environments typically have constrained computational resources, such as the availability of specialized hardware (e.g., DL acceleration cards) and increased memory, which affect the runtime performance of DL inference workloads. Thus, taking into consideration the potential deployment of the final consensus model in such low-resource settings, we decided to proceed with a single 3D-ResU-Net, rather than an ensemble of multiple models. This decision ensured a reduced computational burden when compared with running multiple models, which is typically done in academic research projects[18–19].

To further facilitate use in low-resource environments, we have provided a post-training run-time optimized[114] version of the final consensus model. Graph level optimizations (i.e., operators fusion) were initially applied, followed by optimizations for low precision inference, i.e., converting the floating point single precision model to a fixed precision 8-bit integer model (a process known as "quantization"[115]). In particular, we used accuracy-aware quantization[116], where model layers were iteratively scaled to a lower precision format. These optimizations yielded run-time performance benefits, such as lower inference latency (a platform-dependent 4.48 × average speedup and 2.29 × reduced memory requirement when compared with the original consensus model) and higher throughput (equal to the 4.48 × speedup improvement since the batch size used is equal to 1), while the trade-off was an insignificant ($p_{\mathrm{Average}} < 7 \times 10^{-4}$) drop in the average *DSC*.

**Clinically-deployable consensus models.** To further encourage the reproducibility of our study, and considering enhancing the potential impact for the study of the rare disease of glioblastoma, we publicly released the trained models of this study. We specifically released the final *singlet* and *triplet* consensus models, including the complete source code used in the project. Taking into consideration the potential deployment of these models in clinical settings, we refrained from training an ensemble of models (as typically done in academic

research projects[36–38]), due to the additional computational burden of running multiple models. Furthermore, to facilitate use in low-resource environments, we also provide a post-training run-time optimized[104] version of the final consensus model that obviates the need for any specialized hardware (such as DL acceleration cards) and performs insignificantly different from the final consensus model when evaluated against the centralized out-of-sample data.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The datasets used in this study, from the 71 participating sites, are not made publicly available as a collective data collection due to restrictions imposed by acquiring sites. The public initial model data from 16 sites are publicly available through the BraTS challenge[35–38] and are available from https://www.med.upenn.edu/cbica/brats2020. The data from each of the 55 collaborating sites were neither publicly available during the execution of the study, nor shared among collaborating sites or with the aggregator. They were instead used locally, within each of the acquiring sites, for the training and validation of the global consensus model at each federated round. The anatomical template used for co-registration during preprocessing is the SRI24 atlas[32] and is available from https://www.nitrc.org/projects/sri24.

Source data are provided with this paper. Specifically, we provide the raw data, the associated python scripts, and specific instructions to reproduce the plots of this study in a GitHub repository, at: github.com/FETS-AI/2022_Manuscript_Supplement. The file 'SourceData.tgz', in the top directory holds an archive of csv files representing the source data. The python scripts are provided in the 'scripts' folder which utilize these source data and save '.png' images to disc and/or print latex code (for tables) to stdout. Furthermore, we have provided three sample validation cases, from the publicly available BraTS dataset, to qualitatively showcase the segmentation differences (small, moderate, and large) across the final global consensus model, the public initial model, and the ground truth annotations in the same GitHub repository.

## Code availability

Motivated by findability, accessibility, interoperability, and reusability (FAIR) criteria in scientific research[117], all the code used to design the Federated Tumor Segmentation (FeTS) platform[118] for this study is available through the FeTS Tool[119] and it is available at github.com/FETS-AI/Front-End. The functionality related to preprocessing (i.e., DICOM to NIfTI conversion, population-based harmonized preprocessing, co-registration) and manual refinements of annotation is derived from the open-source Cancer Imaging Phenomics Toolkit (CaPTk, github.com/CBICA/CaPTk)[69,70,71]. The co-registration is performed using the Greedy framework[109], available via CaPTk[69,70,71], ITK-SNAP[57], and the FeTS Tool[119]. The brain extraction is done using the BrainMaGe method[67], and is available at github.com/CBICA/BrainMaGe, and via GaNDLF[98] at github.com/mlcommons/GaNDLF. To generate automated annotations, DeepMedic's[93] integration with CaPTk was used, and we used the model weights and inference mechanism provided by the other algorithm developers (DeepScan[94] and nnU-Net[95] (github.com/MIC-DKFZ/nnunet)). DeepMedic's original implementation is available in github.com/deepmedic/deepmedic, whereas the one we used in this study can be found at github.com/CBICA/deepmedic. The fusion of the labels was done using the Label Fusion tool[120] available at github.com/FETS-AI/LabelFusion. The data loading pipeline and network architecture were developed using the GaNDLF framework[98] by using PyTorch[121]. The data augmentation was done via GaNDLF by leveraging TorchIO[122]. The FL backend developed for this project has been open-sourced as a separate software library, to encourage further research on FL[123] and is available at github.com/intel/openfl. The optimization of the consensus model inference workload was performed via OpenVINO[124] (github.com/openvinotoolkit/openvino/tree/2021.4.1), which is an open-source toolkit enabling acceleration of neural network models through various optimization techniques. The optimizations were evaluated on an Intel Core® i7-1185G7E CPU @ 2.80 GHz with 2 × 8 GB DDR4 3200 MHz memory on Ubuntu 18.04.6 OS and Linux kernel version 5.9.0-050900-generic.

## References


1. Mårtensson, G. et al. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. *Med. Image Anal.* **66**, 101714 (2020).
2. Zech, J. R. et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med.* **15**, e1002683 (2018).
3. Obermeyer, Z. & Emanuel, E. J. Predicting the future-big data, machine learning, and clinical medicine. *New Engl. J. Med.* **375**, 1216 (2016).
4. Marcus, G. Deep learning: a critical appraisal. arXiv preprint arXiv:1801.00631 (2018).
5. Aggarwal, C. C. et al. *Neural Networks and Deep Learning* Vol. 10, 978–983 (Springer, 2018).
6. Thompson, P. M. et al. The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
7. Consortium, T. G. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro-Oncology* **20**, 873–884 (2018).
8. Davatzikos, C. et al. AI-based prognostic imaging biomarkers for precision neuro-oncology: the respond consortium. *Neuro-oncology* **22**, 886–888 (2020).
9. Bakas, S. et al. iglass: imaging integration into the glioma longitudinal analysis consortium. *Neuro-oncology* **22**, 1545–1546 (2020).
10. Rieke, N. et al. The future of digital health with federated learning. *NPJ Digit. Med.* **3**, 1–7 (2020).
11. Sheller, M. J. et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci. Rep. **10**, 1–12 (2020).
12. Annas, G. J. et al. Hipaa regulations—a new era of medical-record privacy? *New Engl. J. Med.* **348**, 1486–1490 (2003).
13. Voigt, P. & Von dem Bussche, A. The EU General Data Protection Regulation (GDPR). In *A Practical Guide* 1st edition, Vol. 10(3152676), 10-5555 (Springer, 2017).
14. McMahan, B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics* (eds Singh, A. & Zhu, J.) 1273–1282 (PMLR, 2017).
15. Sheller, M. J., Reina, G. A., Edwards, B., Martin, J. & Bakas, S.Multi-institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. et al.) 92–104 (Springer, 2018).
16. Dayan, I. et al. Federated learning for predicting clinical outcomes in patients with covid-19. *Nat. Med.* **27**, 1735–1743 (2021).
17. Chang, K. et al. Distributed deep learning networks among institutions for medical imaging. *J. Am. Med. Inform. Assoc.* **25**, 945–954 (2018).
18. Nilsson, A., Smith, S., Ulm, G., Gustavsson, E. & Jirstrand, M. A performance evaluation of federated learning algorithms. In *Proceedings of the Second Workshop on Distributed Infrastructures for Deep Learning*, 1–8 (Association for Computing Machinery, New York, 2018).

19. Sarma, K. V. et al. Federated learning improves site performance in multicenter deep learning without data sharing. *J. Am. Med. Inform. Assoc.* **28**, 1259–1264 (2021).
20. Shen, C. et al. Multi-task federated learning for heterogeneous pancreas segmentation. In *Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning* (eds Laura, C. O. et al.) 101–110 (Springer, 2021).
21. Yang, D. et al. Federated semi-supervised learning for covid region segmentation in chest ct using multi-national data from China, Italy, Japan. *Med. Image Anal.* **70**, 101992 (2021).
22. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).
23. Hannun, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **25**, 65–69 (2019).
24. Griggs, R. C. et al. Clinical research for rare disease: opportunities, challenges, and solutions. *Mol. Genet. Metab.* **96**, 20–26 (2009).
25. Shukla, G. et al. Advanced magnetic resonance imaging in glioblastoma: a review. *Chin. Clin. Oncol.* **6**, 40 (2017).
26. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
27. Verhaak, R. G. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in pdgfra, idh1, egfr, and nf1. *Cancer Cell* **17**, 98–110 (2010).
28. Sottoriva, A. et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl Acad. Sci. USA* **110**, 4009–4014 (2013).
29. Ostrom, Q. T. et al. Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016. *Neuro-oncology* **21**, v1–v100 (2019).
30. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-oncology* **23**, 1231–1251 (2021).
31. Han, W. et al. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *Am. J. Neuroradiol.* **41**, 40–48 (2020).
32. Roth, H. R. et al. Federated learning for breast density classification: a real-world implementation. In *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning* (eds Albarqouni, S. et al.) 181–191 (Springer, 2020).
33. Chaichana, K. L. et al. Multi-institutional validation of a preoperative scoring system which predicts survival for patients with glioblastoma. *J. Clin. Neurosci.* **20**, 1422–1426 (2013).
34. Fathi Kazerooni, A. et al. Cancer imaging phenomics via captk: multi-institutional prediction of progression-free survival and pattern of recurrence in glioblastoma. *JCO Clin. Cancer Inform.* **4**, 234–244 (2020).
35. Menze, B. H. et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2014).
36. Bakas, S. et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. data* **4**, 1–13 (2017).
37. Bakas, S. et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018).
38. Baid, U. et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021).
39. Pati, S. et al. The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research. *Phys Med Biol.* **67**, 204002 (2022).
40. Raghu, M., Zhang, C., Kleinberg, J. & Bengio, S. Transfusion: Understanding transfer learning for medical imaging. *Proceedings of the 33rd International Conference on Neural Information Processing Systems* **32**, 3347–3357 (Association for Computing Machinery, 2019).
41. Young, J. C. & Suryadibrata, A. Applicability of various pre-trained deep convolutional neural networks for pneumonia classification based on x-ray images. *Int. J. Adv. Trends Comput. Sci. Eng.* **9**, 2649–2654 (2020).
42. Stupp, R. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *New Engl. J. Med.* **352**, 987–996 (2005).
43. Beiko, J. et al. Idh1 mutant malignant astrocytomas are more amenable to surgical resection and have a survival benefit associated with maximal surgical resection. *Neuro-oncology* **16**, 81–91 (2014).
44. Olson, J. J. Congress of neurological surgeons systematic review and evidence-based guidelines for the treatment of adults with progressive glioblastoma update: introduction and methods. *J. Neuro-oncol* **158**, 133–137 (2022).
45. Curry, W. T. & Barker, F. G. Racial, ethnic and socioeconomic disparities in the treatment of brain tumors. *J. Neuro-oncol.* **93**, 25–39 (2009).
46. Marsland, S. Novelty detection in learning systems. *Neural Comput. Surv.* **3**, 157–195 (2003).
47. Mazzara, G. P., Velthuizen, R. P., Pearlman, J. L., Greenberg, H. M. & Wagner, H. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *Int. J. Radiat. Oncol.* Biol.* Phys.* **59**, 300–312 (2004).
48. Mitchell, J. R. et al. Deep neural network to locate and segment brain tumors outperformed the expert technicians who created the training data. *J. Med. Imaging* **7**, 055501 (2020).
49. Kamnitsas, K. et al. Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017).
50. Rudie, J. D. et al. Multi-disease segmentation of gliomas and white matter hyperintensities in the brats data using a 3d convolutional neural network. *Front. Comput. Neurosci.* **13**, 84 (2019).
51. Davatzikos, C. et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. *J. Med. Imaging* **5**, 011018 (2018).
52. Yushkevich, P. A. et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116–1128 (2006).
53. Kikinis, R., Pieper, S. D. & Vosburgh, K. G. 3d slicer: a platform for subject-specific image analysis, visualization, and clinical support. In *Intraoperative Imaging and Image-guided Therapy* (ed. Jolesz, F. A.) 277–289 (Springer, 2014).
54. Kofler, F. et al. Brats toolkit: translating brats brain tumor segmentation algorithms into clinical and scientific practice. *Front. Neurosci.* **125**, 125–125 (2020).
55. Kairouz, P. et al. Advances and open problems in federated learning. *Found. Trends* in Mach. Learn.* **14**, 1–210 (2021).
56. Nasr, M., Shokri, R. & Houmansadr, A. Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, 739–753 (IEEE, 2019).
57. Lam, M., Wei, G.-Y., Brooks, D., Reddi, V. J. & Mitzenmacher, M. Gradient disaggregation: breaking privacy in federated learning by reconstructing the user participant matrix. In *International Conference on Machine Learning*, 5959–5968 (PMLR, 2021).
58. Gentry, C. Fully homomorphic encryption using ideal lattices. In *Proc. 41st Annual ACM Symposium on Theory of Computing*, 169–178 (Association for Computing Machinery, New York, 2009).

59. Yao, A. C.Protocols for secure computations. In *23rd Annual Symposium on Foundations of Computer Science* (SFCS 1982), 160–164 (IEEE, 1982).
60. Sabt, M., Achemlal, M. & Bouabdallah, A. Trusted execution environment: what it is, and what it is not. In *2015 IEEE Trustcom/BigDataSE/ISPA* Vol. 1, 57–64 (IEEE, 2015).
61. Schneider, M., Masti, R. J., Shinde, S., Capkun, S. & Perez, R.Sok: Hardware-supported trusted execution environments. arXiv preprint arXiv:2205.12742 (2022).
62. Dwork, C. Differential privacy: a survey of results. In *International Conference on Theory and Applications of Models of Computation* 1–19 (Springer, 2008).
63. Wei, K. et al. Federated learning with differential privacy: algorithms and performance analysis. *IEEE Trans. Inf. Forensics Secur.* **15**, 3454–3469 (2020).
64. Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W. & Tizhoosh, H. R. Federated learning and differential privacy for medical image analysis. *Sci. Rep.* **12**, 1–10 (2022).
65. Tramer, F. & Boneh, D. Slalom: fast, verifiable and private execution of neural networks in trusted hardware. arXiv preprint arXiv:1806.03287 (2018).
66. Kalra, S., Wen, J., Cresswell, J. C., Volkovs, M. & Tizhoosh, H. R. Proxyfl: decentralized federated learning through proxy model sharing. arXiv preprint arXiv:2111.11343 (2021).
67. Lu, M. Y. et al. Federated learning for computational pathology on gigapixel whole slide images. *Med. Image Anal.* **76**, 102298 (2022).
68. Baid, U. et al. Federated learning for the classification of tumor infiltrating lymphocytes. arXiv preprint arXiv:2203.16622 (2022).
69. Linardos, A., Kushibar, K., Walsh, S., Gkontra, P. & Lekadir, K. Federated learning for multi-center imaging diagnostics: a simulation study in cardiovascular disease. *Sci. Rep.* **12**, 1–12 (2022).
70. Rathore, S. et al. Brain cancer imaging phenomics toolkit (brain-captk): an interactive platform for quantitative analysis of glioblastoma. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. et al.) 133–145 (Springer, 2017).
71. Pati, S. et al. The cancer imaging phenomics toolkit (captk): technical overview. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. & Bakas, S.) 380–394 (Springer, 2019).
72. Gilbert, M. R. et al. Rtog 0825: Phase iii double-blind placebo-controlled trial evaluating bevacizumab (bev) in patients (pts) with newly diagnosed glioblastoma (gbm). *J. Clin. Oncol.* **31**(18_suppl18), 1–1 (2013).
73. Gilbert, M. R. et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *New Engl. J. Med.* **370**, 699–708 (2014).
74. Boxerman, J. L. et al. Prognostic value of contrast enhancement and flair for survival in newly diagnosed glioblastoma treated with and without bevacizumab: results from acrin 6686. *Neuro-oncology* **20**, 1400–1410 (2018).
75. Schmainda, K. M. et al. Value of dynamic contrast perfusion mri to predict early response to bevacizumab in newly diagnosed glioblastoma: results from acrin 6686 multicenter trial. *Neuro-oncology* **23**, 314–323 (2021).
76. Pianykh, O. S. *Digital Imaging and Communications in Medicine (DICOM): a Practical Introduction and Survival Guide* (Springer, 2012).
77. Kahn, C. E., Carrino, J. A., Flynn, M. J., Peck, D. J. & Horii, S. C. Dicom and radiology: past, present, and future. *J. Am. College Radiol.* **4**, 652–657 (2007).
78. Mustra, M., Delac, K. & Grgic, M. Overview of the dicom standard. In *2008 50th International Symposium ELMAR* Vol. 1, 39–44 (IEEE, 2008).
79. Cox, R. et al. A (sort of) new image data format standard: Nifti-1. In: *Proc. 10th Annual Meeting of the Organization for Human Brain Mapping* 22 (Wiley, 2004).
80. Li, X., Morgan, P. S., Ashburner, J., Smith, J. & Rorden, C. The first step for neuroimaging data analysis: Dicom to nifti conversion. *J. Neurosci. Methods* **264**, 47–56 (2016).
81. White, T., Blok, E. & Calhoun, V. D. Data sharing and privacy issues in neuroimaging research: opportunities, obstacles, challenges, and monsters under the bed. *Hum. Brain Mapp* **43**, 278–291 (2020).
82. Rohlfing, T., Zahr, N. M., Sullivan, E. V. & Pfefferbaum, A. The sri24 multichannel atlas of normal adult human brain structure. *Hum. Brain Mapp.* **31**, 798–819 (2010).
83. Song, S., Zheng, Y. & He, Y. A review of methods for bias correction in medical images. *Biomed. Eng. Rev.* **1**, 2375–9151 (2017).
84. Sled, J. G., Zijdenbos, A. P. & Evans, A. C. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans. Med. Imaging* **17**, 87–97 (1998).
85. Tustison, N. J. et al. N4itk: improved n3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
86. Schwarz, C. G. et al. Identification of anonymous mri research participants with face-recognition software. *New Engl. J. Med.* **381**, 1684–1686 (2019).
87. Thakur, S. et al. Brain extraction on MRI scans in presence of diffuse glioma: Multi-institutional performance evaluation of deep learning methods and robust modality-agnostic training. *NeuroImage* **220**, 117081 (2020).
88. McKinley, R., Meier, R. & Wiest, R. Ensembles of densely-connected cnns with label-uncertainty for brain tumor segmentation. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. et al.) 456–465 (Springer, 2018).
89. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
90. Rohlfing, T., Russakoff, D. B. & Maurer, C. R. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans. Med. Imaging* **23**, 983–994 (2004).
91. Warfield, S. K., Zou, K. H. & Wells, W. M. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* **23**, 903–921 (2004).
92. Rohlfing, T. & Maurer Jr, C. R. Multi-classifier framework for atlas-based image segmentation. *Pattern Recognit. Lett.* **26**, 2070–2079 (2005).
93. Huo, J., Wang, G., Wu, Q. J. & Thangarajah, A. Label fusion for multi-atlas segmentation based on majority voting. In *International Conference Image Analysis and Recognition* (eds Kamel, M. & Campilho, A.) 100–106 (Springer, 2015).
94. Langerak, T. R. et al. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple). *IEEE Trans. Med. Imaging* **29**, 2000–2008 (2010).
95. Pati, S. et al. Gandlf: a generally nuanced deep learning framework for scalable end-to-end clinical workflows in medical imaging. arXiv preprint arXiv:2103.01006 (2021).
96. Reinhold, J. C., Dewey, B. E., Carass, A. & Prince, J. L. Evaluating the impact of intensity normalization on MR image synthesis. In *Medical Imaging 2019: Image Processing*, Vol. 10949 (eds Angelini, E. D. & Landman, B. A.) 109493H (International Society for Optics and Photonics, 2019).
97. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention* (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A.) 234–241 (Springer, 2015).

98. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-assisted Intervention* (eds Ourselin, S. et al.) 424–432 (Springer, 2016).
99. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (IEEE, 2016).
100. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S. & Pal, C. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications* (eds Carneiro, G. et al.) 179–187 (Springer, 2016).
101. Bhalerao, M. & Thakur, S.Brain tumor segmentation based on 3d residual u-net. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. & Bakas, S.) 218–225 (Springer, 2019).
102. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
103. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. & Cardoso, M. J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (eds Cardoso, M. J. et al.) 240–248 (Springer, 2017).
104. Zijdenbos, A. P., Dawant, B. M., Margolin, R. A. & Palmer, A. C. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans. Med. Imaging* **13**, 716–724 (1994).
105. Chen, L., Qu, H., Zhao, J., Chen, B. & Principe, J. C. Efficient and robust deep learning with correntropy-induced loss function. *Neural Comput. Appl.* **27**, 1019–1031 (2016).
106. Salehi, S. S. M., Erdogmus, D. & Gholipour, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International Workshop on Machine Learning in Medical Imaging* (eds Wang, Q., Shi, Y., Suk, H. & Suzuki, K.) 379–387 (Springer, 2017).
107. Caliva, F., Iriondo, C., Martinez, A. M., Majumdar, S. & Pedoia, V.Distance map loss penalty term for semantic segmentation. arXiv preprint arXiv:1908.03679 (2019).
108. Shamir, R. R., Duchin, Y., Kim, J., Sapiro, G. & Harel, N.Continuous dice coefficient: a method for evaluating probabilistic segmentations. arXiv preprint arXiv:1906.11031 (2019).
109. Horn, R. A. The hadamard product. In *Proc. Symposium on Applied Mathematics*, Vol. 40 (eds Berghel, H. & Talbert, J.) 87–169 (American Mathematical Society, 1990).
110. Barrodale, I. L1 approximation and the analysis of data. *J. R. Stat. Soc.: Ser. C (Appl. Stat.)* **17**, 51–57 (1968).
111. Knauth, T. et al. Integrating remote attestation with transport layer security. arXiv preprint arXiv:1801.05863 (2018).
112. Kaissis, G. A., Makowski, M. R., Rückert, D. & Braren, R. F. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat. Mach. Intell.* **2**, 305–311 (2020).
113. Ekberg, J.-E., Kostiainen, K. & Asokan, N. The untapped potential of trusted execution environments on mobile devices. *IEEE Secur. Priv.* **12**, 29–37 (2014).
114. Rodriguez, A. et al. Lower numerical precision deep learning inference and training. *Intel White Paper* **3**, 1–19 (2018).
115. Lin, D., Talathi, S. & Annapureddy, S. Fixed point quantization of deep convolutional networks. In *International Conference on Machine Learning* (eds Balcan, M. F. & Weinberger, K. Q.) 2849–2858 (PMLR, 2016).
116. Vakili, S., Langlois, J. P. & Bois, G. Enhanced precision analysis for accuracy-aware bit-width optimization using affine arithmetic. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.* **32**, 1853–1865 (2013).
117. Wilkinson, M. D. et al. The fair guiding principles for scientific data management and stewardship. *Sci. data* **3**, 1–9 (2016).
118. Pati, S. & Bakas, S. S. Fets-ai/front-end: release for zenodo https://doi.org/10.5281/zenodo.7036038 (2022)
119. Yushkevich, P. A. et al. Fast automatic segmentation of hippocampal subfields and medial temporal lobe subregions in 3 tesla and 7 tesla t2-weighted MRI. *Alzheimer's Dement.* **12**, P126–P127 (2016).
120. Pati, S. & Bakas, S. LabelFusion: medical image label fusion of segmentations https://doi.org/10.5281/zenodo.4633206 (2021)
121. Paszke, A. et al. Pytorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* vol. 32(eds Wallach, H. M. et al.) 8026–8037 (Neural Information Processing Systems Foundation, Inc., 2019).
122. Pérez-García, F., Sparks, R. & Ourselin, S. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput. Methods Programs Biomed.* **208**, 106236 (2021).
123. Foley, P. et al. OpenFL: the open federated learning library. *Physics in Medicine & Biology* (2022). Online ahead of print.
124. Gorbachev, Y. et al. Openvino deep learning workbench: Comprehensive analysis and tuning of neural networks inference. In *Proc. IEEE/CVF International Conference on Computer Vision Workshops*, 783–787 (IEEE, 2019).

## Acknowledgements

Research and main methodological developments reported in this publication were partly supported by the National Institutes of Health (NIH) under award numbers NIH/NCI:U01CA242871 (S. Bakas), NIH/NINDS:R01NS042645 (C. Davatzikos), NIH/NCI:U24CA189523 (C. Davatzikos), NIH/NCI:U24CA215109 (J. Saltz), NIH/NCI:U01CA248226 (P. Tiwari), NIH/NCI:P30CA51008 (Y. Gusev), NIH:R50CA211270 (M. Muzi), NIH/NCATS:UL1TR001433 (Y. Yuan), NIH/NIBIB:R21EB030209 (Y. Yuan), NIH/NCI:R37CA214955 (A. Rao), and NIH:R01CA233888 (A.L. Simpson). The authors would also like to acknowledge the following NIH funded awards for the multi-site clinical trial (NCT00884741, RTOG0825/ACRIN6686): U10CA21661, U10CA37422, U10CA180820, U10CA180794, U01CA176110, R01CA062500, CA079778, CA080098, CA180794, CA180820, CA180822, CA180868. Research reported in this publication was also partly supported by the National Science Foundation, under award numbers 2040532 (S. Baek), and 2040462 (B. Landman). Research reported in this publication was also supported by i) a research grant from Varian Medical Systems (Palo Alto, CA, USA) (Y.Yuan), (ii) the Ministry of Health of the Czech Republic (Grant Nr. NU21-08-00359) (M.Kerkovský and M.Kozubek), (iii) Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 404521405, SFB 1389, Work Package C02, and Priority Program 2177 "Radiomics: Next Generation of Biomedical Imaging" (KI 2410/1-1 | MA 6340/18-1) (P. Vollmuth), (iv) DFG Project-ID B12, SFB 824 (B. Wiestler), (v) the Helmholtz Association (funding number ZT-I-OO14) (K. Maier-Hein), vi) the Dutch Cancer Society (KWF project number EMCR 2015-7859) (S.R. van der Voort), (vii) the Chilean National Agency for Research and Development (ANID-Basal FB0008 (AC3E) and FB210017 (CENIA)) (P. Guevara), viii) the Canada CIFAR AI Chairs Program (M. Vallières), (ix) Leeds Hospital Charity (Ref: 9R01/1403) (S. Currie), (x) the Cancer Research UK funding for the Leeds Radiotherapy Research Centre of Excellence (RadNet) and the grant number C19942/A28832 (S. Currie), (xi) Medical Research Council (MRC) Doctoral Training Program in Precision Medicine (Award Reference No. 2096671) (J. Bernal), (xii) The European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 757173) (B.Glocker), (xiii) The UKRI London Medical Imaging & Artificial Intelligence Centre for Value-Based Healthcare (K. Kamnitsas), (xiv) Wellcome/Engineering and Physical Sciences Research Council (EPSRC) Center for Medical Engineering (WT 203148/Z/16/Z) (T.C. Booth), (xv) American Cancer Society Research Scholar Grant RSG-16-005-01 (A. Rao), (xvi) the Department of Defense (DOD)

Peer Reviewed Cancer Research Program (PRCRP) W81XWH-18-1-0404, Dana Foundation David Mahoney Neuroimaging Program, the V Foundation Translational Research Award, Johnson & Johnson WIS-TEM2D Award (P. Tiwari), (xvii) RSNA Research & Education Foundation under grant number RR2011 (E.Calabrese), (xviii) the National Research Fund of Luxembourg (FNR) (grant number: C20/BM/14646004/GLASS-LUX/Niclou) (S.P.Niclou), xix) EU Marie Curie FP7-PEOPLE-2012-ITN project TRANSACT (PITN-GA-2012-316679) and the Swiss National Science Foundation (project number 140958) (J. Slotboom), and (xx) CNPq 303808/2018-7 and FAPESP 2014/12236-1 (A. Xavier Falcão). The content of this publication is solely the responsibility of the authors and does not represent the official views of the NIH, the NSF, the RSNA R&E Foundation, or any of the additional funding bodies.

## Author contributions

Study conception: S. Pati, U. Baid, B. Edwards, M. Sheller, G.A. Reina, J. Martin, S. Bakas. Development of software used in the study: S. Pati, B. Edwards, M. Sheller, S. Wang, G.A. Reina, P. Foley, A. Gruzdev, D. Karkada, S. Bakas. Data acquisition: M. Bilello, S. Mohan, E. Calabrese, J. Rudie, J. Saini, R.Y. Huang, K. Chang, T. So, P. Heng, T.F. Cloughesy, C. Raymond, T. Oughourlian, A. Hagiwara, C. Wang, M. To, M. Kerkovský, T. Koprivová, M. Dostál, V. Vybíhal, J.A. Maldjian, M.C. Pinho, D. Reddy, J. Holcomb, B. Wiestler, M. Metz, R. Jain, M. Lee, P. Tiwari, R. Verma, Y. Gusev, K. Bhuvaneshwar, C. Bencheqroun, A. Belousli, A. Abayazeed, A. Abbassy, S. Gamal, M. Qayati, M. Mekhaimar, M. Reyes, R.R. Colen, M. Ak, P. Vollmuth, G. Brugnara, F. Sahm, M. Bendszus, W. Wick, A. Mahajan, C. Balaña, J. Capellades, J. Puig, Y. Choi, M. Muzi, H.F. Shaykh, A. Herrera-Trujillo, W. Escobar, A. Abello, P. LaMontagne, B. Landman, K. Ramadass, K. Xu, S. Chotai, L.B. Chambless, A. Mistry, R.C. Thompson, J. Bapuraj, N. Wang, S.R. van der Voort, F. Incekara, M.M.J. Wijnenga, R. Gahrmann, J.W. Schouten, H.J. Dubbink, A.J.P.E. Vincent, M.J. van den Bent, H.J. Sair, C.K. Jones, A. Venkataraman, J. Garrett, M. Larson, B. Menze, T. Weiss, M. Weller, A. Bink, Y. Yuan, S. Sharma, T. Tseng, B.C.A. Teixeira, F. Spranger, S.P. Niclou, O. Keunen, L.V.M. Dixon, M. Williams, R.G.H. Beets-Tan, H. Franco-Maldonado, F. Loayza, J. Slotboom, P. Radojewski, R. Meier, R. Wiest, J. Trenkler, J. Pichler, G. Necker, S. Meckel, E. Torche, F. Vera, E. Lóópez, Y. Kim, H. Ismael, B. Allen, J.M. Buatti, J. Park, P. Zampakis, V. Panagiotopoulos, P. Tsiganos, E. Challiasos, D.M. Kardamakis, P. Prasanna, K.M. Mani, D. Payne, T. Kurc, L. Poisson, M. Vallières, D. Fortin, M. Lepage, F. Morón, J. Mandel, C. Badve, A.E. Sloan, J.S. Barnholtz-Sloan, K. Waite, G. Shukla, S. Liem, G.S. Alexandre, J. Lombardo, J.D. Palmer, A.E. Flanders, A.P. Dicker, G. Ogbole, D. Oyekunle, O. Odafe-Oyibotha, B. Osobu, M. Shu'aibu, F. Dako, A. Dorcas, D. Murcia, R. Haas, J. Thompson, D.R. Ormond, S. Currie, K. Fatania, R. Frood, J. Mitchell, J. Farinhas, A.L. Simpson, J.J. Peoples, R. Hu, D. Cutler, F.Y. Moraes, A. Tran, M. Hamghalam, M.A. Boss, J. Gimpel, B. Bialecki, A. Chelliah. Data processing: C. Sako, S. Ghodasara, E. Calabrese, J. Rudie, M. Jadhav, U. Pandey, R.Y. Huang, M. Jiang, C. Chen, C. Raymond, S. Bhardwaj, C. Chong, M. Agzarian, M. Kozubek, F. Lux, J. Michálek, P. Matula, C. Bangalore Yogananda, D. Reddy, B.C. Wagner, I. Ezhov, M. Lee, Y.W. Lui, R. Verma, R. Bareja, I. Yadav, J. Chen, N. Kumar, K. Bhuvaneshwar, A. Sayah, C. Bencheqroun, K. Kolodziej, M. Hill, M. Reyes, L. Pei, M. Ak, A. Kotrotsou, P. Vollmuth, G. Brugnara, C.J. Preetha, M. Zenk, J. Puig, M. Muzi, H.F. Shaykh, A. Abello, J. Bernal, J. Gómez, P. LaMontagne, K. Ramadass, S. Chotai, N. Wang, M. Smits, S.R. van der Voort, A. Alafandi, F. Incekara, M.M.J. Wijnenga, G. Kapsas, R. Gahrmann, A.J.P.E. Vincent, P.J. French, S. Klein, H.J. Sair, C.K. Jones, J. Garrett, H. Li, F. Kofler, Y. Yuan, S. Adabi, A. Xavier Falcão, S.B. Martins, D. Menotti, D.R. Lucio, O. Keunen, A. Hau, K. Kamnitsas, L. Dixon, S. Benson, E. Pelaez, H. Franco-Maldonado, F. Loayza, S. Quevedo, R. McKinley, J. Trenkler, A. Haunschmidt, C. Mendoza, E. Ríos, J. Choi, S. Baek, J. Yun, P. Zampakis, V. Panagiotopoulos, P. Tsiganos, E.I. Zacharaki, C. Kalogeropoulou, P. Prasanna, S. Shreshtra, T. Kurc, B. Luo, N. Wen, M. Vallières, D. Fortin, F. Morón, C. Badve, V. Vadmal, G. Shukla, G. Ogbole, D. Oyekunle, F. Dako, D. Murcia, E. Fu, S. Currie, R. Frood, M.A. Vogelbaum, J. Mitchell, J. Farinhas, J.J. Peoples, M. Hamghalam, D. Kattil Veettil, K. Schmidt, B. Bialecki, S. Marella, T.C. Booth, A. Chelliah, M. Modat, C. Dragos, H. Shuaib. Data analysis & interpretation: S. Pati, U. Baid, B. Edwards, M. Sheller, S. Bakas. Site PI/Senior member (of each collaborating group): C. Davatzikos, J. Villanueva-Meyer, M. Ingalhalikar, R.Y. Huang, Q. Dou, B.M. Ellingson, M. To, M. Kozubek, J.A. Maldjian, B. Wiestler, R. Jain, P. Tiwari, Y. Gusev, A. Abayazeed, R.R. Colen, P. Vollmuth, A. Mahajan, C. Balaña, S. Lee, M. Muzi, H.F. Shaykh, M. Trujillo, D. Marcus, B. Landman, A. Rao, M. Smits, H.J. Sair, R. Jeraj, B. Menze, Y. Yuan, A. Xavier Falcão, S.P. Niclou, B. Glocker, J. Teuwen, E. Pelaez, R. Wiest, S. Meckel, P. Guevara, S. Baek, H. Kim, D.M. Kardamakis, J. Saltz, L. Poisson, M. Vallières, F. Morón, A.E. Sloan, A.E. Flanders, G. Ogbole, D.R. Ormond, S. Currie, J. Farinhas, A.L. Simpson, C. Apgar, T.C. Booth.Writing the original manuscript: S. Pati, U. Baid, B. Edwards, M. Sheller, S. Bakas. Review, edit, & approval of the final manuscript: All authors.

## Competing interests

The Intel-affiliated authors (B. Edwards, M. Sheller, S. Wang, G.A. Reina, P. Foley, A. Gruzdev, D. Karkada, P. Shah, J. Martin) would like to disclose the following (potential) competing interests as Intel employees. Intel may develop proprietary software that is related in reputation to the OpenFL open source project highlighted in this work. In addition, the work demonstrates feasibility of federated learning for brain tumor boundary detection models. Intel may benefit by selling products to support an increase in demand for this use-case. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-33407-5.

**Correspondence** and requests for materials should be addressed to Spyridon Bakas.

Sarthak Pati[1,2,3,4,154], Ujjwal Baid[1,2,3,154], Brandon Edwards[5,154], Micah Sheller[5], Shih-Han Wang[5], G. Anthony Reina[5], Patrick Foley[5], Alexey Gruzdev[5], Deepthi Karkada[5], Christos Davatzikos[1,2], Chiharu Sako[1,2], Satyam Ghodasara[2], Michel Bilello[1,2], Suyash Mohan[1,2], Philipp Vollmuth[6], Gianluca Brugnara[6], Chandrakanth J. Preetha[6], Felix Sahm[7,8], Klaus Maier-Hein[9,10], Maximilian Zenk[9], Martin Bendszus[6], Wolfgang Wick[7,11], Evan Calabrese[12], Jeffrey Rudie[12], Javier Villanueva-Meyer[12], Soonmee Cha[12], Madhura Ingalhalikar[13], Manali Jadhav[13], Umang Pandey[13], Jitender Saini[14], John Garrett[15,16], Matthew Larson[15], Robert Jeraj[15,16], Stuart Currie[17], Russell Frood[17], Kavi Fatania[17], Raymond Y. Huang[18], Ken Chang[19], Carmen Balaña[20], Jaume Capellades[21], Josep Puig[22], Johannes Trenkler[23], Josef Pichler[24], Georg Necker[23], Andreas Haunschmidt[23], Stephan Meckel[23,25], Gaurav Shukla[1,26], Spencer Liem[27], Gregory S. Alexander[28], Joseph Lombardo[27,29], Joshua D. Palmer[30], Adam E. Flanders[31], Adam P. Dicker[29], Haris I. Sair[32,33], Craig K. Jones[33], Archana Venkataraman[34], Meirui Jiang[35], Tiffany Y. So[35], Cheng Chen[35], Pheng Ann Heng[35], Qi Dou[35], Michal Kozubek[36], Filip Lux[36], Jan Michálek[36], Petr Matula[36], Miloš Keřkovský[37], Tereza Kopřivová[37], Marek Dostál[37,38], Václav Vybíhal[39], Michael A. Vogelbaum[40], J. Ross Mitchell[41,42], Joaquim Farinhas[43], Joseph A. Maldjian[44], Chandan Ganesh Bangalore Yogananda[44], Marco C. Pinho[44], Divya Reddy[44], James Holcomb[44], Benjamin C. Wagner[44], Benjamin M. Ellingson[45,46], Timothy F. Cloughesy[46], Catalina Raymond[45], Talia Oughourlian[45,47], Akifumi Hagiwara[47], Chencai Wang[47], Minh-Son To[48,49], Sargam Bhardwaj[48], Chee Chong[50], Marc Agzarian[50,51], Alexandre Xavier Falcão[52], Samuel B. Martins[53], Bernardo C. A. Teixeira[54,55], Flávia Sprenger[55], David Menotti[56], Diego R. Lucio[56], Pamela LaMontagne[57], Daniel Marcus[57], Benedikt Wiestler[58,59], Florian Kofler[58,59,60], Ivan Ezhov[4,59,60], Marie Metz[58], Rajan Jain[61,62], Matthew Lee[61], Yvonne W. Lui[61], Richard McKinley[63], Johannes Slotboom[63], Piotr Radojewski[63], Raphael Meier[63], Roland Wiest[63], Derrick Murcia[64], Eric Fu[64], Rourke Haas[64], John Thompson[64], David Ryan Ormond[64], Chaitra Badve[65], Andrew E. Sloan[66,67,68], Vachan Vadmal[68], Kristin Waite[69], Rivka R. Colen[70,71], Linmin Pei[72], Murat Ak[70], Ashok Srinivasan[73], J. Rajiv Bapuraj[73], Arvind Rao[74], Nicholas Wang[74], Ota Yoshiaki[73], Toshio Moritani[73], Sevcan Turk[73], Joonsang Lee[74], Snehal Prabhudesai[74], Fanny Morón[75], Jacob Mandel[51], Konstantinos Kamnitsas[76,77], Ben Glocker[76], Luke V. M. Dixon[78], Matthew Williams[79], Peter Zampakis[80], Vasileios Panagiotopoulos[81], Panagiotis Tsiganos[82], Sotiris Alexiou[83], Ilias Haliassos[84], Evangelia I. Zacharaki[83], Konstantinos Moustakas[83], Christina Kalogeropoulou[80], Dimitrios M. Kardamakis[85], Yoon Seong Choi[86], Seung-Koo Lee[86], Jong Hee Chang[86], Sung Soo Ahn[86], Bing Luo[87], Laila Poisson[88], Ning Wen[87,89], Pallavi Tiwari[90], Ruchika Verma[42,90], Rohan Bareja[90], Ipsa Yadav[90], Jonathan Chen[90], Neeraj Kumar[41,42], Marion Smits[91], Sebastian R. van der Voort[91], Ahmed Alafandi[91], Fatih Incekara[91,92], Maarten M. J. Wijnenga[93], Georgios Kapsas[91], Renske Gahrmann[91], Joost W. Schouten[92], Hendrikus J. Dubbink[94], Arnaud J. P. E. Vincent[92], Martin J. van den Bent[93], Pim J. French[93], Stefan Klein[95], Yading Yuan[96], Sonam Sharma[96], Tzu-Chi Tseng[96], Saba Adabi[96], Simone P. Niclou[97], Olivier Keunen[98], Ann-Christin Hau[97,99], Martin Vallières[100,101], David Fortin[101,102], Martin Lepage[101,103], Bennett Landman[104], Karthik Ramadass[104], Kaiwen Xu[105], Silky Chotai[106], Lola B. Chambless[106], Akshitkumar Mistry[106], Reid C. Thompson[106], Yuriy Gusev[107], Krithika Bhuvaneshwar[107], Anousheh Sayah[108], Camelia Bencheqroun[107], Anas Belouali[107], Subha Madhavan[107], Thomas C. Booth[109,110], Alysha Chelliah[109], Marc Modat[109], Haris Shuaib[111,112], Carmen Dragos[111], Aly Abayazeed[113], Kenneth Kolodziej[113], Michael Hill[113], Ahmed Abbassy[114], Shady Gamal[114], Mahmoud Mekhaimar[114], Mohamed Qayati[114], Mauricio Reyes[115], Ji Eun Park[116], Jihye Yun[116], Ho Sung Kim[116], Abhishek Mahajan[117], Mark Muzi[118], Sean Benson[119], Regina G. H. Beets-Tan[120,121], Jonas Teuwen[119], Alejandro Herrera-Trujillo[122,123], Maria Trujillo[123], William Escobar[122,123], Ana Abello[123], Jose Bernal[123,124], Jhon Gómez[123], Joseph Choi[125], Stephen Baek[126], Yusung Kim[127], Heba Ismael[127], Bryan Allen[127], John M. Buatti[127], Aikaterini Kotrotsou[128], Hongwei Li[129], Tobias Weiss[130], Michael Weller[130], Andrea Bink[131], Bertrand Pouymayou[131], Hassan F. Shaykh[132], Joel Saltz[133], Prateek Prasanna[133], Sampurna Shrestha[133], Kartik M. Mani[133,134], David Payne[135], Tahsin Kurc[133,136], Enrique Pelaez[137], Heydy Franco-Maldonado[138], Francis Loayza[137], Sebastian Quevedo[139], Pamela Guevara[140], Esteban Torche[140], Cristobal Mendoza[140], Franco Vera[140], Elvis Rios[140], Eduardo López[140], Sergio A. Velastin[141], Godwin Ogbole[142], Mayowa Soneye[142], Dotun Oyekunle[142], Olubunmi Odafe-Oyibotha[143], Babatunde Osobu[142], Mustapha Shu'aibu[144], Adeleye Dorcas[145], Farouk Dako[2,146], Amber L. Simpson[112,147], Mohammad Hamghalam[147,148], Jacob J. Peoples[147], Ricky Hu[147], Anh Tran[147], Danielle Cutler[149], Fabio Y. Moraes[150], Michael A. Boss[151], James Gimpel[151], Deepak Kattil Veettil[151], Kendall Schmidt[152], Brian Bialecki[152], Sailaja Marella[151], Cynthia Price[151], Lisa Cimino[151], Charles Apgar[151], Prashant Shah[5], Bjoern Menze[4,129], Jill S. Barnholtz-Sloan[69,153], Jason Martin[5] & Spyridon Bakas[1,2,3]

[1]Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, PA, USA. [2]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [4]Department of Informatics, Technical University of Munich, Munich, Bavaria, Germany. [5]Intel Corporation, Santa Clara, CA, USA. [6]Department of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany. [7]Clinical Cooperation Unit Neuropathology, German Cancer Consortium (DKTK) within the German Cancer Research Center (DKFZ), Heidelberg, Germany. [8]Department of Neuropathology, Heidelberg University Hospital, Heidelberg, Germany. [9]Division of Medical Image Computing, German Cancer Research Center, Heidelberg, Germany. [10]Pattern Analysis and Learning Group, Department of Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany. [11]Neurology Clinic, Heidelberg University Hospital, Heidelberg, Germany. [12]Department of Radiology & Biomedical Imaging, University of California San Francisco, San Francisco, CA, USA. [13]Symbiosis Center for Medical Image Analysis, Symbiosis International University, Pune, Maharashtra, India. [14]Department of Neuroimaging and Interventional Radiology, National Institute of Mental Health and Neurosciences, Bangalore, Karnataka, India. [15]Department of Radiology, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA. [16]Department of Medical Physics, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA. [17]Leeds Teaching Hospitals Trust, Department of Radiology, Leeds, UK. [18]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [19]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA. [20]Catalan Institute of Oncology, Badalona, Spain. [21]Consorci MAR Parc de Salut de Barcelona, Catalonia, Spain. [22]Department of Radiology (IDI), Girona Biomedical Research Institute (IdIBGi), Josep Trueta University Hospital, Girona, Spain. [23]Institute of Neuroradiology, Neuromed Campus (NMC), Kepler University Hospital Linz, Linz, Austria. [24]Department of Neurooncology, Neuromed Campus (NMC), Kepler University Hospital Linz, Linz, Austria. [25]Institute of Diagnostic and Interventional Neuroradiology, RKH Klinikum Ludwigsburg, Ludwigsburg, Germany. [26]Department of Radiation Oncology, Christiana Care Health System, Philadelphia, PA, USA. [27]Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA. [28]Department of Radiation Oncology, University of Maryland, Baltimore, MD, USA. [29]Department of Radiation Oncology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA. [30]Department of Radiation Oncology, The James Cancer Hospital and Solove Research Institute, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. [31]Department of Radiology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA. [32]The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [33]The Malone Center for Engineering in Healthcare, The Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA. [34]Department of Electrical and Computer Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA. [35]The Chinese University of Hong Kong, Hong Kong, China. [36]Centre for Biomedical Image Analysis, Faculty of Informatics, Masaryk University, Brno, Czech Republic. [37]Department of Radiology and Nuclear Medicine, Faculty of Medicine, Masaryk University, Brno and University Hospital Brno, Brno, Czech Republic. [38]Department of Biophysics, Faculty of Medicine, Masaryk University, Brno, Czech Republic. [39]Department of Neurosurgery, Faculty of Medicine, Masaryk University, Brno, and University Hospital and Czech Republic, Brno, Czech Republic. [40]Department of Neuro Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [41]University of Alberta, Edmonton, AB, Canada. [42]Alberta Machine Intelligence Institute, Edmonton, AB, Canada. [43]Department of Radiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [44]University of Texas Southwestern Medical Center, Dallas, TX, USA. [45]UCLA Brain Tumor Imaging Laboratory (BTIL), Center for Computer Vision and Imaging Biomarkers, Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [46]UCLA Neuro-Oncology Program, Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CaA, USA. [47]Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [48]College of Medicine and Public Health, Flinders University, Bedford Park, SA, Australia. [49]Division of Surgery and Perioperative Medicine, Flinders Medical Centre, Bedford Park, SA, Australia. [50]South Australia Medical Imaging, Flinders Medical Centre, Bedford Park, SA, Australia. [51]Department of Neurology, Baylor College of Medicine, Houston, TX, USA. [52]Institute of Computing, University of Campinas, Campinas, São Paulo, Brazil. [53]Federal Institute of São Paulo, Campinas, São Paulo, Brazil. [54]Instituto de Neurologia de Curitiba, Curitiba, Paraná, Brazil. [55]Department of Radiology, Hospital de Clínicas da Universidade Federal do Paraná, Curitiba, Paraná, Brazil. [56]Department of Informatics, Universidade Federal do Paraná, Curitiba, Paraná, Brazil. [57]Department of Radiology, Washington University in St. Louis, St. Louis, MO, USA. [58]Department of Diagnostic and Interventional Neuroradiology, School of Medicine, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany. [59]TranslaTUM (Zentralinstitut für translationale Krebsforschung der Technischen Universität München), Klinikum rechts der Isar, Munich, Germany. [60]Image-Based Biomedical Modeling, Department of Informatics, Technical University of Munich, Munich, Germany. [61]Department of Radiology, NYU Grossman School of Medicine, New York, NY, USA. [62]Department of Neurosurgery, NYU Grossman School of Medicine, New York, NY, USA. [63]Support Center for Advanced Neuroimaging, University Institute of Diagnostic and Interventional Neuroradiology, University Hospital Bern, Inselspital, University of Bern, Bern, Switzerland. [64]Department of Neurosurgery, Anschutz Medical Campus, University of Colorado, Aurora, CO, USA. [65]Department of Radiology, University Hospitals Cleveland, Cleveland, OH, USA. [66]Department of Neurological Surgery, University Hospitals-Seidman Cancer Center, Cleveland, OH, USA. [67]Case Comprehensive Cancer Center, Cleveland, OH, USA. [68]Department of Neurosurgery, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [69]National Cancer Institute, National Institute of Health, Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA. [70]Department of Radiology, Neuroradiology Division, University of Pittsburgh, Pittsburgh, PA, USA. [71]Department of Diagnostic Radiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA. [72]University of Pittsburgh Medical Center, Pittsburgh, PA, USA. [73]Department of Neuroradiology, University of Michigan, Ann Arbor, MI, USA. [74]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [75]Department of Radiology, Baylor College of Medicine, Houston, TX, USA. [76]Department of Computing, Imperial College London, London, UK. [77]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK. [78]Department of Radiology, Imperial College NHS Healthcare Trust, London, UK. [79]Computational Oncology Group, Institute for Global Health Innovation, Imperial College London, London, UK. [80]Department of NeuroRadiology, University of Patras, Patras, Greece. [81]Department of Neurosurgery, University of Patras, Patras, Greece. [82]Clinical Radiology Laboratory, Department of Medicine, University of Patras, Patras, Greece. [83]Department of Electrical and Computer Engineering, University of Patras, Patras, Greece. [84]Department of Neuro-Oncology, University of Patras, Patras, Greece. [85]Department of Radiation Oncology, University of Patras, Patras, Greece. [86]Yonsei University College of Medicine, Seoul, Korea. [87]Department of Radiation Oncology, Henry Ford Health System, Detroit, MI, USA. [88]Public Health Sciences, Henry Ford Health System, Detroit, MI, USA. [89]SJTU-Ruijin-UIH Institute for Medical Imaging Technology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, 200025 Shanghai, China. [90]Case Western Reserve University, Cleveland, OH, USA. [91]Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Centre Rotterdam, Rotterdam, Netherlands. [92]Department of Neurosurgery, Brain Tumor Center, Erasmus MC University Medical Centre Rotterdam, Rotterdam, Netherlands. [93]Department of Neurology, Brain Tumor Center, Erasmus MC Cancer Institute, Rotterdam, Netherlands. [94]Department of Pathology, Brain Tumor Center, Erasmus MC Cancer Institute, Rotterdam, Netherlands. [95]Biomedical Imaging Group Rotterdam, Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Centre Rotterdam, Rotterdam, Netherlands. [96]Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [97]NORLUX Neuro-

Oncology Laboratory, Department of Cancer Research, Luxembourg Institute of Health, Luxembourg, Luxembourg. [98]Translation Radiomics, Department of Cancer Research, Luxembourg Institute of Health, Luxembourg, Luxembourg. [99]Luxembourg Center of Neuropathology, Laboratoire National De Santé, Luxembourg, Luxembourg. [100]Department of Computer Science, Université de Sherbrooke, Sherbrooke, QC, Canada. [101]Centre de Recherche du Centre Hospitalière Universitaire de Sherbrooke, Sherbrooke, QC, Canada. [102]Division of Neurosurgery and Neuro-Oncology, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada. [103]Department of Nuclear Medicine and Radiobiology, Sherbrooke Molecular Imaging Centre, Université de Sherbrooke, Sherbrooke, QC, Canada. [104]Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA. [105]Department of Computer Science, Vanderbilt University, Nashville, TN, USA. [106]Department of Neurosurgery, Vanderbilt University Medical Center, Nashville, TN, USA. [107]Innovation Center for Biomedical Informatics (ICBI), Georgetown University, Washington, DC, USA. [108]Division of Neuroradiology & Neurointerventional Radiology, Department of Radiology, MedStar Georgetown University Hospital, Washington, DC, USA. [109]School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK. [110]Department of Neuroradiology, Ruskin Wing, King's College Hospital NHS Foundation Trust, London, UK. [111]Stoke Mandeville Hospital, Mandeville Road, Aylesbury, UK. [112]Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada. [113]Neosoma Inc., Groton, MA, USA. [114]University of Cairo School of Medicine, Giza, Egypt. [115]University of Bern, Bern, Switzerland. [116]Department of Radiology, Asan Medical Center, Seoul, South Korea. [117]The Clatterbridge Cancer Centre NHS Foundation Trust Pembroke Place, Liverpool, UK. [118]Department of Radiology, University of Washington, Seattle, WA, USA. [119]Netherlands Cancer Institute, Amsterdam, Netherlands. [120]Department of Radiology, Netherlands Cancer Institute, Amsterdam, Netherlands. [121]GROW School of Oncology and Developmental Biology, Maastricht, Netherlands. [122]Clinica Imbanaco Grupo Quirón Salud, Cali, Colombia. [123]Universidad del Valle, Cali, Colombia. [124]The University of Edinburgh, Edinburgh, UK. [125]Department of Industrial and Systems Engineering, University of Iowa, Iowa, USA. [126]Department of Industrial and Systems Engineering, Department of Radiation Oncology, University of Iowa, Iowa City, IA, USA. [127]Department of Radiation Oncology, University of Iowa, Iowa City, IA, USA. [128]MD Anderson Cancer Center, University of Texas, Houston, TX, USA. [129]Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland. [130]Department of Neurology, Clinical Neuroscience Center, University Hospital Zurich and University of Zurich, Zurich, Switzerland. [131]Department of Neuroradiology, Clinical Neuroscience Center, University Hospital Zurich and University of Zurich, Zurich, Switzerland. [132]University of Alabama in Birmingham, Birmingham, AL, USA. [133]Department of Biomedical Informatics, Stony Brook University, Stony Brook, New York, USA. [134]Department of Radiation Oncology, Stony Brook University, Stony Brook, NY, USA. [135]Department of Radiology, Stony Brook University, Stony Brook, NY, USA. [136]Scientific Data Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [137]Escuela Superior Politecnica del Litoral, Guayaquil, Guayas, Ecuador. [138]Sociedad de Lucha Contra el Cancer - SOLCA, Guayaquil Ecuador, Guayaquil, Ecuador. [139]Universidad Católica de Cuenca, Cuenca, Ecuador. [140]Universidad de Concepción, Concepción, Biobío, Chile. [141]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK. [142]Department of Radiology, University College Hospital Ibadan, Oyo, Nigeria. [143]Clinix Healthcare, Lagos, Lagos, Nigeria. [144]Department of Radiology, Muhammad Abdullahi Wase Teaching Hospital, Kano, Nigeria. [145]Department of Radiology, Obafemi Awolowo University Ile-Ife, Ile-Ife, Osun, Nigeria. [146]Center for Global Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [147]School of Computing, Queen's University, Kingston, ON, Canada. [148]Department of Electrical Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran. [149]The Faculty of Arts & Sciences, Queen's University, Kingston, ON, Canada. [150]Department of Oncology, Queen's University, Kingston, ON, Canada. [151]Center for Research and Innovation, American College of Radiology, Philadelphia, PA, USA. [152]Data Science Institute, American College of Radiology, Reston, VA, USA. [153]Center for Biomedical Informatics and Information Technology, National Cancer Institute (NCI), National Institute of Health, Bethesda, MD, USA. [154]These authors contributed equally: Sarthak Pati, Ujjwal Baid, Brandon Edwards.
✉e-mail: sbakas@upenn.edu

doi: 10.14735/amcsnn2017647

# Moderní techniky MR zobrazení u roztroušené sklerózy

## State-of-the-Art MRI Techniques for Multiple Sclerosis


### Souhrn

Magnetická rezonance (MR) je v současnosti klíčovou součástí diagnostiky roztroušené sklerózy. Kromě konvenčních technik založených na hodnocení počtu a lokalizace viditelných lézí mozku a míchy zaznamenáváme v posledních letech rychlý rozvoj nových technik MR zobrazení, které poskytují nové kvantitativní biomarkery lépe charakterizující patologické strukturální změny tkání centrálního nervového systému vzniklé v důsledku demyelinizačního onemocnění. V tomto článku jsou shrnuty nové trendy v MR diagnostice roztroušené sklerózy po stránce technických základů jednotlivých metod, možností analýzy dat i jejich praktického využití.

### Abstract

Magnetic resonance imaging (MRI) is currently a key component of multiple sclerosis diagnostics. In addition to conventional techniques, based on the evaluation of the number and localization of visible brain and spinal cord lesions, in recent years we have seen a rapid development of new MRI techniques providing new quantitative biomarkers which better characterize pathological structural changes in central nervous system tissues occurring due to a demyelinating disease. This article summarizes new trends in MRI diagnostics of multiple sclerosis in terms of the technical foundations of different methods, possibilities for data analysis and their practical use.


Tato práce byla podpořena grantem MZ ČR NV15-32133A a fondy Lékařské fakulty MU na podporu juniorského výzkumníka (M. Keřkovský).

Supported by Czech health research council of the Ministry of Health of the Czech Republic (NV15-32133A) and by funds from the Faculty of Medicine MU to junior researcher (M. Keřkovský).

**Autoři deklarují, že v souvislosti s předmětem studie nemají žádné komerční zájmy.**
The authors declare they have no potential conflicts of interest concerning drugs, products, or services used in the study.
**Redakční rada potvrzuje, že rukopis práce splnil ICMJE kritéria pro publikace zasílané do biomedicínských časopisů.**
The Editorial Board declares that the manuscript met the ICMJE "uniform requirements" for biomedical papers.

**M. Keřkovský[1], J. Stulík[1], I. Obhlídalová[2], P. Praksová[2], J. Bednařík[2], M. Dostál[1,3], M. Kuhn[4–6], A. Šprláková-Puková[3], M. Mechl[1]**

[1] Klinika radiologie a nukleární medicíny LF MU a FN Brno
[2] Neurologická klinika LF MU a FN Brno
[3] Biofyzikální ústav, LF MU a FN Brno
[4] Psychiatrická klinika LF MU a FN Brno
[5] Institut biostatistiky a analýz, LF MU, Brno
[6] Behaviorální a sociální neurovědy, CEITEC – Středoevropský technologický institut MU

**MUDr. Miloš Keřkovský, Ph.D.**
**Klinika radiologie**
**a nukleární medicíny**
**LF MU a FN Brno**
**Jihlavská 20**
**625 00 Brno**
**e-mail: kerkovsky.milos@fnbrno.cz**

Přijato k recenzi: 26. 3. 2017
Přijato do tisku: 25. 10. 2017

### Klíčová slova

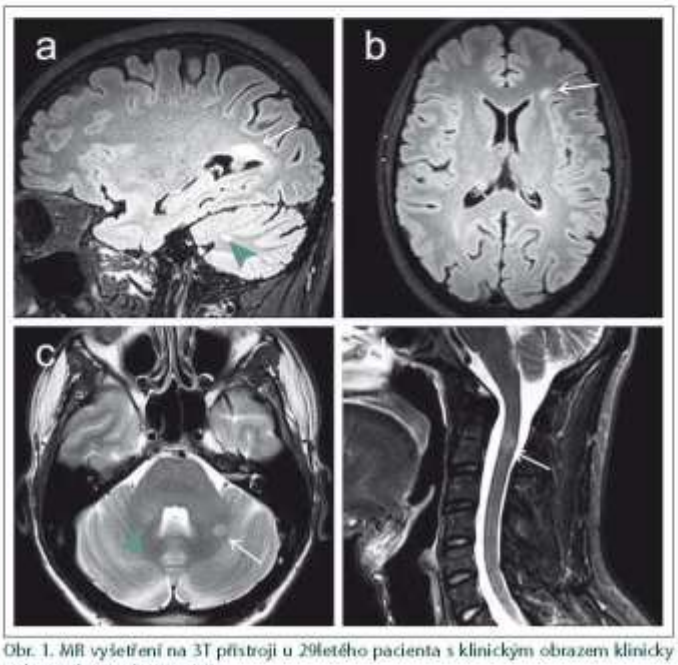
roztroušená skleróza – zobrazení magnetickou rezonancí – zobrazení nervového systému – zobrazení tenzorů difuze – protonová magnetickorezonanční spektroskopie

### Key words

multiple sclerosis – magnetic resonance imaging – neuroimaging – diffusion tensor imaging – proton magnetic resonance spectroscopy

## Úvod

Roztroušená skleróza (RS) představuje chronické zánětlivé onemocnění centrálního nervového systému (CNS), které je z patomorfologického hlediska charakterizováno přítomností zánětlivé infiltrace, demyelinizace, axonálního poškození a gliózy v různých oblastech CNS. Predilekčně jsou postiženy zrakové nervy, mozkový kmen, mozeček, dále periventrikulární a subkortikální bílá hmota mozkových hemisfér [1]. Je též známo, že patologický proces u pacientů s RS není limitován pouze na bílou hmotu, nýbrž postihuje často i oblasti kortikální a hluboké subkortikální šedé hmoty mozku [2].




**Obr. 1. MR vyšetření na 3T přístroji u 29letého pacienta s klinickým obrazem klinicky izolovaného syndromu (CIS).**

Obr. 1a) 3D sekvence FLAIR (fluid attenuated inversion recovery) v sagitální rovině s nálezem periventrikulárního ložiska při trigonu pravé postranní komory (šipka) a dalšího drobnějšího ložiska mozečku (zelená šipka).

Obr. 1b) Rekonstrukce FLAIR obrazu v axiální rovině znázorňující subkortikální lézi vlevo frontálně (šipka).

Obr. 1c) T2 vážený obraz v axiální rovině s nálezem ložiska v levé mozečkové hemisféře (šipka) a v pravém mozečkovém pedunklu (zelená šipka).

Obr. 1d) STIR (short-tau inversion recovery) zobrazení krční míchy v sagitální rovině, šipka označuje míšní ložisko v etáži C2. Ložiska mozečku byla detekována nově v porovnání s minulým vyšetřením, nález tak splňuje kritéria diseminace v prostoru i v čase, což značí progresi CIS do klinicky definitivní roztroušené sklerózy.

**Fig. 1. MRI examination on a 3T device in a 29-year-old patient with a clinically isolated syndrome (CIS) clinical image.**

Fig. 1a) Fluid attenuated inversion recovery (FLAIR) 3D sequence in the sagittal plane indicating a periventricular plaque near the trigone of the right lateral chamber (arrow) and another smaller plaque of the cerebellum (green arrow).

Fig. 1b) Reconstruction of the FLAIR image in the axial plane showing the subcortical lesion in the left frontal lobe (arrow).

Fig. 1c) T2-weighted image in the axial plane indicating a plaque in the left cerebellar hemisphere (arrow) and in the right cerebellar peduncle (green arrow).

Fig. 1d) STIR (short-tau inversion recovery) imaging of the spinal cord in the sagittal plane, arrow marking the spinal cord plaque at the level of C2. Cerebellar plaques were newly detected in comparison to the previous examination, and the finding thus fulfils the criteria of dissemination in space and time indicating the progression of CIS into clinically definitive multiple sclerosis.

Magnetická rezonance (MR) hraje v současnosti klíčovou roli v diagnostice RS. MR diagnostika je založena především na využití T2 vážených sekvencí a zobrazení FLAIR (fluid attenuated inversion recovery), pomocí kterých lze detekovat hyperintenzní léze mozku či míchy. Ukazuje se však, že konvenční techniky MR zobrazení neumožňují zcela komplexní náhled na patofyziologické procesy v rámci RS. Tyto limitace lze dokumentovat např. poměrně chabou korelací MR nálezů s klinickou symptomatikou a diskrepancí mezi MR zobrazením a histopatologickými nálezy [3,4]. Dále je známo, že konvenční techniky MR mají relativně omezené možnosti detekce lézí šedé hmoty a difuzních změn v bílé hmotě [5]. V posledních letech se začínají využívat nejrůznější nové techniky MR zobrazení, jejichž rozvoj je spjat s celkovým vývojem MR technologie a které nabízejí komplexnější náhled na strukturální poškození CNS v rámci RS. U některých technik je nespornou výhodou možnost kvantifikace nejrůznějších parametrů, které se mohou stát cennými biomarkery v diagnostice a sledování vývoje demyelinizačního onemocnění. V dalších odstavcích pojednáme o vybraných technikách MR zobrazování z hlediska základů techniky a analýzy získaných dat i možností praktického využití u pacientů s RS.

## Konvenční techniky

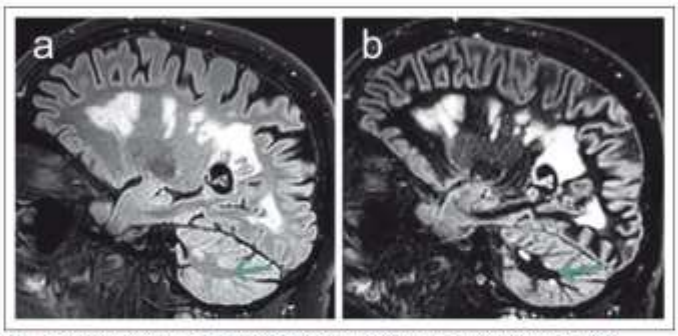
Detekce T2 hyperintenzních ložisek je základem konvenčního radiologického hodnocení MR vyšetření u pacientů s RS nebo s klinicky izolovaným syndromem (CIS), který představuje iniciální stadium demyelinizačního onemocnění [6]. Z hlediska diferenciální diagnostiky a predikce vývoje CIS do klinicky definitivní RS je zásadní zejména zhodnocení počtu a lokalizace ložisek, případně jejich postkontrastního sycení a dynamiky MR nálezu v čase. Tyto atributy jsou součástí původních tzv. McDonaldových kritérií, aktuálně v poslední revizi z roku 2010 s následným upřesněním doporučeními skupiny MAGNIMS pro radiologická diagnostická MR kritéria z 2016 (obr. 1) [7].

I v oblasti těchto tzv. konvenčních technik však dochází k určitému vývoji, v této souvis-

losti je významná zejména otázka senzitivity detekce ložisek. V oblasti zobrazení míchy je kromě T2 zobrazení k dispozici již běžně využívána sekvence STIR (short-tau inversion recovery), která disponuje lepším kontrastním rozlišením demyelinizačních lézí v porovnání s T2 váženým obrazem, a usnadňuje tak jejich detekci [8]. Zejména na 3T MR přístrojích lze s výhodou využít též nové techniky 3D zobrazení sekvencí FLAIR, které disponují vyšší senzitivitou pro detekci demyelinizačních ložisek bílé hmoty mozkové [9].

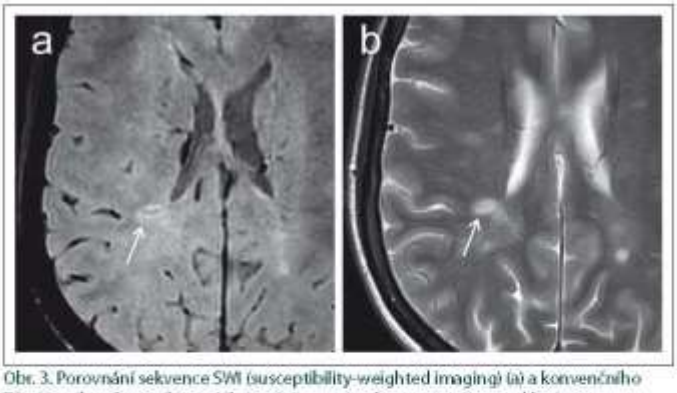
Další zajímavou možností je zobrazení „double inversion recovery" (DIR). Tato sekvence pomocí dvou inverzních pulzů potlačuje zároveň signál mozkomíšního moku i bílé hmoty mozkové, čímž zvyšuje kontrast mezi bílou hmotou a kortexem [10]. Bylo prokázáno, že tato sekvence vykazuje vyšší senzitivitu pro detekci lézí v bílé hmotě oproti T2 a FLAIR vzhledem k vyššímu kontrastu ložisek vůči okolí, umožňuje též lepší detekci lézí infratentoriálních (obr. 2) [11]. Další výhodou techniky DIR je lepší detekce intrakortikálních lézí [12]. Již delší dobu je z histopatologických studií známo, že kortikální postižení je součástí patofyziologie tohoto onemocnění [13] a pomocí techniky DIR byly kortikální léze prokázány řadou autorů, a to i v nejčasnějších stadiích onemocnění [14] nebo u pacientů bez viditelných lézí v bílé hmotě [15]. Někteří autoři též poukazují na signifikantní korelace počtu kortikálních lézí s tíží kognitivního deficitu u pacientů s RS [16] nebo na korelace s mírou fyzické disability [14].

Na hranici konvenčních technik MR zobrazení lze řadit susceptibilně vážené zobrazení (susceptibility-weighted imaging, SWI). Susceptibilitou označujeme fyzikální vlastnost, která charakterizuje míru magnetizace určitého materiálu v magnetickém poli [17]. Celkovou susceptibilitu mozkové tkáně určuje převážně podíl diamagnetické vody ve tkáni, přítomnost paramagnetického železa, stupeň oxygenace krve v kapilárách a žilách a v neposlední řadě zastoupení diamagnetických složek myelinu [18]. Jednou ze zajímavých možností této techniky je zobrazení centrální venuly v rámci demyelinizačních lézí (obr. 3), což umožňují silné paramagnetické vlastnosti deoxyhemoglobinu v těchto žilních strukturách [19]. Histopatologické studie potvrzují perivenózní lokalizaci demyelinizačních plak [20], v souladu s tím řada studií pomocí SWI zobrazení prokázala přítomnost centrální věny u většiny demyelinizačních ložisek u pacientů s RS




**Obr. 2. Srovnání sekvencí 3D FLAIR (fluid attenuated inversion recovery) (a) a DIR (double inversion recovery) (b) u pacienta s pokročilým postižením v rámci roztroušené sklerózy.** Obraz DIR disponuje zřetelně lepším kontrastním rozlišením demyelinizačních lézí vzhledem k potlačení signálu normální bílé hmoty. Lépe jsou detekovatelná zejména drobná ložiska mozečku (šipky).

Fig. 2. Comparison of 3D FLAIR (fluid attenuated inversion recovery) (a) and DIR (double inversion recovery) (b). Sequences in a patient with advanced multiple sclerosis disability. The DIR image shows a markedly better contrast resolution of demyelinating lesions due to suppression of the signal of normal white matter. In particular, small cerebellar plaques are better detected (arrows).

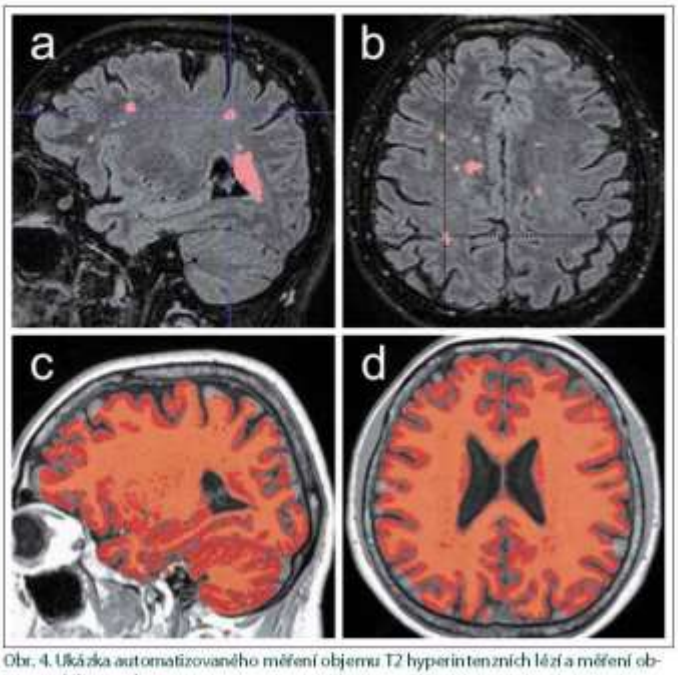



**Obr. 3. Porovnání sekvence SWI (susceptibility-weighted imaging) (a) a konvenčního T2 váženého obrazu (b) v axiální rovině u pacientky s roztroušenou sklerózou.** SWI umožňuje zobrazit centrální venulu v rámci periventrikulárního demyelinizačního ložiska (označeno šipkami) v podobě jemného hypointenzního proužku. V T2 obraze centrální venula prakticky není detekovatelná.

Fig. 3. Comparison of the SWI (susceptibility-weighted imaging) (a) sequence and a conventional T2-weighted image (b) in the axial plane in a patient with multiple sclerosis.

SWI allows for displaying the central venule within the periventricular demyelination plaque (marked with arrows) in the form of a tiny hypointense band. In the T2 image the central venule is practically undetectable.

v porovnání s významně menším zastoupením tohoto nálezu u ložisek odlišné etiologie [21,22]. Autoři Tallantyre et al udávají 80% výskyt perivenulárních ložisek ze všech hodnocených T2 hyperintenzních lézí u skupiny pacientů s RS oproti 19 % u kontrolní skupiny subjektů s T2 hyperintenzními ložisky jiné etiologie. Jako hraniční hodnota pro odlišení pacientů s RS je zde uváděno 40 % perivenulárně lokalizovaných lézí [22]. V další




**Obr. 4. Ukázka automatizovaného měření objemu T2 hyperintenzních lézí a měření objemu celého mozku.**

Obr. 4a, b) Maska segmentace lézí v bílé hmotě mozkové na podkladě rekonstrukce 3D FLAIR (fluid attenuated inversion recovery) zobrazení v sagitální (a) a transverzální (b) rovině. Pro segmentaci je využito několika nástrojů platformy FSL umožňující separátní segmentaci šedé a bílé hmoty mozkové. Za pomoci registrace s normalizovaným obrazem bílé hmoty jsou segmentovány hyperintenzní léze se stanovením jejich celkového objemu [27].

Obr. 4c, d) Maska segmentace celého objemu mozku pomocí aplikace SIENAX. Tento algoritmus vyvinutý pro platformu FSL umožňuje automatizovanou segmentaci celého objemu mozku normalizovaného na velikost hlavy, dále separátní segmentaci a změření objemu šedé a bílé hmoty a objemu kompartmentu mozkomíšního moku [28].

**Fig. 4. Demonstration of automated volume measurement of T2 hyperintense lesions and volume measurement of the entire brain.**

Fig. 4a, b) Mask of lesion segmentation in white brain matter on the basis of a reconstructed 3D FLAIR (fluid attenuated inversion recovery) display in the sagittal (a) and transversal (b) planes. Several instruments of the FSL platform are used for segmentation, enabling separate segmentation of grey and white brain matter. Using registration with a normalized white matter image, hyperintense lesions are segmented and their total volume is determined [27].

Fig. 4c, d) Segmentation mask of the entire brain volume using the SIENAX application. This algorithm developed for the FSL platform enables automated segmentation of the entire brain volume normalized to head size as well as separate segmentation and measurement of the volume of grey and white matter and volume of the cerebrospinal fluid compartment [28].

prácí autorů Kilsdonk et al je dokumentován relativně menší rozdíl v zastoupení lézí s centrální venulou u pacientů s RS (74 %) oproti pacientům s ložisky vaskulární etiologie (47 %). Při stanovení cut-off hodnoty 52 % však při společném hodnocení celkového počtu ložisek a přítomnosti centrální venuly bylo stále možné odlišit jednotlivé etiologicky různé skupiny pacientů se senzitivitou a specificitou 88 % [23]. SWI tak skýtá určitý potenciál pro odlišení demyelinizačních lézí od ložisek odlišné etiologie, což lze považovat v rámci diagnostiky RS za obecný problém. Zároveň je však třeba říci, že specificita tohoto biomarkeru není doposud zcela ověřena a byla studována jen ve vztahu k omezenému množství patologických stavů mimo RS. Jeho spolehlivost proto musí být ještě ověřena [24]. Praktickému využití této techniky též příliš nepřispívá skutečnost, že je v současnosti doménou převážně experimentálních „high-field" MR přístrojů [25].

Další známkou, kterou lze pozorovat na sekvencích typu SWI, jsou plošné hypointenzity v rámci demyelinizačních lézí nebo jemný hypointenzní lem v periferii ložisek. Tyto nálezy jsou též některými autory uváděny jako relativně specifické pro CIS nebo RS v porovnání s jinými druhy neurologických onemocnění [23,26], ačkoli patofyziologický podklad tohoto jevu není doposud zcela objasněn, zvažován je zejména podíl depozit metabolitů železa, případně i role volných radikálů ve spojitosti se zánětlivým procesem [27].

## Volumometrie

V souvislosti s nástupem nových léčebných preparátů jsou vyvíjeny velké snahy najít spolehlivé prognostické markery, pomocí nichž by bylo možné individuálně predikovat budoucí průběh a aktivitu choroby. Jako slibné se v této souvislosti jeví zejména měření počtu a objemu T2 hyperintenzních lézí a kvantifikace stupně mozkové atrofie (obr. 4). Výhodou těchto technik jsou obvykle semiautomatizované postupy a skutečnost, že jsou jako zdrojová data využity konvenční sekvence MR zobrazení, které jsou zároveň použity pro běžné radiologické hodnocení. Pro validní volumometrickou analýzu však i tyto konvenční sekvence musí splňovat určité náležitosti, a to zejména dostatečné prostorové rozlišení, pro účely hodnocení mozkové atrofie je obvykle využívána 3D T1 sekvence gradientního echa s velikostí voxelu kolem 1 $mm^3$. Pro kvantifikaci T2 hyperintenzních lézí lze s výhodou využít 3D FLAIR sekvence turbo-spinového echa s variabilní hodnotou sklápěcího úhlu disponující obdobně vysokým rozlišením [28]. MR protokoly používané pro diagnostiku RS je tedy třeba do určité míry optimalizovat, aby data bylo možné použít pro validní měření objemu. Zejména pro účely longitudinálního sledování vývoje onemocnění u individuálních pacientů je pro dosažení co nejkonzistentnějších výsledků třeba

zdůraznit také potřebu standardizace zobrazovací diagnostiky. Jedná se o provádění kontrolních vyšetření pokud možno na stejném MR přístroji za pomoci standardního diagnostického protokolu a provádění analýz obrazových dat při použití stále stejných softwarových nástrojů. Již v počátku onemocnění může zjištění počtu případně objemu hyperintenzních ložisek v T2 obraze pomoci v odhadu pacientovy prognózy. Samotná přítomnost T2 hyperintenzních ložisek v počátku onemocnění s sebou nese riziko konverze CIS do klinicky definitivní RS v dlouhodobém horizontu v 60–80 % oproti 20% riziku konverze při negativním vstupním MR nálezu [6]. Řada autorů však udává též asociaci počtu a celkového objemu T2 hyperintenzních ložisek detekovaných v počátku onemocnění se zvýšeným rizikem pozdější konverze do klinicky definitivní RS [29].

Podle starších studií při dalším sledování koreluje počet a objem lézí s vývojem dlouhodobé disability již relativně méně. Tento jev je označován jako klinicko-radiologický paradox [30]. Další longitudinální studie nicméně prokázaly korelaci zhoršení klinického stavu v delším časovém období s vyšším počtem a objemem ložisek detekovaných v několika prvních letech onemocnění [31]. Někteří autoři poukazují též na význam objemu T2 hyperintenzních ložisek na vstupním MR vyšetření z hlediska rozvoje krátkodobé disability se zhoršením skóre EDSS (Expanded Disability Status Scale) [32,33]. Dalším parametrem, který lze pomocí výpočetní analýzy obrazu kvantifikovat z konvenčních technik T1 3D zobrazení, je míra mozkové atrofie. Progresivní ztráta objemu mozkové tkáně je obvyklým rysem RS, za níž stojí zejména ztráta myelinu a oligodendrocytů a v neposlední řadě i ztráta neuronů a neuroglie šedé hmoty mozkové [34,35]. Již v roce 1999 autoři Dastidar et al prokázali korelaci objemu kompartmentu mozkomíšního moku vyjadřujícího stupeň atrofie s klinickým skóre EDSS [36]. Mozková atrofie bývá též považována za poměrně časnou známku RS a její vývoj v prvním roce je významným prediktivním faktorem pro budoucí zhoršení [37]. Autoři Minneboo et al poukazují na význam míry atrofie mozku pro predikci vývoje klinického postižení v prvních letech onemocnění RS [32].

Je však třeba se zmínit o tom, že měřené změny objemu mozku nemusí být dány pouze skutečnou atrofií podmíněnou ztrátou mozkové tkáně. V několika studiích byl pozorován pokles objemu mozku především v prvním roce imunomodulační léčby oproti kontrolním skupinám. Naopak v druhém a třetím roce byl pozorován protektivní efekt léčby v podobě redukce ztráty objemu mozku [38]; obdobné změny byly patrny též u pacientů léčených natalizumabem [39]. Tento jev, označovaný jako „pseudoatrofie", by mohl být spojen s ústupem edému mozkové tkáně v úvodu léčby a je třeba s ním počítat při interpretaci výsledků studií zkoumajících protektivní účinky léčby u pacientů s RS [40].

Je též známo, že jednotlivé oblasti mozku mohou atrofovat různou rychlostí, např. atrofie corpus callosum, thalamu, hypothalamu, putamen, nucleus caudatus či mozkového kmene se ukázala jako významný prediktor konverze CIS do klinicky definitivní RS [41,42]. Za významné lze považovat zejména změny objemu thalamů a kalózního tělesa. Řada autorů uvádí silné korelace atrofizace thalamu s klinickou progresí onemocnění [43]. Autoři Vaněčková et al udávají atrofii kalózního tělesa v průběhu prvního roku onemocnění jako významný prediktor rozvoje pozdější disability. Pro kvantifikaci je zde využito jednoduché a v praxi aplikovatelné měření plochy kalózního tělesa na sagitálních MR obrazech [44]. Klinický význam atrofie v ostatních zmiňovaných oblastech je třeba ještě ověřit [43].

Regionální změny objemu mozku je obecně možné kvantifikovat pomocí různých více či méně automatizovaných technik. Jako příklad lze uvést volně dostupné softwarové nástroje VBM (voxel-based morphometry) platformy SPM (statistical parametric mapping) [45]. Obdobná je situace v případě volumometrické analýzy celkového objemu mozku, jako příklad automatizovaného řešení umožňujícího stanovení celkového objemu mozku, objemu šedé a bílé hmoty a objemu kompartmentu mozkomíšního moku můžeme uvést software SIENAX [46].

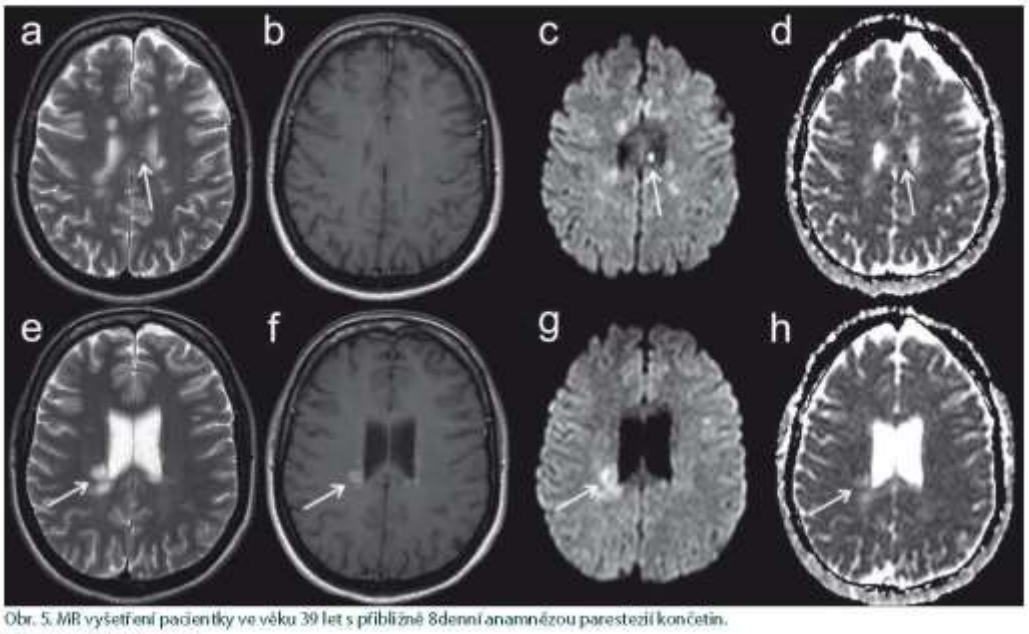
Z výše uvedeného je zřejmé, že kvantifikace počtu a objemu T2 hyperintenzních lézí i měření mozkové atrofie je cenný nástroj umožňující predikovat klinický vývoj u pacientů s RS a sledovat odpověď na léčbu. Tyto techniky tak mají reálnou šanci prosadit se v praxi a stát se standardní součástí MR diagnostiky u pacientů s RS. Většímu praktickému rozšíření těchto metod doposud brání zejména nutnost použití externích softwarových aplikací, což zvyšuje časovou náročnost a celkově v praxi komplikuje hodnocení. Byla by proto žádoucí širší implementace vysoce automatizovaných softwarových nástrojů do komerčních aplikací využívaných pro rutinní hodnocení MR vyšetření.

## Zobrazení difuze

Jako difuze se označuje náhodný pohyb molekul vody ve tkáni označovaný jako tzv. Brownův pohyb. Citlivosti sekvencí spinového i gradientního echa MR zobrazení vůči tomuto jevu lze docílit použitím přídatného magnetického gradientu, jehož charakteristiku (zejm. amplitudu a časový průběh) vyjadřuje tzv. b faktor [47]. Míru difuzivity molekul vody ve tkáni se dá vyjádřit číselně hodnotou ADC (apparent diffusion coefficient). K výpočtu ADC mapy je třeba dvou měření s různou hodnotou b faktoru.

Je známo, že v rámci demyelinizačních lézí dochází ke změnám difuzivity. Autoři Christiansen et al publikovali již v roce 1993 práci analyzující ADC hodnoty u malé skupiny pacientů s RS s nálezem signifikantně vyšších ADC hodnot chronických demyelinizačních lézí v porovnání s normálně vyhlížející bílou hmotou (normal appearing white matter, NAWM), ale též zvýšení ADC hodnot u akutních plak [48]. Obdobně Yurtsever et al udávají zvýšení ADC hodnot u aktivních lézí v porovnání s NAWM [49]. Novější údaje však poukazují na skutečnost, že difuzivita akutních demyelinizačních lézí se rapidně mění zejména v prvních 10 dnech od ataky. Autoři Eisele et al popisují obraz restrikce difuze se snížením ADC hodnot u akutních lézí v rozmezí 0–7 dní od vzniku klinických obtíží, pseudonormalizaci ADC hodnot v 7–10. dni a zvýšení ADC hodnot v rozmezí 10 dní až 4 týdnů (obr. 5) [50]. Lze tak říci, že změny difuzivity patří do obrazu RS a z pohledu neuroradiologa tak mohou mít určitý význam v diferenciální diagnostice tohoto onemocnění. V neposlední řadě můžeme uvést též změny difuzivity u ložisek progresivní multifokální leukoencefalopatie (PML), která je možnou komplikací u pacientů s RS léčených biologickou léčbou [51].

Technika zobrazení tenzoru difuze (diffusion tensor imaging; DTI) je postavena na principech difuzního MR zobrazení. Klíčovým prvkem je zde však detekce směrově závislosti difuzivity molekul vody ve tkáni v závislosti na směru použitého magnetického gradientu, tzv. anizotropie difuzivity [52]. Tuto směrovou závislost lze charakterizovat pomocí opakovaných měření s proměnnou orientací směru magnetického gradientu. Konečným výsledkem je matematická konstrukce 3D elipsoidu, jehož tvar a orientace




**Obr. 5. MR vyšetření pacientky ve věku 39 let s přibližně 8denní anamnézou parestezií končetin.**

Obr. 5a, e) T2 vážený obraz.

Obr. 5b, f) T1 obraz po aplikaci kontrastní látky.

Obr. 5c, g) Izotropní zobrazení difuze (b = 1 000 s/mm²).

Obr. 5d, h) Mapa ADC (apparent diffusion coefficient).

MR nález má charakter demyelinizačního onemocnění, splňuje kritéria diseminace v prostoru i v čase. Šipkami jsou označena ložiska s různou mírou difuzivity, u nichž lze proto usuzovat na různé stáří. Drobné ložisko kalózního tělesa (a–d) se postkontrastně nesytí, vykazuje však výraznou restrikci difuze se sníženým hodnoty ADC; dle literárních údajů lze odhadovat stáří ložiska na méně než 7 dní. Naopak ložisko vpravo periventrikulárně (e–h) se již postkontrastně sytí a ADC hodnoty jsou zvýšené, tato léze vznikla patrně o několik dní dříve [50].

**Fig. 5. MRI examination in a 39-year-old female patient with approximately eight-day history of limb paraesthesia.**

Fig. 5a, e) T2-weighted image.

Fig. 5b, f) T1 image after administration of contrast medium.

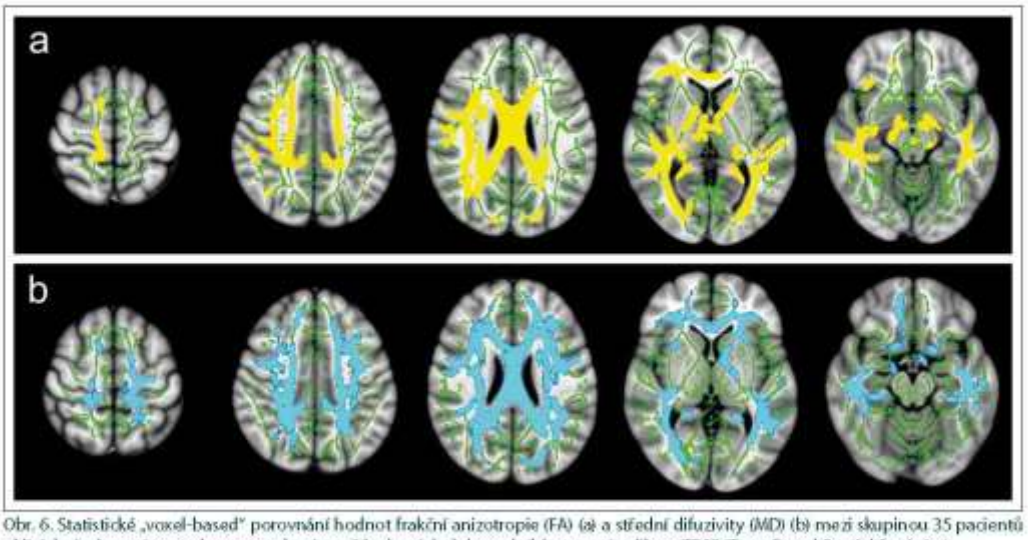
Fig. 5c, g) Isotropic diffusion-weighted imaging (b = 1,000 s/mm).

Fig. 5d, h) ADC (apparent diffusion coefficient) map.

MRI finding has the character of a demyelinating disease and fulfils the criteria of dissemination in space and time. Arrows indicate plaques with various degrees of diffusivity in which various ages can thus be inferred. A small plaque of a callous body (a–d) does not enhance after contrast medium administration, but does demonstrate substantial restriction of diffusion by decreasing the ADC value. According to the literature, its age can be estimated as less than 7 days. On the other hand, the plaque on the right periventricular (e–h) enhances after contrast medium administration and ADC values are increased. This lesion apparently had formed several days earlier [50].

charakterizují anizotropii difuze v jednotlivých voxelech. Tvar elipsoidu může být popsán pomocí tří hodnot, které reprezentují velikost tří jeho hlavních os. Z těchto parametrů se dá vypočíst několik skalárních veličin, z nichž prakticky nejvýznamnější jsou hodnoty frakční anizotropie (FA) udávající relativní míru anizotropie v daném voxelu a střední difuzivita (mean diffusivity; MD) či ADC, jež vyjadřují celkovou míru difuzivity nezávislé na anizotropii difuze [47].

Bylo zjištěno, že pomocí kvantifikace uvedených skalárních parametrů lze citlivě detekovat strukturální patologii tkání CNS u nejrůznějších onemocnění, zejména index FA je v současnosti považován za parametr senzitivní k narušení integrity bílé hmoty a je také jedním z nejčastěji sledovaných parametrů ve studiích využívajících DTI zobrazení [53]. V minulosti byla publikována celá řada studií využívajících DTI pro detekci patologických změn mozku či míchy u pacientů s RS. Existují např. důkazy o korelaci parametrů difuzivity s histopatologickými nálezy demyelinizace a axonální dezintegrace [54]. Tyto patologické změny vedou k nárůstu difuzivity napříč nervovými trakty způsobujícímu mimo jiné pokles hodnot parametru FA v důsledku snížené anizotropie difuzivity. V souladu s tím je zjištění, že dochází k signifikantní změně skalárních parametrů difuzivity v rámci T2 hyperintenzních ložisek v porovnání s NAWM [55]. Patrně nejpozo-




**Obr. 6. Statistické „voxel-based" porovnání hodnot frakční anizotropie (FA) (a) a střední difuzivity (MD) (b) mezi skupinou 35 pacientů s klinicky izolovaným syndromem a skupinou 32 zdravých dobrovolníků pomocí aplikace TBSS (Tract-Based Spatial Statistics).**

Byly nalezeny rozsáhlé oblasti bílé hmoty se statisticky významným ($p < 0,05$) snížením hodnot FA (žluté) a zvýšením hodnot MD (modře) u pacientů v porovnání s dobrovolníky. Zelenou barvou je znázorněn průměrný skeleton hlavních traktů bílé hmoty, na který jsou v průběhu zpracování projektovány hodnoty FA a MD jednotlivých subjektů.

Fig. 6. Statistical "voxel-based" comparison of fractional anisotropy (FA) (a) and mean diffusivity (MD) (b) values between a group of 35 patients with clinically isolated syndrome and a group of 32 healthy volunteers using a TBSS (tract-based spatial statistics) application.

Large areas of white matter with statistically significant ($p < 0.05$) decrease in FA values (yellow) and increase in MD values (blue) were recorded in patients as compared to volunteers. Green colour marks the average skeleton of the main tracts of white matter upon which the FA and MD values of the individual subjects are projected during processing.

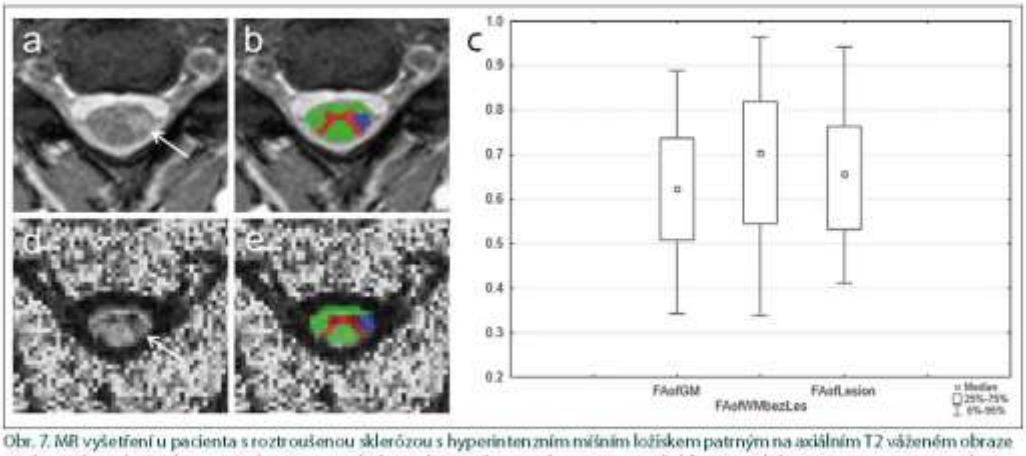
ruhodnější je skutečnost, že lze prokázat změny difuzivity i v rámci samotné NAWM i NAGM (šedá hmota normálního vzhledu) v poměru se zdravými jedinci [56]. DTI je tak možno vnímat jako citlivější metodu pro detekci patologických změn u pacientů s RS v porovnání s konvenčními technikami MR zobrazení. Někteří autoři poukazují též na korelaci změn difuzivity mozku s tíží klinického postižení např. po stránce motorických [57] či kognitivních [58] funkcí. Technika DTI byla v minulých letech aplikována úspěšně i pro zobrazení míchy u pacientů s RS, kdy byla prokázána korelace parametrů DTI s tíží klinického postižení [59] a obdobně jako v případě mozku je poukazováno na abnormality FA v oblastech míchy bez patologického nálezu na konvenčních sekvencích [60].

Existuje celá řada přístupů k metodice analýzy dat DTI. Za nejjednodušší lze považovat měření parametrů difuzivity v rámci manuálně definovaných oblastí zájmu (region of interest; ROI) ve zvolených oblastech mozku či míchy. Nevýhodu tohoto postupu mohou představovat nepřesnosti v umístění ROI a následné zkreslení výsledků. S tímto problémem se snaží vypořádat některé automatizované techniky analýzy DTI dat, jako je např. TBSS (tract-based spatial statistics). Tato metoda umožňuje analýzu difuzivity v celém měřeném objemu a je vhodná zejména pro skupinové statistické zpracování (obr. 6). Je zde využito nelineární registrace map FA následované konstrukcí skeletonu hlavních traktů bílé hmoty a projekcí voxelů jednotlivých subjektů na tento skeleton. Tím je minimalizováno nechtěné zahrnutí voxelů šedé hmoty do analýzy [61]. V případě zobrazení míchy existuje relativně méně možností výpočetní analýzy obrazových dat. I zde však již bylo popsáno několik technik využívajících semiautomatické či automatické postupy při segmentaci celé míchy nebo šedé a bílé hmoty [62], které mohou být využity i v rámci vyhodnocení difuzních parametrů (obr. 7).

Závěrem této části je třeba poznamenat, že DTI je stále spíše v pozici výzkumné aplikace než prakticky používaného diagnostického nástroje. Jednou z limitací, které brání většímu rozšíření této techniky, je poměrně velká časová náročnost DTI sekvencí. Dalším a patrně významnějším problémem je obecný nedostatek standardizace po stránce vlastní akvizice dat a jejich následné softwarové analýzy. Doposud není zcela vyřešena otázka reprodukovatelnosti měřených skalárních parametrů difuzivity při vyšetřeních na různých přístrojích [63] nebo s nastavením různých akvizičních parametrů [64], což komplikuje porovnávání výsledků různých studií a stanovení obecněji platných prahových hodnot.

## Zobrazení magnetizačního transferu

Zobrazení magnetizačního transferu (magnetization transfer imaging; MTI) představuje další metodu, která jde za hranice konvenčních technik MR zobrazení a umožňuje detekovat diskrétní patologické změny tkáně mozku či míchy, k nimž dochází v rámci demyelinizačního onemocnění. Tato technika využívá odlišností v chování volných protonů obsažených v molekulách vody v po-




**Obr. 7. MR vyšetření u pacienta s roztroušenou sklerózou s hyperintenzním mílním ložiskem patrným na axiálním T2 váženém obraze gradientního echa (šipka na panelu a). Léze je sledovatelná i v obraze zobrazení tenzorů difuze v podobě snížení intenzity signálu na mapě frakční anizotropie (FA) (d).**

Obr. 7b, e) Ukázka semiautomatické segmentace bílé hmoty (zeleně), šedé hmoty (červeně) a patologického ložiska (modře) pomocí aplikace ITK-SNAP. Po registraci segmentačních masek mezi T2 a FA obrazem lze kvantifikovat hodnoty FA v jednotlivých tkáních.

Obr. 7c) Rozdíly mezi šedou a bílou hmotou a hyperintenzními ložisky je možné jednoduše graficky vyjádřit pomocí krabicového grafu.

**Fig. 7. MRI examination in a patient with multiple sclerosis with a hyperintense spinal cord plaque apparent on the axial T2-weighted image of gradient echo (arrow on panel a). The lesion is perceptible also in the diffusion tensor imaging image in the form of reduced signal intensity on the fractional anisotropy (FA) map (d).**

Fig. 7b, e) Demonstration of semi-automatic segmentation of white matter (green), grey matter (red), and pathological plaque (blue) using the ITK-SNAP application. After registering the segmentation masks between T2 and FA images, FA values in the individual tissues can be quantified.

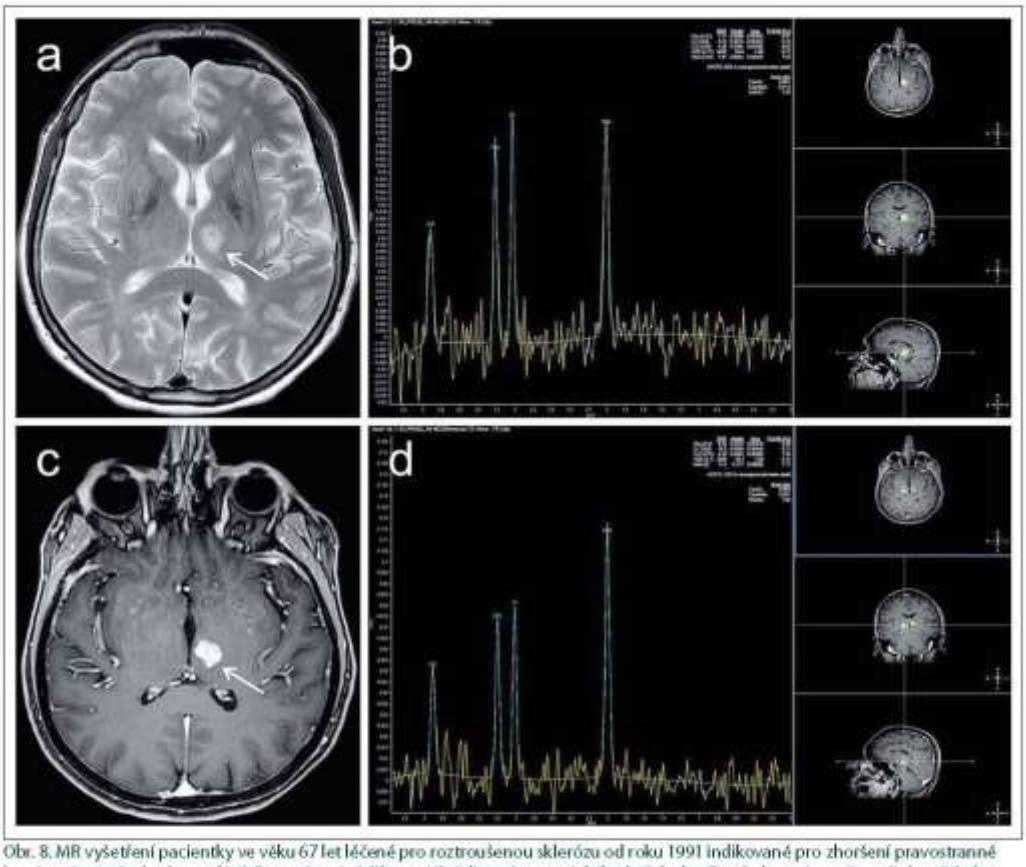
Fig. 7c) Differences between grey and white matter and hyperintense lesions can easily be expressed in a box plot.

rovnání s vázanými protony, u kterých v důsledku magnetických interakcí s okolními makromolekulami dochází k velmi rychlému rozfázování magnetizace a jsou proto charakterizovány velmi krátkými T2 časy. V případě, že jsou vázané protony saturovány pomocí zvláštního saturačního radiofrekvenčního pulzu („off-resonance"), dojde k výměně magnetizace (magnetizačnímu transferu) mezi těmito protony a protony volnými, což ovlivní magnetizaci volných protonů [65]. Tento jev vede k viditelnému snížení intenzity signálu v MR obraze, existují však i možnosti jeho kvantifikace. Patrně nejčastěji používaný a nejjednodušší způsob je využití indexu MTR (magnetization transfer ratio), který představuje relativní rozdíl mezi dvěma měřeními, z nichž pouze u jednoho z nich je aplikován zmiňovaný „off-resonance" saturační pulz. MTR tak nepřímo odráží míru zastoupení makromolekul ve tkáni a umožňuje mimo jiné detekovat ztrátu myelinu a v menší míře axonální dezintegraci [66]. Na využití MTI v diagnostice RS lze nahlížet jako na určitou paralelu DTI. Byla prokázána redukce MTR akutních i chronických demyelinizačních lézí [67], obdobně jako u DTI byla i v případě MTI zaměřena pozornost na změny v rámci NAWM i NAGM s průkazem signifikantní redukce MTR u nejrůznějších fenotypů RS vč. nejčasnějších klinických stadií [68]. Abnormality MTR též dle některých autorů korelují s délkou trvání onemocnění a s tíží klinického postižení [68,69].

## MR spektroskopie

MR spektroskopie je analytická metoda umožňující detekci a kvantifikaci různých sloučenin ve tkáni. K získání MR spektra je možno využít jádra různých prvků, v klinické praxi se však nejčastěji používá jádro vodíku [70]. V praxi se obvykle uplatňuje jedna ze dvou základních technik spektroskopického zobrazení. První z nich je tzv. single-voxel spektroskopie (SVS), která zobrazuje spektrum metabolitů z jediného předem definovaného voxelu. Druhou možností je spektrální akvizice z více voxelů širší oblasti mozku označovaná někdy jako „magnetic resonance spectroscopy imaging" (MRSI) nebo „chemical shift imaging". Pomocí obou zmiňovaných technik lze detekovat koncentraci nejrůznějších metabolitů v mozkové tkáni, jako jsou např. tuky, aminokyseliny (zejm. alanin, glutamin), laktát, N-acetylaspartát (NAA), kreatin (Cr), a myo-inositol [71].

Pro potřeby diagnostiky pacientů s RS se nejčastěji užívá detekce a kvantifikace NAA, resp. poměru NAA/Cr, dále cholinu, laktátu a myo-inositolu. NAA je obsažen v mitochondriích neuronů, v axonech a dendritech, jeho pokles proto nastává při jejich poškození či zániku. Ke snížení jeho koncentrace ale může dojít i relativně při edému či přechodně při omezení neuronální funkce. Pokles NAA je patrný zejména v akutní fázi onemocnění (obr. 8), během remise potom může docházet k jeho částečné normalizaci. Cholin je za normálních okolností vázán v buněčných membránách, k nárůstu detekovatelné volné porce dochází při rozpadu buněk. K tomu může docházet např. při zánětu, tedy i v při-




**Obr. 8. MR vyšetření pacientky ve věku 67 let léčené pro roztroušenou sklerózu od roku 1991 indikované pro zhoršení pravostranné hemiparézy s přechodnou afázií. Pro upřesnění diferenciální diagnózy atypického ložiska levého thalamu (označeno šipkami) bylo provedeno též spektroskopické vyšetření technikou „single-voxel".**

Léze má převážně vysoký signál v T2 obraze (a), postkontrastně se intenzivně sytí prakticky v celém objemu (c). Spektroskopie v oblasti ložiska (b) prokazuje relativní snížení koncentrace N-acetylaspartátu (NAA) vůči kreatinu (Cr) (NAA/Cr = 1,24) v porovnání s normálním spektrem získaným při kontrolním měření v oblasti kontralaterálního thalamu (NAA/Cr = 1,62) (d).

Absence elevace koncentrace cholinu svědčí proti tumorózního původu ložiska, jako nejpravděpodobnější etiologie je označena aktivní demyelinizační plaka. V korelaci s tím došlo na kontrolním vyšetření za 2 týdny k významné regresi T2 hyperintenzity i postkontrastního sycení.

**Fig. 8. MRI examination in a 67-year-old female patient treated for multiple sclerosis since 1991, indicated for deterioration of right-side hemiparesis with transient aphasia. To increase the accuracy of differential diagnosis of an atypical focal lesion of the left thalamus (marked with arrows) a spectroscopic "single-voxel" examination was also performed.**

The lesion has predominantly a high signal in the T2 image (a) and enhanced practically in its entire volume after contrast medium administration (c). Spectroscopy in the area of the lesion (b) demonstrates a relative decrease in N-acetylaspartate (NAA) to creatine (Cr) ratio (NAA/Cr = 1.24) in comparison to the normal spectrum obtained in a control measurement in the area of the contralateral thalamus (NAA/Cr = 1.62) (d). Absence of an elevated choline concentration reduces the probability that the lesion is of tumorous origin. Active demyelinating plaque is marked as the most probable aetiology. In correlation with this, at a follow-up examination 2 weeks later, a substantial regression of T2 hyperintensity and post-contrast enhancement occurred.

padě akutní fáze RS. Obdobně v akutní fázi RS pozorujeme nárůst hodnot laktátu následkem zvýšené anaerobní glykolýzy v terénu zánětu. Bylo zjištěno, že nárůst hodnot myoinositolu je spojen s vyšší metabolickou či proliferační aktivitou astrocytů v plakách RS na rozdíl od lézí bílé hmoty jiné etiologie [72].

Výsledky studií vč. metaanalýz nejsou zatím jednotné. Přesto naznačují, že by pokles koncentrace NAA mohl být markerem predikce klinického postižení pacientů s RS,

a to zejména u již léčených pacientů vzhledem k jeho částečné normalizaci v průběhu terapie [71,73]. Detekce sníženého obsahu NAA a zároveň zvýšených hodnot cholinu a laktátu se jeví jako potenciální marker akutního postižení bílé hmoty časově předcházející korelát v konvenčním obraze MR [72].

## Závěr

Přestože konvenční radiologické hodnocení přítomnosti viditelných lézí mozku a míchy zůstává doposud hlavním nástrojem v oblasti MR diagnostiky RS, existuje nepřeberné množství důkazů o novějších technikách, jako jsou např. DTI, MTI či spektroskopie, které umožňují přesněji detekovat patologické změny tkání CNS. Tyto techniky proto mají potenciál stát se cenným nástrojem v rámci diferenciální diagnostiky v iniciálních stadiích demyelinizačního onemocnění a v dalším průběhu objektivním markerem či prediktorem klinického vývoje a odpovědi na léčbu. Určitou nevýhodou těchto technik jsou nároky na skenovací čas (zejm. DTI a spektroskopie) a obecným problémem výše diskutovaných metod je doposud nedostatek standardizace akvizičních protokolů a v některých případech i omezená reprodukovatelnost měřených dat mezi různými MR přístroji [63]. Dalším aspektem, který do určité míry brání většímu praktickému rozšíření těchto metod, je nejednotnost technik softwarové analýzy naměřených dat a poměrně malá míra implementace automatizovaných technik analýzy obrazu do komerčního softwarového vybavení dodávaného hlavními výrobci MR přístrojů. Budoucí výzkum by proto měl být zaměřen na tyto otázky úzce spojené s technikou vyšetření a na hledání nových dobře reprodukovatelných a kvantifikovatelných parametrů. Půjde-li vývoj tímto směrem, lze v budoucnu očekávat inkorporaci hodnocení těchto nových biomarkerů v rámci standardní diagnostiky u pacientů s RS.

## Literatura


1. Confavreux C, Vukusic S, Moreau T, et al. Relapses and progression of disability in multiple sclerosis. NEngJ Med 2000;343(20):1430–8. doi: 10.1056/NEJM200011163432003.
2. Calabrese M, Favaretto A, Martini V, et al. Grey matter lesions in MS: from histology to clinical implications. Prion 2013;7(3):20–7. doi: 10.4161/pri.23580.
3. Filippi M, Agosta F. Imaging biomarkers in multiple sclerosis. J Magn Reson Imaging 2010;31(4):770–88. doi: 10.1002/jmri.22102.
4. van Waesberghe JH, Kamphorst W, De Groot CJ, et al. Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability. Ann Neurol 1999;46(5):747–54.
5. Bakshi R, Thompson AJ, Rocca MA, et al. MRI in multiple sclerosis: current status and future prospects. Lancet Neurol 2008;7(7):615–25. doi: 10.1016/S1474-4422(08)70137-6.
6. Miller DH, Chard DT, Ciccarelli O. Clinically isolated syndromes. Lancet Neurol 2012;11(2):157–69. doi: 10.1016/S1474-4422(11)70274-5.
7. Filippi M, Rocca MA, Ciccarelli O, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. Lancet Neurol 2016;15(3):292–303. doi: 10.1016/S1474-4422(15)00393-2.
8. Nayak NB, Salah R, Huang JC, et al. A comparison of sagittal short T1 inversion recovery and T2-weighted FSE sequences for detection of multiple sclerosis spinal cord lesions. Acta Neurol Scand 2014;129(3):198–203. doi: 10.1111/ane.12168.
9. Patzig M, Burke M, Brückmann H, et al. Comparison of 3D cube FLAIR with 2D FLAIR for multiple sclerosis imaging at 3 Tesla. Rofo 2014;186(5):484–8. doi: 10.1055/s-0033-1355896.
10. Redpath TW, Smith FW. Technical note: use of a double inversion recovery pulse sequence to image selectively grey or white brain matter. Br J Radiol 1994;67(804):1258–63. doi: 10.1259/0007-1285-67-804-1258.
11. Wattjes MP, Lutterbey GG, Gieseke J, et al. Double inversion recovery brain imaging at 3T: diagnostic value in the detection of multiple sclerosis lesions. Am J Neuroradiol 2007;28(1):54–9.
12. Geurts JJ, Pouwels PJ, Uitdehaag BM, et al. Intracortical lesions in multiple sclerosis: improved detection with 3D double inversion-recovery MR imaging. Radiology 2005;236(1):254–60. doi: 10.1148/radiol.2361040450.
13. Bø L, Vedeler CA, Nyland HI, et al. Subpial demyelination in the cerebral cortex of multiple sclerosis patients. J Neuropathol Exp Neurol 2003;62(7):723–32.
14. Calabrese M, De Stefano N, Atzori M, et al. Detection of cortical inflammatory lesions by double inversion recovery magnetic resonance imaging in patients with multiple sclerosis. Arch Neurol 2007;64(10):1416–22. doi: 10.1001/archneur.64.10.1416.
15. Calabrese M, Gallo P. Magnetic resonance evidence of cortical onset of multiple sclerosis. Mult Scler 2009;15(8):933–41. doi: 10.1177/1352458509106510.
16. Rinaldi F, Calabrese M, Grossi P, et al. Cortical lesions and cognitive impairment in multiple sclerosis. Neurol Sci 2010;31(Suppl 2):235–7. doi: 10.1007/s10072-010-0368-4.
17. Liu C, Li W, Tong KA, et al. Susceptibility-weighted imaging and quantitative susceptibility mapping in the brain. J Magn Reson Imaging 2015;42(1):23–41. doi: 10.1002/jmri.24768.
18. Schenck JF. The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. Med Phys 1996;23(6):815–50. doi: 10.1118/1.597854.
19. Haacke EM, Mittal S, Wu Z, et al. Susceptibility-weighted imaging: technical aspects and clinical applications, part 1. Am J Neuroradiol 2009;30(1):19–30. doi: 10.3174/ajnr.A1400.
20. Fog T. On the vessel-plaque relationships in the brain in multiple sclerosis. Acta Neurol Scand Suppl 1964;40(Suppl 10):9–15.
21. Mistry N, Dixon J, Tallantyre E, et al. Central veins in brain lesions visualized with high-field magnetic resonance imaging: a pathologically specific diagnostic biomarker for inflammatory demyelination in the brain. JAMA Neurol 2013;70(5):623–8. doi: 10.1001/jamaneurol.2013.1405.
22. Tallantyre EC, Dixon JE, Donaldson I, et al. Ultra-high-field imaging distinguishes MS lesions from asymptomatic white matter lesions. Neurology 2011;76(6):534–9. doi: 10.1212/WNL.0b013e318206ca30.
23. Kilsdonk ID, Wattjes MP, Lopez-Soriano A, et al. Improved differentiation between MS and vascular brain lesions using FLAIR* at 7 Tesla. Eur Radiol 2014;24(4):841–9. doi: 10.1007/s00330-013-3080-y.
24. Enzinger C, Barkhof F, Ciccarelli O, et al. Nonconventional MRI and microstructural cerebral changes in multiple sclerosis. Nat Rev Neurol 2015;11(12):676–86. doi: 10.1038/nrneurol.2015.194.
25. Tallantyre EC, Morgan PS, Dixon JE, et al. A comparison of 3T and 7T in the detection of small parenchymal veins within MS lesions. Invest Radiol 2009;44(8):491–4. doi: 10.1097/RLI.0b013e318b4c144.
26. Kelly JE, Mar S, D'Angelo G, et al. Susceptibility-weighted imaging helps to discriminate pediatric multiple sclerosis from acute disseminated encephalomyelitis. Pediatr Neurol 2015;52(1):36–41. doi: 10.1016/j.pediatrneurol.2014.10.014.
27. Absinta M, Sati P, Gaitán MI, et al. Seven-tesla phase imaging of acute multiple sclerosis lesions: a new window into the inflammatory process. Ann Neurol 2012;74(5):669–78. doi: 10.1002/ana.23959.
28. Wetter NC, Hubbard EA, Motl RW, et al. Fully automated open-source lesion mapping of T2-FLAIR images with FSL correlates with clinical disability in MS. Brain Behav 2016;6(3):e00440. doi: 10.1002/brb3.440.
29. Odenthal C, Coulthard A. The prognostic utility of MRI in clinically isolated syndrome: a literature review. AJNR Am J Neuroradiol 2015;36(3):425–31. doi: 10.3174/ajnr.A3954.
30. Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. Curr Opin Neurol 2002;15(3):239–45.
31. Rudick RA, Lee JC, Simon J, et al. Significance of T2 lesions in multiple sclerosis: A 13-year longitudinal study. Ann Neurol 2006;60(2):236–42. doi: 10.1002/ana.20883.
32. Minneboo A, Jaspers B, Barkhof F, et al. Predicting short-term disability progression in early multiple sclerosis: added value of MRI parameters. J Neurol Neurosurg Psychiatry 2008;79(8):917–23. doi: 10.1136/jnnp.2007.124123.
33. Gauthier SA, Mandel M, Guttmann CR, et al. Predicting short-term disability in multiple sclerosis. Neurology 2007;68(24):2059–65. doi: 10.1212/01.wnl.0000264890.97479.b1.
34. De Stefano N, Arnold DL. Towards a better understanding of pseudoatrophy in the brain of multiple sclerosis patients. Mult Scler 2015;21(6):675–6. doi: 10.1177/1352458514564494.
35. Filippi M, Rocca MA, Barkhof F, et al. Association between pathological and MRI findings in multiple sclerosis. Lancet Neurol 2012;11(8):349–60. doi: 10.1016/S1474-4422(12)70003-0.
36. Dastidar P, Heinonen T, Lehtimäki T, et al. Volumes of brain atrophy and plaques correlated with neurological disability in secondary progressive multiple sclerosis. J Neurol Sci 1999;165(1):36–42.
37. Vaněčková M, Seidl Z, Krásenský J, et al. Naše zkušenosti s MR monitorováním pacientů s roztroušenou sklerózou v klinické praxi. Cesk Slov Neurol N 2010;73/106(4):716–20.
38. Zivadinov R, Stosic M, Cox JL, et al. The place of conventional MRI and newly emerging MRI techniques in monitoring different aspects of treatment outcome. J Neurol 2008;255(Suppl 1):61–74. doi: 10.1007/s00415-008-1009-1.
39. Miller DH, Soon D, Fernando KT, et al. MRI outcomes in a placebo-controlled trial of natalizumab in relapsing MS. Neurology 2007;68(17):1390–401. doi: 10.1212/01.wnl.0000260064.77700.f1.
40. Koudriavtseva T, Mainiero C. Brain Atrophy as a Measure of Neuroprotective Drug Effects in Multiple Sclerosis: Influence of Inflammation. Front Hum Neurosci 2016;10:226. doi: 10.3389/fnhum.2016.00226.

**41.** Henry RG, Shieh M, Okuda DT, et al. Regional grey matter atrophy in clinically isolated syndromes at presentation. J Neurol Neurosurg Psychiatry 2008;79(11):1236–44. doi: 10.1136/jnnp.2007.134825.

**42.** Kalincik T, Vaneckova M, Tyblova M, et al. Volumetric MRI markers and predictors of disease activity in early multiple sclerosis: a longitudinal cohort study. PLoS One 2012;7(11):e50101. doi: 10.1371/journal.pone.0050101.

**43.** Kaunzner UW, Gauthier SA. MRI in the assessment and monitoring of multiple sclerosis: an update on best practice. Ther Adv Neurol Disord 2017;10(9):247–61. doi: 10.1177/1756285617708911.

**44.** Vaneckova M, Kalincik T, Krasensky J, et al. Corpus callosum atrophy – a simple predictor of multiple sclerosis progression: a longitudinal 9-year study. Eur Neurol 2012;68(1):23–7. doi: 10.1159/000337683.

**45.** Mechelli A, Price CJ, Friston KJ, et al. Voxel-based morphometry of the human brain: Methods and applications. Current Medical Imaging Reviews 2005;1(2):105–13.

**46.** Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 2002;17(1):479–89.

**47.** Mukherjee P, Berman JI, Chung SW, et al. Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. Am J Neuroradiol 2008;29(4):632–41. doi: 10.3174/ajnr.A1051.

**48.** Christiansen P, Gideon P, Thomsen C, et al. Increased water self-diffusion in chronic plaques and in apparently normal white matter in patients with multiple sclerosis. Acta Neurol Scand 1993;87(3):195–9.

**49.** Yuntsever I, Hakyemez B, Taskapilioglu O, et al. The contribution of diffusion-weighted MR imaging in multiple sclerosis during acute attack. Eur J Radiol 2008;65(3):421–6. doi: 10.1016/j.ejrad.2007.05.002.

**50.** Eisele P, Szabo K, Griebe M, et al. Reduced diffusion in a subset of acute MS lesions: a serial multiparametric MRI study. AJNR Am J Neuroradiol 2012;33(7):1369–73. doi: 10.3174/ajnr.A2975.

**51.** Vaněčková M, Seidl Z, Čáp F, et al. Návrh bezpečnostní MR monitorace u pacientů s roztroušenou sklerózou léčených natalizumabem. Cesk Slov Neurol N 2016;79(6):663–9.

**52.** Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. Biophys J 1994;66:259–67. doi: 10.1016/S0006-3495(94)80775-1.

**53.** Assaf Y, Pasternak O. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. J Mol Neurosci 2008;34(1):51–61. doi: 10.1007/s12031-007-0029-0.

**54.** Mottershead JP, Schmierer K, Clemence M, et al. High field MRI correlates of myelin content and axonal density in multiple sclerosis – a post-mortem study of the spinal cord. J Neurol 2003;250(11):1293–301. doi: 10.1007/s00415-003-0192-3.

**55.** Filippi M, Iannucci G, Cercignani M, et al. A quantitative study of water diffusion in multiple sclerosis lesions and normal-appearing white matter using echo-planar imaging. Arch Neurol 2000;57(7):1017–21.

**56.** Yu CS, Lin FC, Liu Y, et al. Histogram analysis of diffusion measures in clinically isolated syndromes and relapsing-remitting multiple sclerosis. Eur J Radiol 2008;68(2):328–34. doi: 10.1016/j.ejrad.2007.08.036.

**57.** Hubbard EA, Weiner NC, Sutton BP, et al. Diffusion tensor imaging of the corticospinal tract and walking performance in multiple sclerosis. J Neurol Sci 2016;363:225–31. doi: 10.1016/j.jns.2016.02.044.

**58.** Meijer KA, Muhlert N, Cercignani M, et al. White matter tract abnormalities are associated with cognitive dysfunction in secondary progressive multiple sclerosis. Mult Scler 2016;22(11):1429–37. doi: 10.1177/1352458515622694.

**59.** Valsasina P, Rocca MA, Agosta F, et al. Mean diffusivity and fractional anisotropy histogram analysis of the cervical cord in MS patients. Neuroimage 2005;26(3):822–8. doi: 10.1016/j.neuroimage.2005.02.033.

**60.** Hesseltine SM, Law M, Babb J, et al. Diffusion tensor imaging in multiple sclerosis: assessment of regional differences in the axial plane within normal-appearing cervical spinal cord. Am J Neuroradiol 2006;27(6):1189–93.

**61.** Smith SM, Jenkinson M, Johansen-Berg H, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 2006;31(4):1487–505. doi: 10.1016/j.neuroimage.2006.02.024.

**62.** De Leener B, Taso M, Cohen-Adad J, et al. Segmentation of the human spinal cord. MAGMA 2016;29(2):125–53. doi: 10.1007/s10334-015-0507-2.

**63.** Kirnak AS, Paksoy Y, Erol C, et al. Comparison of apparent diffusion coefficient values among different MRI platforms: a multicenter phantom study. Diagn Interv Radiol 2013;19(6):433–7. doi: 10.5152/dir.2013.13034.

**64.** Santarelli X, Garbin G, Ulmar M, et al. Dependence of the fractional anisotropy in cervical spine from the number of diffusion gradients, repeated acquisition and voxel size. Magn Reson Imaging 2010;28(1):70–6. doi: 10.1016/j.mri.2009.05.046.

**65.** Henkelman RM, Stanisz GJ, Graham SJ. Magnetization transfer in MRI: a review. NMR Biomed 2001;14(2):57–64.

**66.** Schmierer K, Scaravilli F, Altmann DR, et al. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. Ann Neurol 2004;56(3):407–15. doi: 10.1002/ana.20202.

**67.** Ropele S, Fazekas F. Magnetization transfer MR imaging in multiple sclerosis. NeuroImaging Clin N Am 2009;19(1):27–36. doi: 10.1016/j.nic.2008.09.004.

**68.** Filippi M, Rocca MA. Magnetization transfer magnetic resonance imaging of the brain, spinal cord, and optic nerve. Neurotherapeutics 2007;4(3):401–13. doi: 10.1016/j.nurt.2007.03.002.

**69.** Hayton T, Furby J, Smith KJ, et al. Grey matter magnetization transfer ratio independently correlates with neurological deficit in secondary progressive multiple sclerosis. J Neurol 2009;256(3):427–35. doi: 10.1007/s00415-009-0110-4.

**70.** Bertholdo D, Watcharakorn A, Castillo M. Brain proton magnetic resonance spectroscopy: introduction and overview. Neuroimaging Clin N Am 2013;23(3):399–80. doi: 10.1016/j.nic.2012.10.002.

**71.** Aboul-Enein F. MR Spectroscopy in Multiple Sclerosis – A New Piece of the Puzzle or Just a New Puzzle In: Kim DH et al, eds. Magnetic Resonance Spectroscopy. Rijeka (Croatia): InTech 2012:49–72.

**72.** Ge Y. Multiple sclerosis: the role of MR imaging. AJNR Am J Neuroradiol 2006;27(6):1165–76.

**73.** Liufriu S, Kornak J, Ratiney H, et al. Magnetic resonance spectroscopy markers of disease progression in multiple sclerosis. JAMA Neurol 2014;71(7):840–7. doi: 10.1001/jamaneurol.2014.805.

# MR Diffusion Properties of Cervical Spinal Cord as a Predictor of Progression to Multiple Sclerosis in Patients with Clinically Isolated Syndrome

Marek Dostál, Miloš Keřkovský, Jakub Stulík, Josef Bednařík, Petra Praksová, Monika Hulová, Yvonne Benešová, Eva Koriťáková, Andrea Šprláková-Puková, and Marek Mechl

From the Department of Radiology and Nuclear Medicine, University Hospital Brno and Masaryk University, Czech Republic (MD, MK, JS, AŠ-P, MM); Department of Neurology, University Hospital Brno and Masaryk University, Czech Republic (JB, PP, MH, YB); and Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Brno, Czech Republic (EK)

## ABSTRACT

**BACKGROUND AND PURPOSE:** This study's aim was to investigate diffusion properties of the cervical spinal cord in patients with clinically isolated syndrome (CIS) through analysis of diffusion tensor imaging (DTI) data and thereby to assess the capacity of this technique for predicting the progression of CIS to clinically definite multiple sclerosis (CDMS).

**METHODS:** The study groups were comprised of 47 patients with CIS (15 of them with progression to CDMS within 2 years of follow-up) and 57 asymptomatic controls. All patients and controls had undergone magnetic resonance imaging (MRI) of the cervical spine including DTI and brain MRI. Methodological approaches included histogram analysis of the cervical cord's diffusion parameters and evaluation of T2 hyperintense lesions of the spinal cord and brain. All parameters were compared between the study groups. Sensitivity and specificity calculations were then performed with a view to predicting conversion to CDMS.

**RESULTS:** The patient subgroups defined by progression to CDMS differed significantly in values of fractional anisotropy (FA) kurtosis measured within white matter (WM) and normal-appearing WM (NAWM). The same parameters also differed significantly when patients with progression to CDMS were compared to healthy controls. Receiver operating characteristic (ROC) analysis revealed sensitivity and specificity of FA kurtosis of WM and NAWM of 93% and 72%, respectively, in terms of predicting CIS to CDMS progression.

**CONCLUSION:** This study presents evidence that histogram analysis of diffusion parameters of the cervical spinal cord in patients with CIS may be helpful in predicting conversion to CDMS.

**Keywords:** clinically isolated syndrome, diffusion tensor imaging, multiple sclerosis, spine.

**Acceptance:** Received May 4, 2020, and in revised form October 13, 2020. Accepted for publication October 26, 2020.

**Correspondence:** Address correspondence to Miloš Keřkovský, Department of Radiology and Nuclear Medicine, University Hospital Brno and Masaryk University, Jihlavská 340/20, 625 00 Brno, Czech Republic. E-mail: kerkovsky.milos@fnbrno.cz.

**Acknowledgments and Disclosures:** This work was supported by the Czech Health Research Council (AZV-15-32133A) and by the Ministry of Health, Czech Republic - conceptual development of research organization (FNBr, 65269705). The authors declare that there is no conflict of interest regarding the publication of this article. We would like to thank Gale A. Kirking, Editor-in-Chief at English Editorial Services, who edited the manuscript.

J Neuroimaging 2021;31:108-114.
DOI: 10.1111/jon.12808

## Introduction

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS). It is characterized histopathologically by inflammatory infiltration, demyelination, axonal loss, and gliosis in various areas of the CNS.[1] Clinically isolated syndrome (CIS) is the term used to describe the first clinical event exhibiting symptoms of either an inflammatory demyelinating attack or the initial stage of MS.[2]

Magnetic resonance imaging (MRI) is currently the key method for obtaining data used in the diagnostic algorithm for MS. Conventional MRI is based on $T_1$-weighted sequences and is able to detect hyperintense lesions of the brain and spinal cord tissues. These findings carry recognized prognostic value in terms of CIS to clinically definite MS (CDMS) conversion as based on the McDonald criteria.[3]

Other MRI imaging techniques are also available, however, such as those known to detect the structural pathology of white matter (WM) with a higher degree of sensitivity than do the more commonly used sequences. One of these is diffusion tensor imaging (DTI), which is based on the analysis of anisotropy and the directional characteristics of water diffusivity in tissue.[4] The remarkable sensitivity of this technique has been documented in a number of reports detailing cases of occult damage occurring in WM in the absence of abnormalities visible on $T_2$-weighted images (normal-appearing WM [NAWM]) in the brain.[5–6] Although DTI scans of the spine are technically more demanding and comparatively less intensively studied, several reports have drawn attention to the power of DTI in detecting demyelinating changes in the spinal cord.[7–9] However, the value of DTI of the spinal cord in terms of predicting the conversion of CIS to CDMS remains as yet unknown.

The main purpose of this prospective study, therefore, was to investigate diffusion properties of the cervical spinal cord in patients with early-stage CIS through histogram analysis of DTI data and to establish the strength of this technique in predicting the conversion of CIS to CDMS.

Table 1. Clinical and Demographic Characteristics of Patients and Healthy Controls

| | | Nonprogressed (n = 32) | Progressed (n = 15) | Controls (n = 57) |
|---|---|---|---|---|
| Age (years) | Mean (SD) | 32.9 (7.4) | 36.6 (10.4) | 34.6 (8.6) |
| | Median (MMR) | 32.5 (21-47) | 37 (20-61) | 34 (21-62) |
| Sex | Males | 13 (40.6%) | 5 (33.3%) | 20 (35.1%) |
| | Females | 19 (59.4%) | 10 (66.7%) | 37 (64.9%) |
| Treatment | Yes | 17 (53.1%) | 12 (80%) | - |
| | No | 15 (46.9%) | 3 (20%) | - |
| Clinical lesion | Visual pathway | 16 | 4 | - |
| | Motor pathway | 9 | 7 | - |
| | Sensory pathway | 5 | 3 | - |
| | Brainstem | 5 | 1 | - |
| | Spinal cord | 5 | 9 | - |

Mean (SD = standard deviation) and median (MMR = minimum-maximum range) age is shown. Absolute values (relative values) are shown in cases of sex and treatment. The categories of clinical symptoms overlap, as some of the patients displayed multiple clinical symptoms, and the presence of spinal cord symptoms was evaluated independently in all subjects. n, number of subjects in a group.

## Methods

This cross-sectional study included a group of 47 patients with CIS (18 males and 29 females) with median and minimum-maximum range (MMR) age of 34 (20-61) years and a group of 57 neurologically asymptomatic controls (20 males and 37 females) with median and MMR age of 34 (21-62) years. All patients underwent a clinical neurological examination including evaluation of expanded disability status scale (EDSS), laboratory testing of blood and cerebrospinal fluid, and such electrophysiological or other testing as designed to exclude other causes of particular episodes. CIS was diagnosed on the basis of typical clinical findings, such as sensitivity or motor disorders, vertigo, and optic neuritis (Table 1). The patients were clinically monitored at 3-month intervals to register any further clinical attacks. The exclusion criteria for healthy controls, as verified by a questionnaire, included any history of symptoms suggestive of MS lasting longer than 24 hours. Also excluded were subjects with history or suspicion of meningoencephalitis, stroke, transitory ischemic attacks, epilepsy, and systemic inflammatory diseases, as well as subjects with known blood relatives suffering from MS. Two patient subgroups were subsequently classified according to progression to CDMS emerging during a 2-year period of clinical follow-up after their initial MRI scan. These subgroups eventually consisted of 32 patients without progression (13 males and 19 females), median (MMR) age 33 (21-47) years, and 15 patients with progression to CDMS (5 males and 10 females), median (MMR) age 37 (20-61) years. The progression to CDMS was defined in accordance with the Poser criteria indicating a further clinical demyelinating disease attack.[10] Treatment with first-line disease-modifying drugs (interferon beta or glatiramer acetate) was initiated during the follow-up period in 12 (80%) of the progressed patients and in 17 (53%) of the nonprogressed patients.

All subjects signed informed consent forms in order to participate in the study, which was approved by the Ethics Committee of the University Hospital.

All participants underwent an MRI scan of the brain and cervical spine with a 1.5T MR device (Philips Achieva) using a 16-channel head and neck coil. The patient examination was performed within the first couple of months after the first recorded clinical attack (median [MMR] 7 [0-19] months). In 39 patients treated with corticosteroids, the examinations were performed after an interval of at least 3 weeks from the last such dose. The imaging protocol of the spine consisted of STIR (short-tau inversion recovery), $T_1$- and $T_2$-weighted images in the sagittal plane, $T_2^*$-weighted fast field echo (FFE) images dually covering C1-C3 and C3-C7 spinal segments in the axial plane, and DTI images implementing the single-shot echo planar technique acquired with the same geometry and coverage as that of the $T_2^*$-FFE images in the axial plane with 4-mm slice thickness. The DTI sequence magnetic gradient was applied in 15 directions with a b-value of 800 s/mm$^2$. A total of six excitations (NEX) were acquired for diffusion images and a total of three NEX for the B0 image. The imaging protocol for the brain consisted of $T_2$ and 3D $T_1$-weighted images in the transverse plane and 3D fluid-attenuated inversion recovery (FLAIR) images in the sagittal plane (Table 2).

Two experienced radiologists (Miloš Keřkovský and Jakub Stulík) evaluated by consensus the spinal cord and brain images for the presence of $T_2$, $T_2^*$, or FLAIR hyperintense lesions and classified the findings in terms of their dissemination in space (DIS) in accordance with the 2017 McDonald criteria.[9] Final decisions in doubtful cases were made by group consensus while including the other coauthors (Andrea Šprláková-Puková and Marek Mechl).

The next step in data analysis was segmentation of the spinal cord in axial $T_2^*$-weighted images using ITK-SNAP 3.4 software.[11] Segmentation of the entire spinal cord (ESC) was performed, with WM and gray matter (GM) separated, by use of the semiautomatic classification mode. This method uses machine-learning algorithms (random forest classifier and geodesic active contour) based on the training of the contextual information about intensity of neighboring voxels, and coordinates of voxels from multiple image layers were derived. Because training is done on the same subject as is segmentation, the dataset need not to be so large as those required by fully automatic segmentation algorithms, and it is a more personal approach. This method produces broad overlapping with manual segmentation.[12] Furthermore, $T_2^*$-hyperintense lesions of the spinal cord, if present, were segmented in manual mode by Miloš Keřkovský in $T_2^*$-weighted axial images. The remaining WM volume was then classified as NAWM (Fig 1).

DTI data analysis was conducted using the Functional MRI of the Brain Software Library (FSL),[13] which calculated four different maps of diffusion scalar parameters: mean diffusivity (MD), fractional anisotropy (FA), radial diffusivity (RD), and

Table 2. Magnetic Resonance Image Acquisition Protocol

| | Sequence | Orientation | Slices | TR (ms) | TE (ms) | Acq. voxel size (mm) | Recon voxel size (mm) |
|---|---|---|---|---|---|---|---|
| Spine | $T_1$ TSE | Sagittal | 11 | 400 | 7.8 | $0.9 \times 1 \times 3$ | $0.36 \times 0.36 \times 3$ |
| | $T_2$ TSE | Sagittal | 11 | 3500 | 120 | $0.75 \times 0.9 \times 3$ | $0.29 \times 0.29 \times 3$ |
| | STIR TSE | Sagittal | 11 | 5000 | 80 | $1 \times 1.12 \times 3$ | $0.37 \times 0.37 \times 3$ |
| | $T_2^*$ FFE | Transverse | $2 \times 14$ | 334 | 9.21 | $0.6 \times 0.6 \times 4$ | $0.39 \times 0.39 \times 4$ |
| | DTI | Transverse | $2 \times 14$ | 3200 | 92.3 | $1 \times 1 \times 4$ | $0.88 \times 0.88 \times 4$ |
| Brain | $T_2$ TSE | Transverse | 22 | 4850 | 110 | $0.9 \times 1.1 \times 5$ | $0.9 \times 0.9 \times 5$ |
| | FLAIR 3D | sagittal | 257 | 8000 | 275 | $1.4 \times 1.2 \times 1.2$ | $0.7 \times 0.63 \times 0.63$ |

Parameters of MRI sequences.
Abbreviations: Acq., acquisition; DTI, diffusion tensor imaging; FFE, fast field echo; FLAIR, fluid-attenuated inversion recovery; Recon, reconstruction; STIR, short tau inversion recovery; TE, time to echo; TR, repetition time; TSE, turbo spin echo.

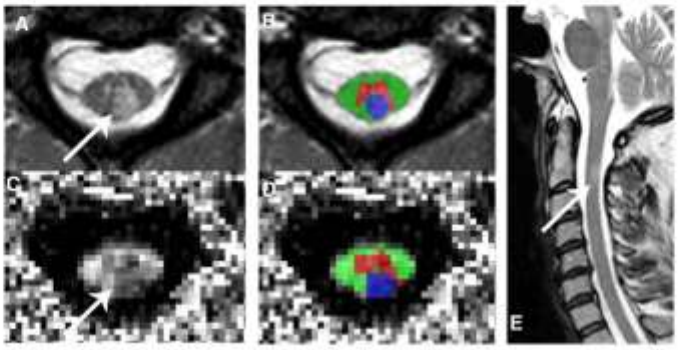



**Fig 1.** MR examination in a patient with hyperintense spinal cord plaque apparent on the axial $T_2^*$-weighted image (A) of gradient echo (arrows in panels A, C, and E) and on the sagittal $T_2$-weighted image of fast spin echo (E). The lesion is also perceptible in the diffusion tensor image in the form of reduced signal intensity on the fractional anisotropy map (C). Demonstration of semiautomatic white matter (green), gray matter (red), and $T_2^*$-hyperintense lesion (blue) segmentation using the ITK-SNAP application (B and D).

axial diffusivity (AD). Segmentation of ESC was also performed in DTI (FA) images using the same technique as described above. $T_2^*$ and DTI volumes were subsequently registered using the contours of the segmented spinal cord as landmarks. The technique of segmentation and subsequent DTI data analysis has been further described and evaluated in a recent study.[12]

Various tissue masks (ESC, WM, GM, NAWM, and $T_2^*$ hyperintense lesions) rendered at $T_2^*$-weighted axial images were applied across the registered DTI images. Histograms of the aforementioned DTI parameters were determined in all subjects as a set of values measured in all voxels constituting the segmented volumes. Further statistical analyses were conducted by taking into consideration parameters characterizing those histograms of the various DTI scalars: mean; median; minimum; maximum; 5th, 25th, 75th, and 95th quantiles; standard deviation (SD); skewness (Eq. 1); and kurtosis (Eq. 2).

$$\frac{n}{(n-1)(n-2)} \sum \frac{(X_i - \bar{X})^3}{s^3} \quad (1)$$

$$\left\{ \frac{n(n+1)}{(n-1)(n-2(n-3)} \sum \frac{(X_i - \bar{X})^4}{s^4} \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (2)$$

where $n$ is the sample size, $X_i$ is the $i$th X value, $\bar{X}$ is the mean, and $s$ is the sample SD.[14]

A separate analysis (comprising 44 statistical tests given by combination of the four maps of diffusion scalar parameters with the 11 histogram parameters) was performed for each differing tissue sample: ESC, WM, NAWM, and GM. Comparisons of the measured parameters between subgroups of patients defined by progression to CDMS as well as between each patient subgroup and controls were performed using Mann-Whitney U test. Although age and gender were comparable among all subject groups based on Mann-Whitney U test and Fisher's exact test, respectively, the effects of these two potential confounding factors were eliminated from the data prior to statistical testing using linear regression in order to avoid biased results. Furthermore, a receiver operating characteristic (ROC) analysis was conducted in order to evaluate the sensitivity and specificity of significantly differing diffusion parameters between the patient subgroups. The subjects' quantitative characteristics (such as age) and EDSS values were described as mean with SD and/or median with MMR and were compared between the groups using Mann-Whitney U test. The subjects' categorical characteristics (such as sex, treatment, incidence of clinical spinal cord symptoms, presence of spinal cord lesions, or McDonald DIS criteria) were summarized using absolute count and percentage and tested between groups using Fisher's exact test. In addition, sensitivity and specificity of McDonald DIS criteria, and particularly of the recorded presence of spinal cord lesion(s), were calculated on the basis of predicted progression to CDMS. The significance level for all statistical tests was set at $P < .05$. In cases of comparing diffusion parameters between the groups, false discovery rate (FDR) correction for

Table 3. Fractional Anisotropy Kurtosis

| | Median (MMR) Fractional Anisotropy kurtosis | | |
|---|---|---|---|
| Tissue | Controls (n = 57) | Nonprogressed (n = 32) | Progressed (n = 15) |
| WM | 0.18 (−0.13-1.82) | 0.26 (−0.23-0.68) | 0.01 (−0.17-0.37) |
| NAWM | 0.18 (−0.13-1.82) | 0.26 (−0.23-0.71) | 0.02 (−0.15-0.34) |

Median (MMR = minimum–maximum range) values of fraction anisotropy kurtosis in white matter (WM) and normal-appearing white matter (NAWM) of all groups. Differences of progressed patients from nonprogressed patients as well as controls were statistically significant even after the false discovery rate correction for multiple comparisons.
Abbreviation: n, number of subjects in group.

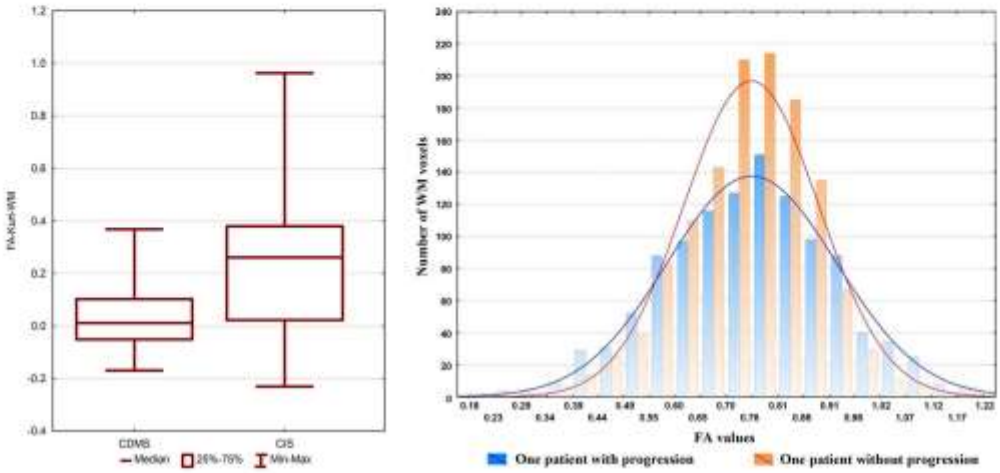



**Fig 2.** (A) Box plot of the kurtosis values of fractional anisotropy histograms measured within white matter. These values demonstrate the comparison between patients with progression to clinically definite multiple sclerosis (CDMS), at left, and those without progression (clinically isolated syndrome [CIS]) at right. (B) Histograms of fractional anisotropy (FA) values measured in the white matter (WM) of the spinal cord in two patients selected to demonstrate the differences in kurtosis (Kurt). The histogram belonging to the patient without progression to CDMS (orange) appears to present a higher kurtosis level (0.71) compared to the kurtosis level (−0.13) of the patient in whom the conversion to CDMS was noted after an 8-month period following the initial MRI scan (blue).

multiple comparisons was further applied for each tissue sample. The statistical analyses were performed using IBM SPSS Statistics 25, R 3.4.1, and Statistica 12 (StatSoft).

## Results

No significant differences were observed with regard to the age (all $P > .23$) and sex (all $P > .65$) of the subjects between the patient subgroups and the control group. The discrepancy in the number of treated patients across the two subgroups, mentioned in the Methods section, was not statistically significant ($P = .11$). The median (MMR) interval between the MRI scan and clinical progression to CDMS in the subgroup of 15 CIS patients who developed CDMS was seven (0-19) months.

Mean (SD) and median (MMR) baseline EDSS scores in patients without progression and in patients with recorded progression to CDMS were 1.73 (1.17); 2 (0-5) and 2.33 (0.70); 2 (1-3.5), respectively. This difference was found to be statistically significant ($P = 0.026$). The incidence of clinical spinal cord symptoms (Table 1) differed significantly between the patient subgroups defined by their progression to CDMS ($P = .006$).

Comparisons drawn between healthy controls and nonprogressed patients revealed no statistically significant differences between those groups after FDR correction.

The patient subgroup with progression to CDMS differed significantly from the control group after FDR correction in kurtosis of FA measured within both WM and NAWM (both $P < .0005$). Similarly, kurtosis of FA within WM and NAWM was significantly different after FDR correction when comparisons were made within the patient subgroups defined by their progression to CDMS (both $P < .001$) (Table 3, Fig 2A, B). The remaining parameters did not demonstrate significant differences.

ROC analysis revealed that a histogram showing kurtosis of FA values appeared as a powerful parameter for distinguishing patients with progression to CDMS, reaching sensitivity of 93% and specificity of 72% with cutoff 0.14 when measured within both WM and NAWM. The areas under curve (confidence intervals) of FA kurtosis for NAWM and WM were 0.798 (0.667-0.929) and 0.796 (0.661-0.930), respectively.

In the nonprogressed patient subgroup, $T_2^*$-hyperintense spinal cord lesions were found in 25% of subjects, while in

Table 4. $T_2$ Hyperintense Lesions in Nonprogressed and Progressed Patients

| Parameter | | Nonprogressed (n = 32) | Progressed (n = 15) | *P*-value |
|---|---|---|---|---|
| SC lesions | Yes | 8 (25.0%) | 9 (60.0%) | .027 |
| | No | 24 (75.0%) | 6 (40.0%) | |
| DIS | Yes | 15 (46.9%) | 12 (80.0%) | .056 |
| | No | 17 (53.1%) | 3 (20.0%) | |

Parameters of the patient subgroups defined by presence or absence of progression to clinically definite multiple sclerosis in absolute counts (relative values).
Abbreviations: DIS, McDonald dissemination in space criteria evaluating both spinal cord and brain images; n, number of subjects in group; SC lesions, presence of $T_2^*$ hyperintense spinal cord lesion(s).

the progressed patient subgroup they were present in 60% of subjects. This difference was statistically significant ($P = .027$) (Table 4).

The presence of spinal cord $T_2^*$-hyperintense lesion(s) predicted conversion to CDMS at sensitivity and specificity of 60% and 75%, respectively. The McDonald criteria also integrated into the evaluation of brain scans demonstrated sensitivity of 80% and specificity of 53% in terms of conversion prediction.

## Discussion

The aim of this study was to investigate the potential for quantitative analysis of diffusion scalar parameters histograms pertaining to the cervical spinal cord in order to predict potential conversion to CDMS in early-stage CIS patients. From a clinical perspective, the identification of high-risk patients with CIS present is a matter of high importance. Several other research groups have attempted to establish biomarkers that would indicate the clinical conversion based on conventional brain MRI scans while using different analytical methods. Wottschel et al achieved relatively high sensitivity and specificity (77% and 66%, respectively) when predicting conversion in CIS-diagnosed patients by analyzing 11 different predictors derived from proton density and $T_2$-weighted images while using advanced machine-learning algorithms.[15]

It has previously been demonstrated that quantitative histogram analysis of diffusion parameters, as has been used in this study, may provide more information in comparison to conventional analysis that evaluates only mean or median values.[16] Histogram analysis of DTI data was also used by Yu et al to evaluate a case of occult damage to normal-appearing white and gray brain matter in patients with CIS and CDMS.[5]

Considering that the presence of spinal cord lesions that are detectable using a conventional MRI scan can serve as an independent predictor of CDMS development in patients with CIS,[17] it is a logical step to focus on the ultrastructural pathology of the cervical spinal cord in CIS-diagnosed patients. Nonetheless, considerably fewer authors have focused on diffusion MRI of the spinal cord in patients with CIS or CDMS. A review by Martin et al reports a total of 69 studies actively addressing DTI of the spinal cord, but only a very low number of these are directed to patients diagnosed with CIS or MS.[18] All present somewhat varying results. Certain authors using ROI-based analysis appear to have demonstrated differences in the RD values recorded in the spinal cord of MS-diagnosed patients when compared to healthy controls without significant changes in mean FA values.[19] Toosy et al have revealed several voxels pertaining to the upper segments of the cervical spinal cord with significantly differing FA and RD values in MS patients compared to healthy controls.[20] Valsasina et al have used histogram analysis of FA and MD in MS-diagnosed patients and revealed significant decrease in mean FA compared to healthy controls.[9] So far as we are able to establish, however, there exists no previous study aiming to investigate the predictive power of a DTI scan of the spinal cord in relation to CIS to CDMS conversion.

This study has demonstrated significant differences in kurtosis of FA measured in WM and NAWM when comparing the subgroup of patients with progression to CDMS to the subgroup of progression-free patients and to the group of healthy controls. Changes in anisotropic diffusion may generally reflect distinct ultrastructural tissue changes constituting nerve tract disintegration.[21]

However, the mean and median FA values did not differ significantly between those study groups investigated. This may be explained in part by heterogeneity of the diffusivity changes present in the spinal cord tissue samples, and minor abnormalities located within a certain spinal segment may have been missed by averaging measurements across the whole cervical cord volume. These may, however, be detected with a higher degree of accuracy by other histogram parameters reflecting their distribution with greater precision.

Based on the results presented herein, it appears that normal spinal cord tissue is characterized by the predominant peak of what we may term "ideal" medium FA values, complemented by relatively low numbers of "extreme" voxels related possibly to the effect of crossing fibers or other features of the neural tissue structure's physiological complexity, susceptibility artefacts, and other influences. This results in high measured kurtosis values. Contrarily, in pathological spinal cord tissue, which is assumed to be analyzed in patients with progression to CDMS, this middle peak of what we may term "normal" values seems to be reduced, and relatively more frequently occurring voxels with lower or higher values may more likely be related to the real structural disintegration of tissue (Fig 2B). Linking the alterations of FA kurtosis specifically with the ultrastructural changes of the spinal cord in CIS or MS patients is nevertheless not straightforward and, so far as we are able to establish, no specific explanation can be found for this in previous studies. Further investigation (e.g., by studies using animal models) is therefore required.

It has previously been proven that the kurtosis parameter is dependent on sample size, however, and the variance of kurtosis is small when the sample size is greater than 750.[14] In this paper, mean (SD) sample size (voxel count) of WM and NAWM was 1475.10 (252.06) and 1456.97 (266.44), respectively. The histograms may therefore be considered sufficiently robust to enable reliable analysis of kurtosis. Some authors have discussed the theoretical basis of skewness and kurtosis and questioned the reliability and usability of those parameters.[14,22] On

the other hand, several recently published studies have used these metrics (e.g., for tumor differentiation) and, upon evaluating their repeatability and correlations with clinical findings, have reported promising results.[23–25] Based on these findings, kurtosis may be regarded as a reliable biomarker with certain clinical potential if the analyzed data set is sufficiently large.

The diagnostic power of FA kurtosis in terms of CIS progression to CDMS was very similar when measured within WM and NAWM, reaching sensitivity of 93% and specificity of 72% for both tissues. Thus, FA kurtosis of WM, in particular, may be an interesting biomarker of later clinical progression, and one whose calculation does not require manual segmentation of $T_2^*$ hyperintense lesions. We similarly attempted to find some significant parameters also using ESC volume analysis, which would permit dispensing with separate segmentation of WM and GM. Unfortunately, we did not discover such an ESC parameter and so GM and WM segmentation remains a crucial step within the analysis.

The presence of spinal cord $T_2^*$-hyperintense lesions was of relatively lower sensitivity (60%) in terms of clinical conversion to CDMS compared with the diffusion parameters. McDonald criteria (DIS) in this study demonstrated high sensitivity (80%) but comparatively lower specificity (53%). Similar data verifying the predictive power of the McDonald criteria have been presented in other studies. Hyun et al report a sensitivity and specificity of 88% and 43%, respectively.[26] Another recent study by Filippi et al reports sensitivity of 91% and specificity of 33% in terms of predicting CIS progression to CDMS.[27]

Based on the aforementioned facts, it would appear that analysis of diffusion data of the cervical spinal cord can be superior to conventional MRI techniques. Addition of brain DTI analysis might yield results even more robust.

This study had a number of limitations. Among these was use of the 1.5T MRI device, which may, when compared to 3T systems, provide poorer signal-to-noise ratios. On the other hand, a lower static magnetic field may reduce susceptibility artefacts and distortions that present fundamental problems in spinal cord diffusion MRI. Also, a relatively low number of subjects were included in this single-center study. Although the number was high enough to provide statistically significant results, result verification using larger subject groups would be beneficial. Furthermore, a 2-year follow-up period is a relatively short time frame within which to evaluate the potential for CDMS conversion. As shown by long-term studies, however, while progression to CDMS may occur after an interval of many years, it also happens to occur most frequently during the first few years after the first clinical symptoms emerge.[28] With this in mind, a 2-year monitoring period was deemed sufficient when taking into consideration the previous notion that relapsing activity during the initial 1-2 years of the disease's course in patients with MS carries crucial prognostic importance in anticipating the severity level of future damage.[29] A certain bias may also have arisen out of the effects of treatments initiated during the follow-up period in patients belonging to both subgroups. Although the numbers of treated patients were not significantly different between the patient subgroups as defined by their progression to CDMS, the rate among treated patients does appear somewhat unequal and may prove significant in the case of a larger data set. Although this fact could to some extent influence the measured diffusion parameters, any observation of this disease's natural progression, while a theoretically optimal methodology, is unacceptable from an ethical standpoint.

In conclusion, this study presents evidence that diffusion parameter histogram analysis of the cervical cord in patients with early stage CIS may be helpful in predicting the conversion of CIS to CDMS during the initial 2-year period.

## References


1. Confavreux C, Vukusic S, Moreau T, et al. Relapses and progression of disability in multiple sclerosis. N Engl J Med 2000;343:1430-8.
2. Miller DH, Chard DT, Ciccarelli O. Clinically isolated syndromes. Lancet Neurol 2012;11:157-69.
3. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018;17:162-73.
4. Ranzenberger LR, Snyder T. *Diffusion Tensor Imaging*. Treasure Island, FL: StatPearls Publishing; 2019.
5. Klistorner A, Wang C, Yiannikas C, et al. Diffusivity in the core of chronic multiple sclerosis lesions. PLoS One 2018;13:e0194142
6. Yu CS, Lin FC, Liu Y, et al. Histogram analysis of diffusion measures in clinically isolated syndromes and relapsing-remitting multiple sclerosis. Eur J Radiol 2008;68:328-34.
7. Hesseltine SM, Law M, Babb J, et al. Diffusion tensor imaging in multiple sclerosis: assessment of regional differences in the axial plane within normal-appearing cervical spinal cord. AJNR Am J Neuroradiol 2006;27:1189-93.
8. Tanenbaum LN. Clinical applications of diffusion imaging in the spine. Magn Reson Imaging Clin N Am 2013;21:299-320.
9. Valsasina P, Rocca MA, Agosta F, et al. Mean diffusivity and fractional anisotropy histogram analysis of the cervical cord in MS patients. Neuroimage 2005;26:822-28.
10. Poser CM, Paty DW, Scheinberg L, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 1983;13:227-31.
11. Yushkevich PA, Gao Y, Gerig G. ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images. Annu Int Conf IEEE Eng Med Biol Soc 2016;2016:3342-5.
12. Dostál M, Keřkovský M, Koriťáková E, et al. Analysis of diffusion tensor measurements of the human cervical spinal cord based on semiautomatic segmentation of the white and gray matter. J Magn Reson Imaging 2018;48:1217–27.
13. Jenkinson M, Beckmann CF, Behrens TEJ, et al. FSL. Neuroimage 2012;62:782-90.
14. McNeese B. Are the skewness and kurtosis useful statistics? 2020. www.spcforexcel.com/knowledge/basic-statistics/are-skewness-and-kurtosis-useful-statistics. Accessed July 21, 2020.
15. Wottschel V, Alexander DC, Kwok PP, et al. Predicting outcome in clinically isolated syndrome using machine learning. Neuroimage Clin 2015;7:281-7.
16. Wagner MW, Narayan AK, Bosemani T, et al. Histogram analysis of diffusion tensor imaging parameters in pediatric cerebellar tumors. J Neuroimaging 2016;26:360-5.
17. Patrucco L, Rojas JI, Cristiano E. Assessing the value of spinal cord lesions in predicting development of multiple sclerosis in patients with clinically isolated syndromes. J Neurol 2012;259:1317-20.
18. Martin AR, Aleksanderek I, Cohen-Adad J, et al. Translating state-of-the-art spinal cord MRI techniques to clinical use: a systematic review of clinical studies utilizing DTI, MT, MWF, MRS, and fMRI. Neuroimage Clin 2016;10:192-238.
19. Oh J, Sotirchos ES, Saidha S, et al. Relationships between quantitative spinal cord MRI and retinal layers in multiple sclerosis. Neurology 2015;84:720-8.
20. Toosy AT, Kou N, Altmann D, et al. Voxel-based cervical spinal cord mapping of diffusion abnormalities in MS-related myelitis. Neurology 2014;83:1321-5.
21. Mottershead JP, Schmierer K, Clemence M, et al. High field MRI correlates of myelincontent and axonal density in multiple sclerosis. J Neurol 2003;250:1293-301.

22. Westfall PH. Kurtosis as peakedness, 1905-2014. R.I.P. Am Stat 2014;68:191-5.
23. Borguezan BM, Lopes AJ, Saito EH, et al. Solid indeterminate nodules with a radiological stability suggesting benignity: a texture analysis of computed tomography images based on the kurtosis and skewness of the nodule volume density histogram. Pulm Med 2019;2019:4071762.
24. Nagano H, Sakai K, Tazoe J, et al. Whole-tumor histogram analysis of DWI and QSI for differentiating between meningioma and schwannoma: a pilot study. Jpn J Radiol 2019;37:694-700.
25. Xiao Z, Tang Z, Zhang J, et al. Whole-tumor histogram analysis of monoexponential and advanced diffusion-weighted imaging for sinonasal malignant tumors: correlations with histopathologic features. J Magn Reson Imaging 2020;51:273-85.
26. Hyun J-W, Kim W, Huh S-Y, et al. Application of the 2017 McDonald diagnostic criteria for multiple sclerosis in Korean patients with clinically isolated syndrome. Mult Scler 2019;25:1488-95.
27. Filippi M, Preziosa P, Meani A, et al. Prediction of a multiple sclerosis diagnosis in patients with clinically isolated syndrome using the 2016 MAGNIMS and 2010 McDonald criteria: a retrospective study. The Lancet Neurology 2018;17:133-42.
28. Hely MA, McManis PG, Doran TJ, et al. Acute optic neuritis: a prospective study of risk factors for multiple sclerosis. J Neurol Neurosurg Psychiatry 1986;49:1125-30.
29. Weinshenker BG, Bass B, Rice GP, et al. The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. Brain 1989;112:1419-28.

Original Investigation

# Evaluating Magnetic Resonance Diffusion Properties Together with Brain Volumetry May Predict Progression to Multiple Sclerosis

Jakub Stulík, MD, Miloš Keřkovský, MD, PhD, Matyáš Kuhn, MSc, Monika Svobodová, MD, Yvonne Benešová, MD, PhD, Josef Bednařík, MD, PhD, Andrea Šprláková-Puková, MD, PhD, Marek Mechl, MD, PhD, MBA, Marek Dostál, MSc, PhD

**Rationale and Objectives:** Although the gold standard in predicting future progression from clinically isolated syndrome (CIS) to clinically definite multiple sclerosis (CDMS) consists in the McDonald criteria, efforts are being made to employ various advanced MRI techniques for predicting clinical progression. This study's main aim was to evaluate the predictive power of diffusion tensor imaging (DTI) of the brain and brain volumetry to distinguish between patients having CIS with future progression to CDMS from those without progression during the following 2 years and to compare those parameters with conventional MRI evaluation.

**Materials and Methods:** All participants underwent an MRI scan of the brain. DTI and volumetric data were processed and various parameters were compared between the study groups.

**Results:** We found significant differences between the subgroups of patients differing by future progression to CDMS in most of those DTI and volumetric parameters measured. Fractional anisotropy of water diffusion proved to be the strongest predictor of clinical conversion among all parameters evaluated, demonstrating also higher specificity compared to evaluation of conventional MRI images according to McDonald criteria.

**Conclusion:** Conclusion: Our results provide evidence that the evaluation of DTI parameters together with brain volumetry in patients with early-stage CIS may be useful in predicting conversion to CDMS within the following 2 years of the disease course.

**Key Words:** Brain volumetry; Clinically isolated syndrome; Diffusion tensor imaging; Magnetic resonance imaging; Multiple sclerosis.



## INTRODUCTION

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system arising from inflammatory infiltration and causing demyelination of axons and their subsequent destruction in various areas of the central nervous system (1). Clinically isolated syndrome (CIS) is a monophasic clinical episode with patient-reported symptoms and objective findings reflecting a focal or multifocal inflammatory demyelinating event in the central nervous system, developing acutely or subacutely and lasting at least 24 hour. There may be one or multiple affected areas (2). CIS patients may or may not progress into clinically definite multiple sclerosis (CDMS).

The gold standard in predicting future progression from CIS to multiple sclerosis consists is the McDonald criteria (2), which combine clinical findings, laboratory results, and findings on MRI examinations of the brain and spinal cord. The MRI criteria evaluate dissemination in space, which means to assess the number and distribution of hyperintense lesions in the white matter visible in T2-weighted or fluid attenuated inversion recovery (FLAIR) images, and dissemination in time, which means the presence of new lesions or lesions

**Acad Radiol 2022; 29:1493–1501**

From the Department of Radiology and Nuclear Medicine, University Hospital Brno, Jihlavská 20 Brno, 62500, Czech Republic (J.S., M.K., A.Š.-P., M.M., M.D.); Department of Neurology, University Hospital Brno, Brno, Czech Republic (M.S., Y.B., J.B.); Department of Biophysics, Masaryk University, Brno, Czech Republic (M.D.); Department of Psychiatry, University Hospital Brno, Brno, Czech Republic (M.K.); Behavioural and Social Neuroscience, CEITEC Masaryk University, Brno, Czech Republic (M.K.); Institute of Biostatistics and Analyses, Masaryk University, Czech Republic (M.K.); Faculty of Medicine, Masaryk University, Brno, Czech Republic (J.S., M.K., M.S., Y.B., J.B., A.Š.-P., M.M.). Received September 15, 2021; revised November 29, 2021; accepted December 11, 2021. **Address correspondence to:** M.K. e-mail: keskovsky.milos@fnbrno.cz


https://doi.org/10.1016/j.acra.2021.12.015

enhancing after gadolinium contrast administration. Due to the evolving capabilities of advanced MRI techniques, an effort is being made to use them to refine the prediction of clinical progression in patients with CIS to definitive multiple sclerosis. The main aim of this prospective single-centre study was to evaluate the predictive power of diffusion MRI of the brain and brain volumetry to distinguish patients with CIS with future progression to CDMS from those without progression and compare those parameters with evaluation of conventional MRI.

## MATERIALS AND METHODS

The study was approved by the ethics committee of the university hospital, and all subjects signed informed consent forms in order to participate in the study. Two groups of subjects were included in the prospective study. The first group comprised 72 patients with CIS, who were further divided into two subgroups: 50 patients without clinical progression to CDMS and 22 patients with progression during the observed 2-year period since the initial MRI examination. The second group consisted of 62 community-recruited healthy controls. More detailed information about the study participants is shown in (Tables 1 and 2).

All patients underwent a clinical neurological examination including evaluation of expanded disability status scale (EDSS), laboratory testing of blood and cerebrospinal fluid, and such electrophysiological or other testing as designed to exclude other causes of particular episodes. CIS was diagnosed on the basis of typical clinical findings, such as sensitivity or motor disorders, vertigo, optic neuritis, etc. The progression to CDMS was defined in accordance with the Poser criteria indicating a further clinical demyelinating disease attack (3). The patients were clinically monitored at 3-month intervals to register any further clinical attacks. The 2 patient subgroups were subsequently classified according to progression to CDMS emerging during a 2-year period of clinical follow-up after their initial MRI scan. Treatment with first-line disease-modifying drugs (interferon beta or glatiramer acetate) was initiated during the follow-up period in 24 of the non-progressed patients and in 16 of the progressed patients (Table 2). The exclusion criteria for healthy controls, as verified by a questionnaire, included any history of symptoms suggestive of MS. Also excluded were subjects with history or suspicion of meningoencephalitis, stroke, transitory ischemic attacks, epilepsy, and systemic inflammatory diseases, as well as subjects with known blood relatives suffering from MS and subjects with significant load of incidentally

**TABLE 1. Characteristics of Patient and Control Groups**

| | | Healthy Volunteers ($n$ = 62) | Patients ($n$ = 72) | $p$ |
|---|---|---|---|---|
| Sex | Women | 45 (72.6%) | 48 (66.7%) | 0.573 |
| | Men | 17 (27.4%) | 24 (33.3%) | |
| Age (ys) | Mean ± SD | 33.3 ± 7.4 | 34.1 ± 8.4 | 0.779 |
| | Median (min–max) | 32.5 (19.9–51.3) | 33.1 (19.9–61.3) | |
| MRI device | MR1 | 47 (75.8%) | 64 (88.9%) | 0.065 |
| | MR2 | 15 (24.2%) | 8 (11.1%) | |

Demographic characteristics of study participants and number of examinations performed on two different MRI devices (MR1 and MR2). Statistical significance of differences between groups is given by $p$-values calculated using Fisher's exact test or Mann–Whitney $U$ test, as applicable.

**TABLE 2. Characteristics of Patients Subgroups**

| | | Patients without progression ($n$ = 50) | Patients with progression ($n$ = 22) | $p$ |
|---|---|---|---|---|
| Sex | Women | 32 (64.0%) | 16 (72.7%) | 0.591 |
| | Men | 18 (36.0%) | 6 (27.3%) | |
| Age (ys) | Mean ± SD | 33.7 ± 7.8 | 34.9 ± 9.9 | 0.774 |
| | Median (min–max) | 32.9 (21.3–50.8) | 34.4 (19.9–61.3) | |
| Treatment | Yes | 24 (48.0%) | 16 (72.7%) | 0.072 |
| | No | 26 (52.0%) | 6 (27.3%) | |
| EDSS | Mean ± SD | 1,7 ± 1,2 | 2,2 ± 0,8 | 0,063 |
| | Median (min–max) | 2 (0-5) | 2 (1-3,5) | |
| MRI device | MR1 | 48 (96%) | 16 (72.7%) | 0.008 |
| | MR2 | 2 (4%) | 6 (27.3%) | |

Demographic characteristics of patients classified according to their later clinical progression within 2 years of observation. Numbers of examinations performed on two different MRI devices (MR1 and MR2) are also shown as well as numbers of treated patients within both subgroups. Expanded disability status scale (EDSS) at the first MR examination. Statistical significance of differences between groups is given by $p$-values calculated using Fisher's exact test or Mann–Whitney $U$ test, as applicable.

found T2 hyperintensities fulfilling the dissemination in space (DIS) criteria according to the 2017 McDonald criteria.

All participants underwent an MRI scan of the brain using one of two 1.5T MR devices (an older Philips Achieva device or a newer Philips Ingenia device) using a 16-channel head and neck coil. Detailed proportions of examinations performed on each scanner among all study groups are shown in (Tables 1 and 2). Patient examinations were performed within the first couple of months after the first recorded clinical attack (median [minimum−maximum] 2 [0 −8] months). In patients treated with corticosteroids, the examinations were performed after an interval of at least 3 weeks from the last such dose. The protocol comprised sequences for structural imaging and subsequent volumetric analyses (T2, FLAIR 3D, and T1 3D) and diffusion tensor imaging (DTI) sequence. Details about the imaging protocol are shown in (Table 3). DTI was acquired with b factor 0 and 1000s/mm$^2$ using 32 directions of the magnetic gradient. Two experienced radiologists (MK and JS) evaluated by consensus the images for the presence of T2/FLAIR hyperintense lesions and classified the findings in terms of the DIS in accordance with the 2017 McDonald criteria. Final decisions in doubtful cases were made by group consensus while including the other co-authors (AŠ and MM).

DTI data processing was done using FMRIB's Software Library (FSL) (4), starting with brain extraction, eddy current, and movement correction and calculation of maps of the scalar diffusion parameters (fractional anisotropy [FA], mean diffusivity [MD], axial diffusivity [AD], and radial diffusivity [RD]). The data were subsequently analyzed by voxel-wise tract-based spatial statistics (TBSS) (13) and its statistical module (Randomize) to compare the diffusion parameters between the study groups with multiple comparisons correction and subjects' age, sex, and MR device set as covariates. Mean values of the aforementioned scalar parameters were extracted from the sum of white matter voxels differing significantly ($p < 0.05$) between those groups being compared, and those values were then entered into further statistical analyses.

The total brain volume and volumes of white matter (WM) and grey matter (GM), normalized for subject head size, were estimated with SienaX (6), which is part of FSL (7), using T1-weighted 3D images. Normalization was done by SienaX scaling factor. Pre-processing included also automatic lesion segmentation and lesion filling procedure using the lesion growth algorithm (5,8) as implemented in the LST toolbox version 3.0.0 (www.statistical-modelling.de/lst.html) for statistical parametric mapping.

For overall evaluation of brain abnormalities in patients with CIS, all DTI and volumetric parameters were compared between the whole group of patients and healthy controls. To evaluate the predictive value of those parameters, we mutually compared the data of individual subgroups of patients defined according to progression to CDMS using Mann−Whitney *U* test. The subjects' age was described as mean with SD and/or median with minimum−maximum values and was compared between the groups using Mann−Whitney *U* test. The subjects' categorical characteristics (such as sex, treatment, or scanner type) were tested between groups using Fisher's exact test. Effects of potential confounding factors (age and sex of the subjects and MRI scanner type) were eliminated from the data prior to statistical testing using linear regression in order to avoid biased results. Similarly, these parameters were set as covariates for voxel-based analysis of diffusion parameters using Randomize. Furthermore, a receiver operating characteristic (ROC) analysis was conducted in order to evaluate the sensitivity and specificity of significantly differing diffusion and volumetric parameters between the patient subgroups.

To evaluate the possible influence of the two different MRI scanners on the power of DTI and volumetric parameters to predict clinical progression, we included also separate analysis of those modalities evaluating only the group of 64 patients (41 women, 23 men, mean age 34.7 years), who were examined solely on the MR1 device (more details about the group are shown in Table 2S in the supplementary material).

In addition, sensitivity and specificity of McDonald DIS criteria were calculated on the basis of predicted progression to CDMS. The significance level for all statistical tests was set at $p < 0.05$. Statistical analyses were performed using IBM SPSS Statistics 25, R 3.4.1, and Statistica 12 (StatSoft).

**TABLE 3. Parameters of Magnetic Resonance Imaging Protocol**

| Sequence | Orientation | TR (ms) | TE (ms) | Acquisition Voxel Size (mm) |
|---|---|---|---|---|
| T2 TSE | transverse | 4851 | 110 | 0.9 × 1.12 × 5 |
| FLAIR 3D | sagittal | 8000 | 275 | 1.2 × 1.2 × 1.4 |
| T1 3D FFE | transverse | 25 | 4.1 | 0.9 × 0.9 × 1.6 |
| DTI | transverse | 21000 | 62 | 2 × 2 × 2 |

Parameters of magnetic resonance imaging protocol. TSE, turbo spin echo; TR, repetition time; TE, echo time; FLAIR, fluid attenuation inversion recovery; FFE, fast field echo; DTI, diffusion tensor imaging.

## RESULTS

There were no statistically significant differences in the age or sex of the subjects between patients and controls or between the patient subgroups. Employment of the two MRI scanners yielded significantly different outcomes between the two patient subgroups (Tables 1 and 2). The differences in number of treated patients were not statistically significant between the subgroups defined by progression to CDMS (Table 2). The median (min−max) interval between the MRI scan and clinical progression to CDMS in the subgroup of 22 CIS patients who developed CDMS was 11.5 (0 −19) months.

TBSS analysis of the diffusion data identified extensive areas within the brain WM differing significantly in FA, MD,

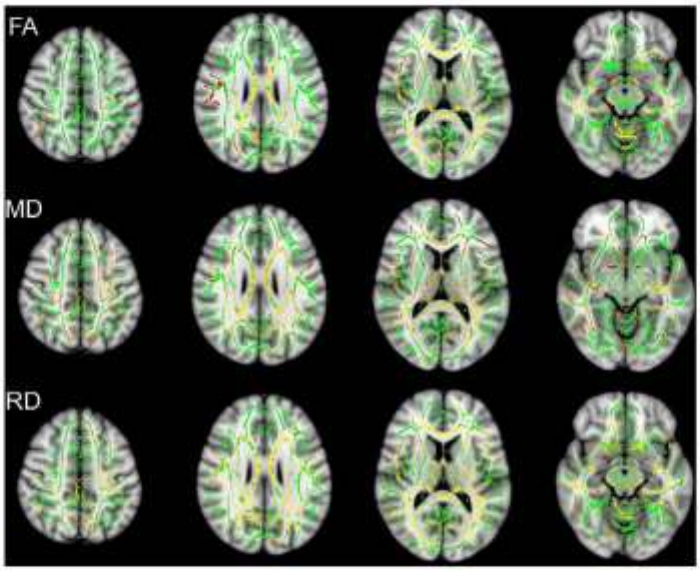



**Figure 1.** Graphic representation of the results of tract-based spatial statistics (TBSS) analyses. The selected brain images show red–yellow areas where significant differences were found between subgroups of CIS patients with and without progression to CDMS in the parameters of fractional anisotropy (FA), mean diffusivity (MD), and radial diffusivity (RA). The skeleton of white matter tracts used in data processing is marked in green.(Color version of figure is available online).

and RD between the subgroups of CIS patients as defined by their later progression to CDMS (Fig 1). Several areas within the brain also differed significantly in FA, MD, and AD values when all patients were compared with controls. The differences in FA and MD were apparently more widespread across the brain tissues in patients with progression compared to the progression-free subgroup than in all patients compared to healthy controls (Fig 2). The number of voxels (percentage of all significant voxels) with significantly different FA and MD values overlapping between the two TBSS analyses came to 9872 (19.63%) and 5202 (9.59%) voxels, respectively. The median values of the diffusion parameters extracted from the significant voxels (based on TBSS analyses) differed significantly between the groups also according to Mann–Whitney *U* test (Table 4). ROC analysis of the diffusion parameters in terms of predicting progression from CIS to CDMS revealed FA as the strongest predictor (sensitivity 77.3%, specificity 90%) against MD (sensitivity 63.6%, specificity 78%) and RD (sensitivity 63.6%, specificity 86%) (Table 5, Fig 3).

The analyses of the DTI data of the group of 64 patients excluding those examined on the MR2 device revealed similar results in terms of prediction of the clinical conversion (FA: sensitivity 68.8% and specificity 93.7%, MD: sensitivity 68.8% and specificity 87.5%, RD: sensitivity 68.8% and specificity 85.4%). More detailed results are shown in tables 4S and 5S in the supplementary material.

By conventional evaluation of T2-w and FLAIR images, we identified 18 (81.8%) subjects among the progressed patients who met DIS criteria, while in the group of non-progressed patients DIS criteria were met in 26 (52%) subjects. Thus, the sensitivity of DIS in terms of predicting clinical progression was 81.8% and the specificity 48%.

All measured volumetric parameters (whole brain, WM, and GM) differed significantly between patients and controls, revealing generally lower volumes in patients (Table 6). Similarly, the volume of brain WM and whole brain volume were significantly reduced in patients with progression compared to non-progressed patients, but the volume of GM did not differ significantly between these subgroups (Table 6). In the subsequent ROC analysis, the volume of WM was able to predict the conversion of CIS to CDMS with sensitivity 90.9% and specificity 58.0% (Table 5). Very similar results were obtained when only the 64 patients examined on MR1 were included in the analysis of the WM volume parameter to differentiate between CIS and CDMS subgroups (revealing sensitivity 87.5% and specificity 56.2%). More detailed results are shown in tables 5S and 6S in the supplementary material.

## DISCUSSION

This study's aim was to explore the potential of diffusion scalar parameters and brain volumetry analysis to predict the clinical conversion to CDMS in patients with early-stage CIS. From a clinical point of view, identifying patients with CIS having high-risk of conversion to

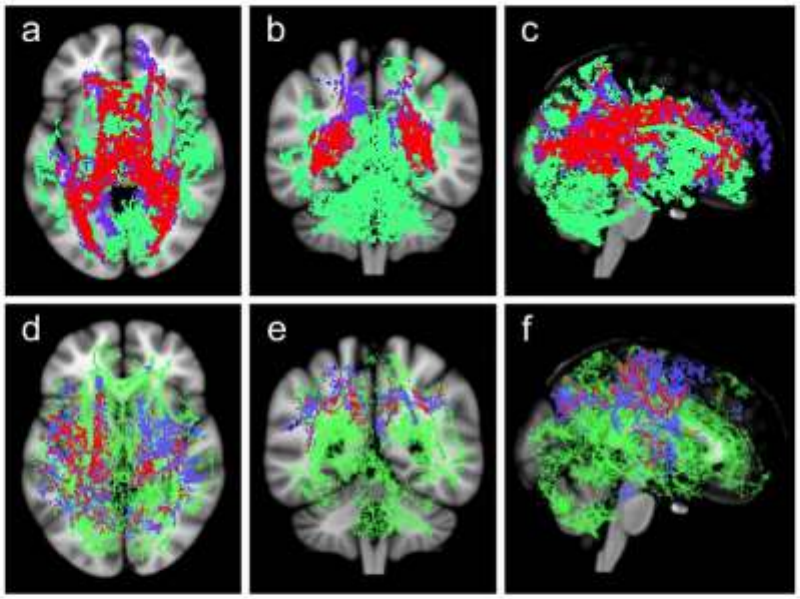



**Figure 2.** Graphic presentation of differences in distribution of diffusivity changes between the two tract-based spatial statistics (TBSS) analyses comparing patients and healthy controls and both subgroups of patients. Distribution of voxels with significantly different fractional anisotropy (FA) (a–c) and mean diffusivity (MD) (d–f) values between the groups compared in axial (a, d), coronal (b, e), and sagittal (c, f) projection. Blue colour represents voxels with significant differences between all patients and healthy controls. Green colour represents voxels with significant differences between the subgroups of patients defined by their later clinical progression, and red colour marks the significant voxels common for both analyses.(Color version of figure is available online).

CDMS is a matter of great importance. In recent years, a number of other groups of researchers have endeavoured to introduce methods for predicting the progression of CIS to CDMS based on clinical, electrophysiological, laboratory, or imaging findings (9,10).

If we focus on those studies using diagnostic imaging methods, efforts are being made either to find new possibilities for evaluating and quantifying findings on conventional MRI sequences or to find new advanced methods of MRI and possibly introduce them into routine clinical practice. An example of a study employing advanced techniques of structural MRI data analysis is that of Bendfeldt et al. (11), which evaluates combination of clinical and demographic data with evaluation of image-based lesion-specific geometry and brain volume. The highest prediction accuracy of 70.4% was achieved by a combination of lesion-specific geometric

**TABLE 4. Diffusion Tensor Imaging Parameters in Patients and Healthy Controls**

| Parameter | Patients Median (MM) | Controls Median (MM) | *p* Voxels† | Patients with Progression Median (MM) | Patients without Progression Median(MM) | *p* Voxels‡ |
|---|---|---|---|---|---|---|
| FA | 0.532 (0.398–0.574) | 0.545 (0.472–0.59) | <0.001 19303 | 0.394 (0.343–0.429) | 0.426 (0.391–0.455) | <0.001 50278 |
| MD [$10^{-6}$] | 730 (562–800) | 736 (699–773) | <0.001 15271 | 835 (702–969) | 811 (647–864) | <0.001 54227 |
| AD [$10^{-6}$] | 1168 (881–1235) | 1175 (1117–1234) | <0.001 49301 | - | - | - |
| RD [$10^{-6}$] | - | - | - | 659 (587–780) | 624 (515–683) | <0.001 75755 |

FA, fractional anisotropy; MD, mean diffusivity; AD, axial diffusivity; RD, radial diffusivity; MM, minimum–maximum. p-values representing statistical significance of differences between patients and controls (p†) and between the patient subgroups (p‡) were calculated using Mann–Whitney U test with correction for age, sex,and scanner. Number of voxels identified as significantly ($p<0.05$) different between patients and controls and between both subgroups of patients as revealed by tract-based spatial statistics (TBSS) (voxels† and voxels‡ respectively). ROIs used for the two analyses († and ‡) only partly overlap;therefore, the values entering those analyses are not fully comparable.

**TABLE 5. Receiver Operating Characteristic (ROC) Analysis of Significant Diffusion Tensor Imaging and Volumetric Parameters**

| Parameter | ROC area (CI) | *p* | Cut-Off | Sensitivity | Specificity |
|---|---|---|---|---|---|
| FA | 0.890 (0.806–0.974) | <0.001 | 0.408698 | 77.3 | 90.0 |
| MD | 0.707 (0.562–0.852) | <0.005 | 0.000826 | 63.6 | 78.0 |
| RD | 0.750 (0.615–0.885) | <0.001 | 0.000645 | 63.6 | 86.0 |
| White matter | 0.753 (0.643 – 0.863) | <0.001 | 725609.4 | 90.9 | 58.0 |
| Whole brain | 0.628 (0.488 – 0.768) | 0.085 | 1480222.8 | 50.0 | 70.0 |

FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity.

(image-based) and demographic and/or clinical features (11). Similarly, Wotschell et al. (12) report the potential of machine-learning algorithms for predicting conversion to CDMS by analysis of conventional proton density and T2-weighted images, revealing sensitivity of 77% and specificity of 66% during 1 year of observation.

We have chosen TBSS, a module of FSL, for the analysis of DTI data. This tool represents a widely accepted approach to diffusion data analysis using nonlinear registration of FA maps, reconstruction of main WM tracts, and projection of FA values of individual subjects onto this skeleton (13). This fully automatic method is virtually operator independent and also time efficient, which are the major benefits over manual techniques based upon region-of-interest that may be prone to subjective error and may provide poorer reproducibility compared to TBSS (14).

Our results generally correspond with previous findings of changes in DTI parameters of brain WM in patients with multiple sclerosis (15) or CIS (16) in comparison to healthy controls. In the normal, healthy population, brain WM, unlike the GM, has a comparatively high FA value and low MD values (17). As the brain ages, gradual physiological degeneration has been observed, and thus there is gradual decrease in FA and increase in MD in the WM (18). Similarly, in our study, mean values of MD were higher and values of FA lower in patients with progression compared to those without progression, which, from the perspective of diffusion properties of the brain WM, may resemble an accelerated process of aging. In contrast, we observed moderately lower MD values in some WM areas (mostly parietal and frontal lobe WM, Fig 2d–f) within the group of all patients compared to the control group. It is necessary to realize that the reported values of scalar parameters are not fully comparable between the two analyses, because they were measured within different areas of WM as given by TBSS analysis. Physiological structural heterogeneity probably plays some role here. Furthermore, according to previous studies, early demyelinating changes may demonstrate some degree of restricted diffusion (19) in contrast to chronic changes in CDMS patients, within whom higher MD values were found (20). Inasmuch as the MRI examinations were performed quite early after the initial clinical attack (approximately 42% of patients examined within the first month, median 2 months), we may speculate that lower MD values in the whole group of early-stage CIS patients compared to controls may generally be related to early demyelination, but comparatively higher MD values in selected patients with further clinical progression compared to non-progressed patients may be due to underlying (possibly subclinical) chronic ultrastructural abnormalities that are associated with risk of future clinical progression.

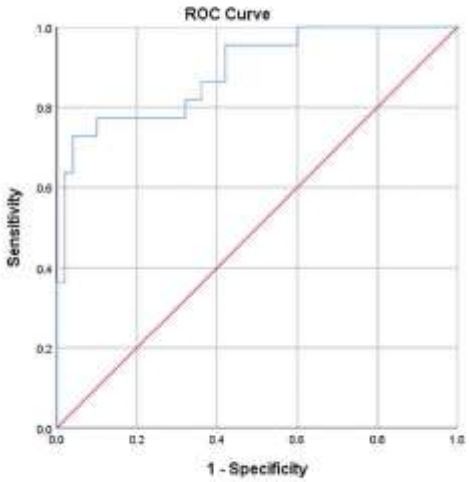



**Figure 3.** Receiver operating characteristic (ROC) curve of the fractional anisotropy (FA) parameter as a predictor of progression to clinically definite multiple sclerosis in patients with clinically isolated syndrome.

The differences in FA and MD were generally more widespread across the brain (eg, cerebellum) when the subgroups of patients defined by later progression to CDMS were mutually compared than in the case of comparing patients to healthy controls (Fig 2). Moreover, patients differed from the control group in AD, and, conversely, we proved significant differences in RD in patients with progression compared to those without progression. Those two parameters may be attributed to different ultrastructural abnormalities of WM, where RD is recognized as a marker of demyelination while changes in AD may more likely reflect axonal disintegration (21). From the perspective of diffusion properties, therefore, the changes in patients with later clinical progression appear moderately specific and show different characteristics compared to general abnormalities in CIS patients found in comparison to healthy subjects.

**TABLE 6. Volumetric Parameters in Patients and Healthy Controls**

| Volume | Patients Median (MM) | Controls Median (MM) | $p^1$ (value) | Patients with Progression Median (MM) | Patients without Progression Median (MM) | $p^2$ (value) |
|---|---|---|---|---|---|---|
| GM [$cm^3$] | 787 (674.2–900.6) | 812.5 (686.1–939.3) | <0.001 | 780 (705.4–897.6) | 787 (674.2–900.6) | 0.932 |
| WM [$cm^3$] | 717.5 (644.5–799.8) | 731.4 (624.3–810.6) | <0.05 | 694.2 (644.5–749.2) | 735 (655.4–799.8) | <0.001 |
| WB [$cm^3$] | 1511.9 (1348–1627) | 1537.6 (1379–1722) | <0.001 | 1491.3 (1365–1623) | 1515.1 (1348–1627) | <0.05 |

GM, grey matte; WM, white matter; WB, whole brain; MM, minimum–maximum; *p*-values representing statistical significance of differences between patients and controls ($p^1$) and between patient subgroups ($p^2$) were calculated using Mann–Whitney U test with correction for age, sex and scanner.

The FA scalar parameter quantifying anisotropy of diffusion within brain tissue may be perceived as a general marker of nerve fibre integrity disruption (22). In our study, FA of brain WM appeared as the strongest predictor of conversion to CDMS, with sensitivity of 77.3% and specificity of 90%. There is not much data in the literature about the potential of MRI diffusion techniques for predicting CIS to CDMS conversion. One previous study from Gallo et al. (23) found significant DTI abnormalities within normal appearing white matter (NAWM) in the brains of patients with CIS, but it did not find significant differences between a subgroup of patients with CIS initially fulfilling McDonald criteria for dissemination in space with later progression to CDMS and a subgroup without progression. Conversely, Kugler et al. (24), in their study (and similarly to our study), proved alterations of FA in cerebellar tissues as a predictor of conversion to CDMS. Moreover, histogram analysis of cervical cord's diffusion parameters also has been used in a recent study to predict CIS to CDMS progression with sensitivity and specificity of FA kurtosis of both WM and NAWM of 93% and 72%, respectively (25). Conventional evaluation of structural MRI data according to McDonald DIS criteria revealed sensitivity of 81.8% in terms of prediction to CDMS. That was comparable to the predictive power of FA, but the specificity of the conventional criteria and evaluation was substantially lower, at 48%. From this perspective, it may appear that the analysis of diffusion data is more accurate than is conventional MRI.

Several studies have been published confirming brain volume reduction in patients with CDMS. Moreover, the rate of volume reduction correlates with progression of the disease's clinical symptoms (26). Some studies have also evaluated cerebral atrophy in CIS patients (27) or even considered using brain volume measurements as a predictor of progression from CIS to CDMS while taking into consideration separately GM and WM volumes (26). One of the studies of a nature similar to that of ours is the study of Dalton et al. (28). In this work, the only statistically significant predictor of clinical progression was reduction in the volume of GM. The volume of WM was not significantly different in the two investigated groups. These findings contradict the results of our study, which evidence statistically significant reduction of whole brain volume and of WM to be most pronounced in patients with progression to CDMS within the next 2 years. The reason for these discrepancies may relate to a smaller sample of patients in the case of the first study and especially differences in methodology, as the cited older study uses not a T1 3D sequence for segmentation but only a 2D T2 sequence. In such case, a poorer WM / GM contrast can be expected and the resulting weaker spatial resolution may provide less precise volumetric data. We believe that our results are logical, given that demyelination generally affects WM more than GM (29) and WM atrophy correlates with the clinical state of the patients (30). WM volume demonstrated comparatively low specificity (58.0%) with respect to predicting clinical conversion in patients with CIS, but the sensitivity was comparatively higher (90.9%) and the discrimination power of volumetry was generally weaker compared to those of DTI parameters.

This study had several limitations. Some may consider as a limitation the use of a 1.5T MRI device that provides images with generally lower signal-to-noise ratio compared to 3T systems. On the other hand, this shortcoming is partially offset by the fact that a lower magnetic field, by its physical nature, produces smaller numbers of susceptibility artefacts compared to machines with higher induction that may become important especially in anatomical areas near the skull base. Furthermore, one of the recent multicentre studies has shown that most of the diffusion MRI-derived parameters are robust even across 1.5T and 3T scanners (31).

Another limitation is the use of two different MRI devices, as the hardware had been replaced during the study. Although both scanners were 1.5T devices from the same manufacturer and the examinations on the newer MRI device were performed using exactly the same acquisition parameter settings, the influence of different MRI hardware, especially on the diffusion scalar parameters, may be significant (32). Because the proportions of examinations performed on the two MRI devices were not equal among the study groups, the MRI device was included as a covariate into all statistical analyses to correct for possible influence of this factor; such approach has already been reported in the literature (33). Moreover, to further validate the results, we provide also the key analysis comparing progressed and non-progressed patients restricted merely to the group of 64 patients examined on a single MRI device. Inasmuch as these data do not differ substantially from the whole-group analysis, we believe that the influence of the different MRI hardware is not crucial. In any case, the reproducibility among different

MRI devices with different field strengths should be studied in relation to this topic before the techniques investigated here can be used in routine diagnostics.

Another limitation is the relatively small size of the study group due to recruitment of patients from only one multiple sclerosis centre. Although the number of patients is sufficient to provide statistically significant results, it would be appropriate to verify the results on a larger number of patients. In addition, the 2-year follow-up period is a relatively short time frame within which to evaluate the potential for CDMS conversion. The time to conversion to CDMS in CIS patients reported in the literature is somewhat variable. In a large study with more than 1,000 patients, for example, the median time to conversion was 1,096 days (34). Another study investigating a smaller group of patients indicated mean time to conversion of 11 months (9). With this in mind, a 2-year monitoring period was considered acceptable, especially given that relapsing activity during the initial 1 −2 years of the disease's course in patients with MS is of crucial prognostic importance in anticipating the severity level of future damage (35). It is important to note, however, that the conversion rate in CIS patients in the next 20 years reported in long-term studies reaches up to 50% −60% (36,37). Considering the lower conversion rate established in our study (30.6%), we may expect that some of the patients who remained clinically stable for 2 years may develop further clinical attack in future. Thus, the predictive power and longitudinal evolution of DTI and volumetric parameters need to be further investigated by long-term studies.

A certain bias may also have arisen from the effects of treatments initiated during the follow-up period in patients belonging to both subgroups. Although the numbers of treated patients were not significantly different between the patient subgroups as defined by their progression to CDMS, this fact could to some extent influence the measured diffusion parameters. Nevertheless, any intentional observation of this disease's natural progression, while a theoretically optimal methodology, would be wholly unacceptable from an ethical standpoint.

## CONCLUSION

This study provides evidence that the evaluation of DTI parameters together with brain volumetry in patients with early-stage CIS may be useful in predicting CIS conversion to CDMS within the following 2 years of the disease.

## ACKNOWLEDGEMENTS

Contract grant sponsor: Czech Health Research Council; contract number: AZV-15 32133A + Supported by Ministry of Health, Czech Republic – conceptual development of research organization (FNBr, 65269705)

## REFERENCE


1. Confavreux C, Vukusic S, Moreau T, Adeleine P. Relapses and progression of disability in multiple sclerosis. N Engl J Med 2000; 343:1430–1438.
2. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018; 17:162–173.
3. Poser CM, Paty DW, Scheinberg L, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 1983; 13:227–231.
4. Jenkinson M, Beckmann CF, Behrens TE, Woolrich W, Smith SM. Fsl. Neuroimage 2012; 62:782–790.
5. Battaglini M, Jenkinson M, De Stefano N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. Hum brain mapp 2012; 33:2062–2071.
6. Smith SM, Rao A, De Stefano N, et al. Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: cross-validation of BSI, SIENA and SIENAX. Neuroimage 2007; 36:1200–1206.
7. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004; 23:S208–S219.
8. Schmidt P, Gaser C, Arsic M, et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. Neuroimage 2012; 59:3774–3783.
9. Ciner BP, Özakbaş S. Prediction of conversion from clinically isolated syndrome to multiple sclerosis according to baseline characteristics: a prospective study. Arch Neuropsychiatry 2018; 55:15.
10. Kočova J, Kočica J, Hulová M, et al. Conversion of clinically isolated syndrome to multiple sclerosis: a prospective study. Mult Scler Relat Disord 2020; 44:102262.
11. Bendfeldt K, Taschler B, Gaetano L, et al. MRI-based prediction of conversion from clinically isolated syndrome to clinically definite multiple sclerosis using SVM and lesion geometry. Brain imaging Behav 2019; 13:1381–1374.
12. Wottschel V, Alexander DC, Kwok PP, et al. Predicting outcome in clinically isolated syndrome using machine learning. Neuroimage Clin 2015; 7:281–287.
13. Smith SM, Jenkinson M, Johansen-Berg H, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 2006; 31:1487–1505.
14. Lilja Y, Gustafsson O, Ljungberg M, Nilsson D, Starck G. Impact of region-of-interest method on quantitative analysis of DTI data in the optic tracts. BMC Med Imaging 2016; 16:1–10.
15. Metcalf M. Advanced MR technique development for improved characterization of multiple sclerosis (Doctoral dissertation, UCSF). 2008.
16. Cappellani R, Bergsland N, Weinstock-Guttman B, et al. Diffusion tensor MRI alterations of subcortical deep gray matter in clinically isolated syndrome. J Neurol Scienc 2014; 338:128–134.
17. Kodiweera C, Alexander AL, Harezlak J, McAllister TW, Wu YC. Age effects and sex differences in human brain white matter of young to middle aged adults: A DTI, NODDI, and q-space study. Neuroimage 2016; 128:180–192.
18. Draganski B, Ashburner J, Hutton C, et al. Regional specificity of MR contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). Neuroimage 2011; 55:1423–1434.
19. Bisele P, Szabo K, Griebe M, et al. Reduced diffusion in a subset of acute MS lesions: a serial multiparametric MRI study. Am J Neuroradiol 2012; 33:1369–1373.
20. Sbardella E, Tona F, Petsas N, Pantano P. DTI measurements in multiple sclerosis: evaluation of brain damage and clinical implications. Mult Scler Int 2013; 2013:671730.
21. Song SK, Sun SW, Ju WK, Lin SJ, Cross AH, Neufeld AH. Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. Neuroimage 2003; 20:1714–1722.
22. Seewann A, Vrenken H, van der Valk P, et al. Diffusely abnormal white matter in chronic multiple sclerosis: imaging and histopathologic analysis. Arch Neurol 2009; 66:601–609.
23. Gallo A, Rovaris M, Riva R, et al. Diffusion-tensor magnetic resonance imaging detects normal-appearing white matter damage unrelated to short-term disease activity in patients at the earliest clinical stage of multiple sclerosis. Arch Neurol 2005; 62:803–808.

24. Kugler AV, Deppe M. Non-lesional cerebellar damage in patients with clinically isolated syndrome: DTI measures predict early conversion into clinically definite multiple sclerosis. NeuroImage Clin 2018; 19:633–639.
25. Dostal M, Keřkovský M, Stulík J, et al. MR diffusion properties of cervical spinal cord as a predictor of progression to multiple sclerosis in patients with clinically isolated syndrome. J Neuroimaging 2021; 31:106–114.
26. Giorgio A, De Stefano N. Clinical use of brain volumetry. J Magn Reson Imaging 2013; 37:1–14.
27. Rojas JI, Patrucco L, Besada C, Bengules L, Cristiano E. Brain atrophy in clinically isolated syndrome. Neurologia (English Edition) 2010; 25:430–434.
28. Dalton CM, Chard DT, Davies GR, et al. Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. Brain 2004; 127:1101–1107.
29. Lassmann H. Multiple sclerosis pathology. Cold Spring Harb Perspect Med 2018; 8:a028936.
30. Sbardella E, Petsas N, Tona F, et al. Assessing the correlation between grey and white matter damage with motor and cognitive impairment in multiple sclerosis patients. PloS One 2013; 8:e63250.
31. Grech-Sollars M, Hales PW, Miyazaki K, et al. Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain. NMR Biomed 2015; 28:468–485.
32. Landman BA, Farrell JA, Jones CK, Smith SA, Prince JL, Mori S. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. NeuroImage 2007; 36:1123–1138.
33. Takao H, Hayashi N, Ohtomo K. Sex dimorphism in the white matter: fractional anisotropy and brain size. J Magn Reson Imaging 2014; 39:917–923.
34. Kuhle J, Disanto G, Dobson R, et al. Conversion from clinically isolated syndrome to multiple sclerosis: a large multicentre study. Mult Scler J 2015; 21:1013–1024.
35. Weinshenker BG, Bass B, Rice GPA, et al. The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. Brain 1989; 112:1419–1428.
36. Fisniku LK, Brex PA, Altmann DR, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain 2008; 131:808–817.
37. Miller D, Barkhof F, Montalban X, Thompson A, Filippi M. Clinically isolated syndromes suggestive of multiple sclerosis, part 2: non-conventional MRI, recovery processes, and management. Lancet Neurol 2005; 4:341–348.

## SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.acra.2021.12.015.