Ultra-low-field brain MRI morphometry: test-retest reliability and correspondence to high-field MRI

František Váša[1*], Carly Bennallick[1], Niall J. Bourke[1], Francesco Padormo[2], Levente Baljer[1], Ula Briski[1], Paul Cawley[3,4], Tomoki Arichi[3,4], Tobias C. Wood[1], David J. Lythgoe[1], Flavio Dell'Acqua[1], Thomas C. Booth[6,7], Ashwin V. Venkataraman[1,8], Emil Ljungberg[9,1], Sean C.L. Deoni[10], Rosalyn J. Moran[1], Robert Leech[1†], Steven C.R. Williams[1†]

[1] Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.

[2] Hyperfine, Inc., Guilford, Connecticut, USA.

[3] Research Department of Early Life Imaging, School of Imaging Sciences and Biomedical Engineering, King's College London, London, UK.

[4] Medical Research Council Center for Neurodevelopmental Disorders, King's College London, London, UK.

[5] Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.

[6] School of Biomedical Engineering & Imaging Sciences, King's College London, 9th Floor, Becket House, London, UK.

[7] Department of Neuroradiology, Ruskin Wing, King's College Hospital NHS Foundation Trust, London, UK.

[8] Centre for Healthy Brain Ageing, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.

[9] Department of Medical Radiation Physics, Lund University, Lund, Sweden.

[10] Bill and Melinda Gates Foundation, Seattle, Washington, USA.

*Corresponding author. E-mail: frantisek.vasa@kcl.ac.uk;

†These authors contributed equally to this work.

Abstract

Magnetic resonance imaging (MRI) enables non-invasive monitoring of healthy brain development and disease. Widely used higher field (>1.5 T) MRI systems are associated with high energy and infrastructure requirements, and high costs. Recent ultra-low-field (<0.1T) systems provide a more accessible and cost-effective alternative. However, it is not known whether anatomical ultra-low-field neuroimaging can be used to extract quantitative measures of brain morphometry, and to what extent such measures correspond to high-field MRI. Here we scanned 23 healthy adults aged 20-69 years on two identical 64 mT systems and a 3 T system, using T_1w and T_2w scans across a range of (64 mT) resolutions. We segmented brain images into 4 global tissue types and 98 local structures, and systematically evaluated between-scanner reliability of 64 mT morphometry and correspondence to 3 T measurements, using correlations of tissue volume and Dice spatial overlap of segmentations. We report high 64 mT reliability and correspondence to 3 T

across 64 mT scan contrasts and resolutions, with highest performance shown by combining three T_2w scans with low through-plane resolution into a single higher-resolution scan using multi-resolution registration. Larger structures show higher 64 mT reliability and correspondence to 3 T. Finally, we showcase the potential of ultra-low-field MRI for mapping neuroanatomical changes across the lifespan, and monitoring brain structures relevant to neurological disorders. Raw images and code are publicly available (*upon publication*), enabling systematic validation of pre-processing and analysis approaches for ultra-low-field neuroimaging.

Introduction

Magnetic resonance imaging (MRI) can provide detailed insight into the anatomy and physiology of a wide range of organs, in health and disease. However, conventional higher field MRI systems (>1.5 T) have significant infrastructure requirements (including shielding, stable power and cryogenic cooling) as well as associated installation and running costs.

The recent development of ultra-low-field MRI scanners, such as the 64 mT Hyperfine Swoop, promises to revolutionize neuroimaging (Arnold et al., 2023; Kimberly et al., 2023). Ultra-low-field MRI systems run using a standard electrical socket and are portable, enabling scanning at the bedside (Mazurek et al., 2021; Balaji et al., 2024) and in low- and middle-income settings with limited access to MRI (Chetcuti et al., 2022; Abate et al., 2024).

Initial studies have shown good correspondence of quantitative estimates of brain morphometry, following synthetic enhancement of ultra-low-field scans using deep-learning image quality transfer. Iglesias et al. (2023) reported significant correlations between volumetric measurements from adult ultra-low-field and high-field scans. Subsequently, Cooper et al. (2024) reported good agreement of tissue volume and surface area, and moderate correspondence of cortical thickness estimates, in young adult scans. However, both studies only reported moderate correspondence to 3 T data, with results at least somewhat dependent on prior enhancement of scans using deep learning (Iglesias et al., 2021). Moreover, these prior results were dependent on the simultaneous use of multiple contrasts (i.e., both T_1w and T_2w) for image enhancement and/or morphometry. Other recent studies included concurrent acquisition of paired ultra-low-field and high-field data, across a range of participants including infants (Baljer et al., 2024) or patients with multiple sclerosis (Arnold et al., 2022); however, these studies were not primarily focused on the correspondence of morphometric estimates. Notably, there have been no prior studies on the test-retest reliability of 64 mT imaging - whether between or across scanners - and derived morphometric measurements, which are crucial to enable pooling of data across sites (Abate et al., 2024). Additionally, no prior studies have publicly shared the underlying MRI data in a readily available form, hindering the development of new approaches for analyzing ultra-low-field MR images.

Here, we present the results of morphometric analyses of healthy adults scanned on two identical 64 mT scanners, and a 3 T scanner, without prior image enhancement (e.g., using image quality transfer). We scanned 23 healthy adult participants using T_1w and T_2w scans, including four different voxel resolutions at ultra-low-field, three of which were combined to generate a fifth higher-resolution volume. We segmented and parcellated all scans into 98 structures across 4 tissue types, and extracted corresponding tissue volumes, using SynthSeg+ (Billot et al., 2023). We used these estimates to quantify between-scanner test-retest reliability of 64 mT scans, and their correspondence to 3 T scans, using both correlations of volume across participants, and Dice overlap of segmentations within participants (following rigid-registration of 64 mT to 3 T scans within participants). For details, see Fig. 1 and Methods. We report excellent between-scanner test-retest reliability of 64 mT scan morphometry, and correspondence to 3 T MRI, across contrasts and resolutions. The underlying raw MRI scans and analysis code are publicly available (*upon publication*), to help accelerate the development and optimization of methods for the pre-processing and analysis of ultra-low-field brain MRI data.

Results

Data acquisition and quality

We recruited 23 healthy adult participants aged 20-69 years, with 2-3 males and 2-3 females per decade (for participant distribution by age and sex, see supplementary Figure S1). We acquired both T_1w and T_2w scans on a 3 T high-field scanner (GE Healthcare SIGNA Premier, Waukesha, WI) and two identical 64 mT ultra-low-field scanners (Hyperfine Swoop, Guildford, CT; hardware version: 1.7, software version: 8.6.0).

Ultra-low-field scans were acquired at four resolutions, including three product sequences with higher resolution within the axial, coronal or sagittal plane ($1.5\times1.5\times5$ mm^3 or $1.6\times1.6\times5$ mm^3, hereafter referred to as AXI, COR, SAG respectively) and a custom isotropic sequence of approximately equivalent voxel volume (2.2 mm^3 or 2.3 mm^3 isotropic, hereafter referred to as ISO). We further combined AXI, COR and SAG scans into one within participants using multi-resolution registration (Deoni et al., 2022), yielding a single higher-resolution "MRR" volume per contrast (of effective resolution 1.5 mm^3 or 1.6 mm^3 isotropic). High-field scans were acquired at 1 mm^3 isotropic resolution. For details regarding scan parameters, see supplementary Table S1.

Identical contrasts were acquired on both 64 mT scanners, except for T_2w ISO scans, which were only acquired at one site and therefore omitted from test-retest analyses. Moreover, the T_1w SAG scan did not properly reconstruct for one participant, resulting in analyses on T_1w SAG and T_1w MRR scans being carried out on a sample of 22 (out of 23) participants.

Ultra-low-field MRI scans are of high quality on visual inspection, particularly within the high-resolution plane of acquisition, but also inherently of lower visual quality compared to high-field MRI, characterized by lower resolution, reduced contrast between tissue types and increased noise (Fig. 1A,B). We visually controlled the quality of all raw scans, their SynthSeg+ segmentations (Billot et al., 2023) and FSL FLIRT rigid-registrations within participants (Jenkinson and Smith, 2001). For details, see Methods and Supplementary Information.

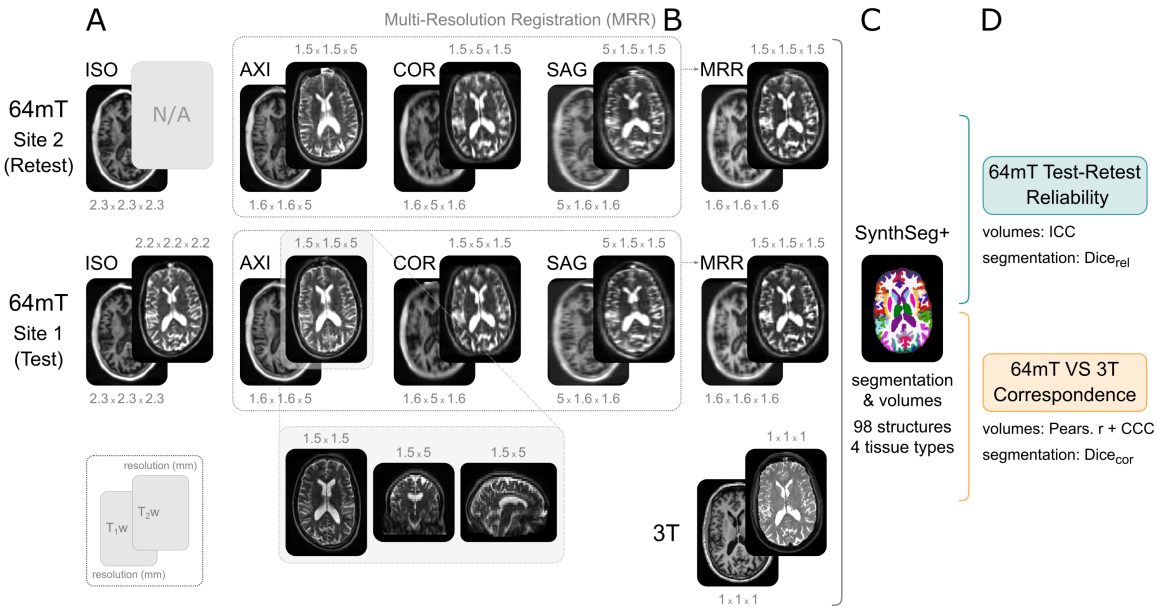

Figure 1: Data and pre-processing. A) Each participant was scanned on two identical 64 mT scanners and a 3 T scanner, using T_1w and T_2w scans, across a range of 64 mT resolutions (listed in mm above or below each scan). The second 64 mT site did not include a T_2w ISO scan. For details regarding acquisition parameters, see supplementary Table S1. B) The three scans with high in-plane resolutions were combined into a single higher-resolution volume using multi-resolution registration (MRR; Deoni et al., 2022). C) All scans were segmented and parcellated into 98 structures across 4 tissue types using SynthSeg+ (Billot et al., 2023). D) The resulting volumes and segmentations were used to quantify the test-retest reliability of 64 mT scans, and their correspondence to 3 T MRI (ICC: intraclass correlation coefficient; CCC: Lin's concordance correlation coefficient).

Results were broadly consistent across contrasts and 64 mT resolutions. In the main text, we focus primarily on results obtained using T_2w MRR 64 mT scans, which show the highest between-scanner reliability and correspondence to 3 T scans. We summarize the reliability and correspondence to 3 T for both T_1w and T_2w scans and across resolutions in Table 1 and supplementary Figures S3-S6.

Global reliability of 64 mT and correspondence to 3 T

Ultra-low-field T_2w MRR scans combined using multi-resolution registration (Deoni et al., 2022) showed excellent reliability between both 64 mT scanners, across all four tissue types (cortical grey matter, GM_{cort}; subcortical grey matter, GM_{subc}; white matter, WM; cerebrospinal fluid, CSF; for details regarding the assignment of individual structures to tissue types, see supplementary Table S4). Correlations of tissue volumes across participants were particularly high, with intraclass correlation coefficient (ICC) values of 0.98-1.0 (Fig. 1A). Accurate voxel-level labeling of tissues is a more difficult task, with commensurately lower but still very high reliability, including a reliability Dice overlap (referred to as $Dice_{rel}$) of underlying segmentations within participants showing median values of 0.76-0.92 (Fig. 1B).

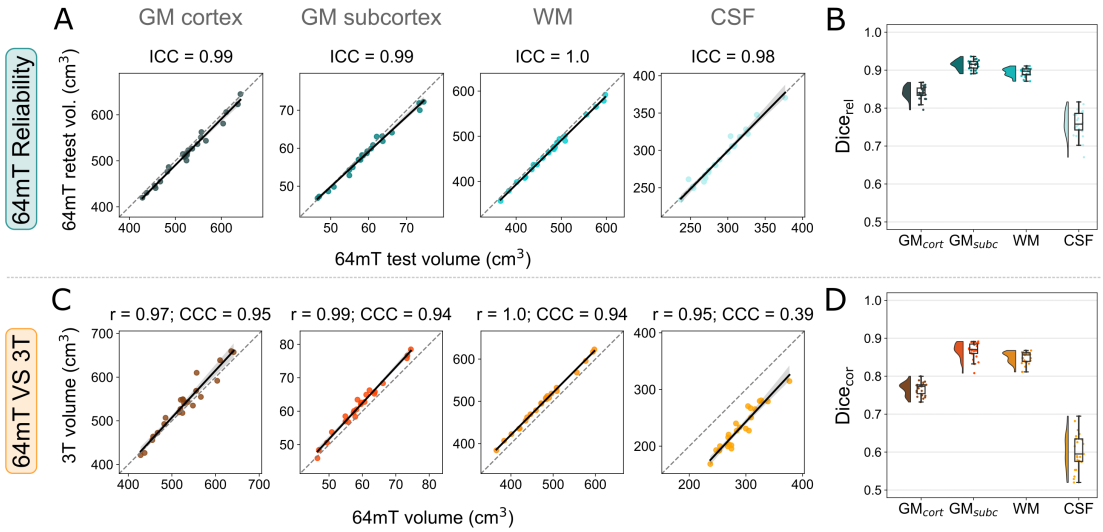

Figure 2: Global reliability of 64 mT scans, and correspondence to 3 T MRI. A) Correlations of global tissue volumes between both 64 mT sites, across participants. B) The Dice overlap of underlying rigid-registered segmentations within participants. C) Correspondence of 64 mT (site 1) data to 3 T, across participants. Statistics indicate the linear correlation (Pearson's r) and exact agreement, or alignment with the $x = y$ identity line (Lin's concordance correlation coefficient; CCC). D) Overlap of segmentations across field strengths, within participants. For results across all contrasts and resolutions, see supplementary Figures S3-S6.

Global tissue volumes extracted from 64 mT T_2w MRR scans showed both high linear correlations to 3 T counterparts, with Pearson's r across tissue types in the range 0.95-1.0, but also under- or over-estimates resulting in Lin's concordance correlation coefficient (CCC), quantifying exact agreement, in the range 0.39-0.95 (Fig. 1C). The correspondence Dice overlap (referred to as $Dice_{cor}$) of globally-aggregated segmentation labels across field strengths was lower than between both 64 mT scanners, but remained high, with median values in the range 0.60-0.87 (Fig. 1D).

For results of global 64 mT reliability and correspondence to 3 T across contrasts and resolutions, see supplementary Figures S3-S6.

Assessment of percentage difference in volume estimates at 64 mT, relative to 3 T, confirmed median underestimates of volume in grey and white matter ($GM_{cort} = -1.5\%$, $GM_{subc} = -3.7\%$, $WM = -4.4\%$), and overestimates in the CSF (+24.6%), apparent in global correspondence plots (Fig. 1C). Individual structures showed a similar trend, with median values ranging from −25.7% to +28.0% (Md $[Q_1,Q_3] = -1.6\ [-4.5, -0.9]\%$). For a summary of percentage differences across contrasts and resolutions, see supplementary Table S2.

Local reliability of 64 mT and correspondence to 3 T

We next repeated analyses locally, across all 98 structures provided by the SynthSeg+ segmentation and cortical parcellation (Billot et al., 2023). In Figures 3 and 4, we show both an overview of reliability and correspondence statistics across regions (left), as well as detailed results for three exemplar regions for each statistic, showcasing structures with minimum, median and maximum values (right).

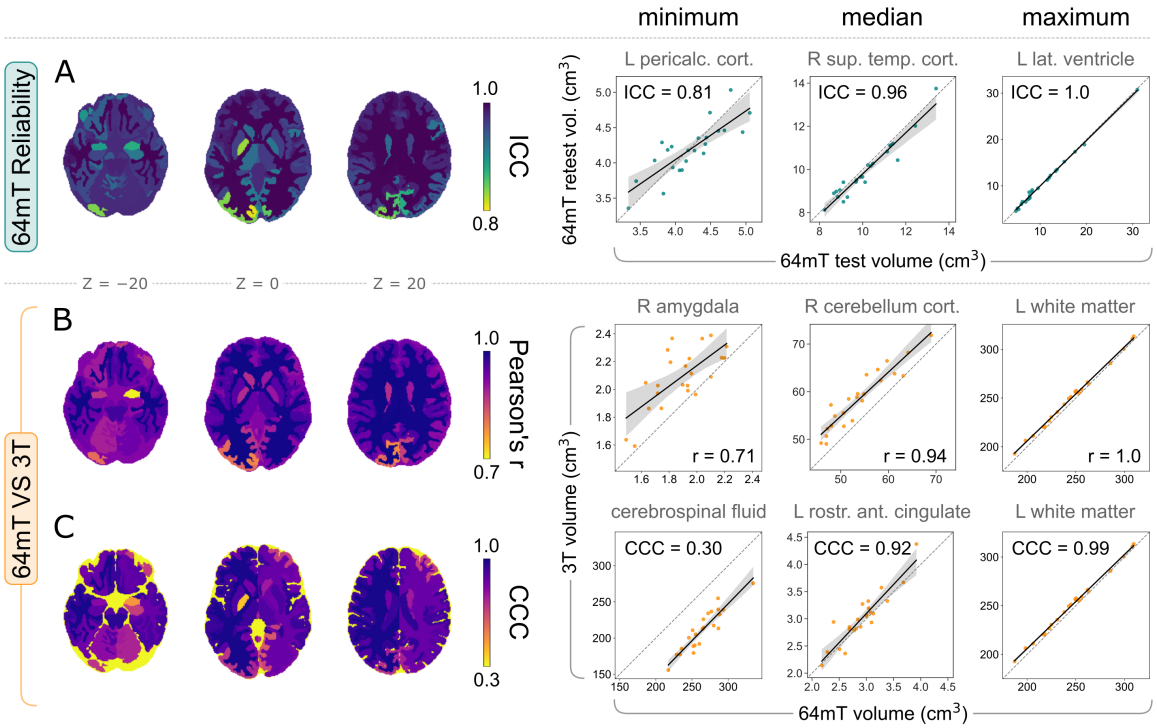

Figure 3: Local reliability of 64 mT tissue volumes, and correspondence to 3 T MRI. Left: summary of correlation statistics across regions, on three axial slices of a segmented high-field MNI template. Right: the corresponding regions with minimum, median and maximum volume correlations. A) 64 mT between-scanner reliability of local volumes, quantified using the intraclass correlation coefficient (ICC). Below, the correspondence of local 64 mT volumes to 3 T, assessed using B) Pearson's r, which quantifies linear correlation, and C) Lin's concordance correlation coefficient (CCC), which quantifies exact agreement.

Between-scanner reliability of 64 mT local volumes demonstrated variability across regions, but remained very high overall, with median ICC = 0.96 (Fig. 3A). The correspondence of volumes extracted from 64 mT scans to 3 T counterparts remained high at the individual structure level as well, with both high linear correlations exemplified by median r = 0.94 (Fig. 3B) and good agreement, shown by median CCC = 0.92 (Fig. 3C). However, some structures showed lower values, including for example minimum r = 0.71 and minimum CCC = 0.30, in the right amygdala and extra-cortical CSF respectively (Fig. 3B,C).

Similarly to global tissue types, we next assessed the median (across participants) Dice overlap of segmentations at the local level, both within 64 mT scans (Dice_{rel}) and across field strengths (Dice_{cor}) (Fig. 4). Segmentations derived from both 64 mT sites showed high overlap, including median $\text{Dice}_{\text{rel}} = 0.81$ across regions, but also moderate variability, with a Dice_{rel} coefficient range of 0.63-0.94 (Fig. 4A). Similarly to global volumes, the overlap of 64 mT segmentations to 3 T counterparts was slightly lower (than reliability), but remained high, with median $\text{Dice}_{\text{cor}} = 0.74$ and a range of 0.53-0.92 (Fig. 4B), indicating that ultra-low-field scans give rise to generally reliable and accurate voxel-wise segmentations at the local level as well.

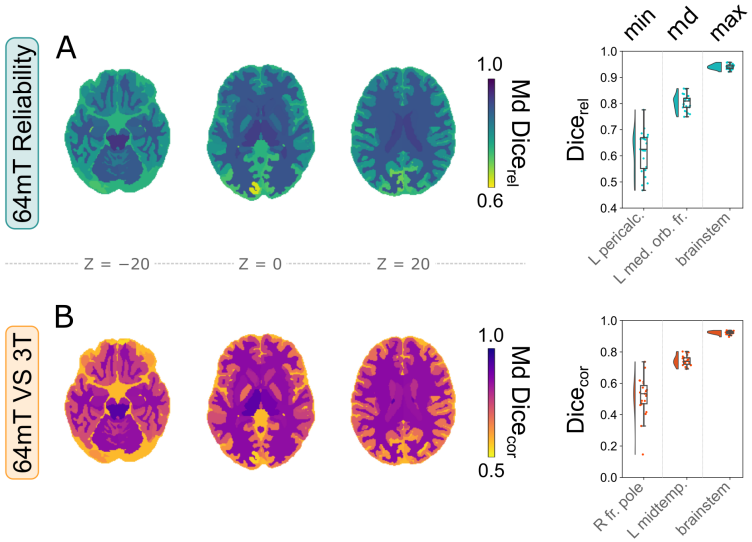

Figure 4: Local reliability of 64 mT segmentations, and correspondence to 3 T MRI. Left: median Dice overlap of all structures across participants, on three axial slices of a segmented high-field MNI template. Right: the corresponding regions with minimum, median and maximum Dice overlap. A) Reliability of local segmentations between 64 mT scanners. B) Correspondence of local segmentations between 64 mT and 3 T scans.

Relationships between regional statistics

We next wished to ascertain whether structures with more reliable 64 mT volume estimates and segmentations show higher correspondence to 3 T scans, and whether these quantities are additionally related to the size of regions. We therefore calculated Spearman's rank correlations between the 98 regional values of each statistic, across all pairs of six measures, including both measures of 64 mT reliability (ICC and median Dice$_{rel}$), three measures of 3 T correspondence (Pearson's r, Lin's CCC and median Dice$_{cor}$) and the median volume of regions across all participants, as estimated using 3 T MRI scans (Fig. 5).

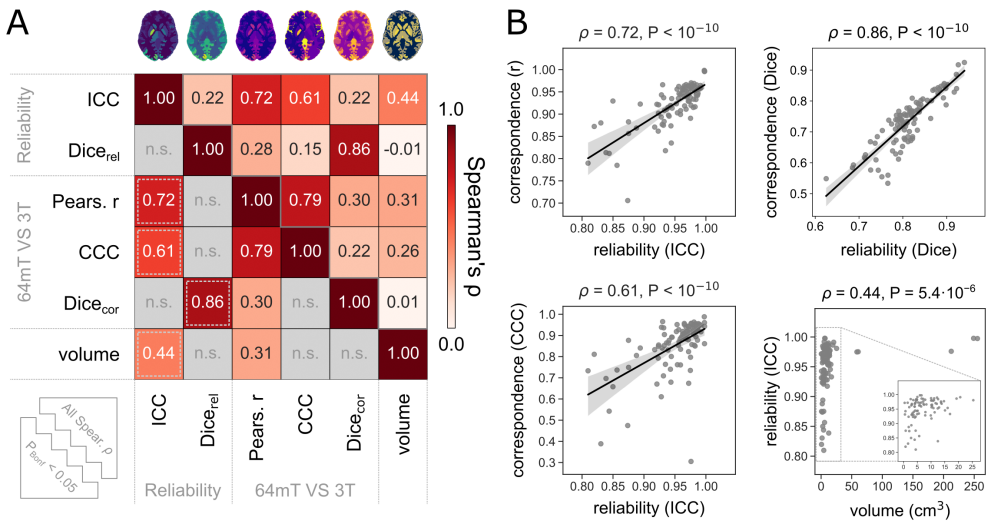

Figure 5: Larger structures are more reliably estimated using 64 mT MRI and show greater correspondence to 3 T MRI. A) Spearman's rank correlations between local maps, of 64 mT reliability, correspondence to 3 T, and average volume of each structure (extracted from high-field scans). In the lower triangular part, only correlations which remain significant after Bonferroni correction for multiple comparisons are shown (i.e. $p < 0.05/15 = 0.0033$); the remaining non-significant (n.s.) correlations are greyed out. Four exemplar significant relationships, indicated by dashed grey boxes in the lower triangular part of matrix A, are plotted in part B). (All reliability and correspondence results stem from T$_2$w MRR scans.)

Measures were positively correlated, with 7/15 relationships surviving Bonferroni correction for multiple comparisons (i.e. $p < 0.05/15 = 0.0033$), indicating that larger structures can be more reliably estimated using 64 mT scans, and show greater correspondence to 3 T scans. Relationships between different measures of 64 mT reliability were strong and relatively linear, as were relationships between measures of 64 mT reliability and correspondence across fields; however, relationships between regional volume and remaining statistics were commensurately weaker and showed nonlinear relationships. For scatterplot illustrations of four exemplar relationships, see Fig. 4B.

Reliability and correspondence across contrasts and 64 mT resolutions

So far, we focused on results obtained using 64 mT T_2w MRR scans, generated from three orthogonal scans using multi-resolution registration (Deoni et al., 2022), which showed the highest median between-scanner reliability and correspondence to 3 T 1 mm^3 isotropic counterparts. Below, we summarise 64 mT reliability and correspondence to 3 T scans across all five 64 mT resolutions, for both T_1w and T_2w scans (Table 1). We focus on median values, and first and third quartiles (i.e. Md [Q_1, Q_3]) across the 98 regional values of all five main reliability and correspondence statistics.

Table 1: Summary of regional reliability and correspondence, across contrasts and resolutions. Main values correspond to median [first, third quartile], i.e. Md [Q_1, Q_3], across 98 SynthSeg+ labels. The last column is the difference between the maximum and minimum Md value within each row. Maximum values for each statistic are highlighted in bold.

	contrast	**MRR** T_1w 1.6×1.6×1.6 T_2w 1.5×1.5×1.5	**ISO** T_1w 2.3×2.3×2.3 T_2w 2.2×2.2×2.2	**AXI** T_1w 1.6×1.6×5 T_2w 1.5×1.5×5	**COR** T_1w 1.6×5×1.6 T_2w 1.5×5×1.5	**SAG** T_1w 5×1.6×1.6 T_2w 5×1.5×1.5	**max - min**
ICC	T_1w	0.96 [0.93,0.98]	0.95 [0.90,0.96]	0.91 [0.86,0.94]	0.93 [0.88,0.96]	0.92 [0.82,0.94]	0.05
	T_2w	**0.97** [0.93,0.98]	N/A	0.95 [0.91,0.97]	0.94 [0.91,0.97]	0.92 [0.84,0.95]	0.05
Md $Dice_{rel}$	T_1w	0.79 [0.77,0.82]	0.72 [0.68,0.77]	0.75 [0.73,0.79]	0.74 [0.71,0.77]	0.75 [0.73,0.78]	0.07
	T_2w	**0.81** [0.79,0.84]	N/A	0.76 [0.71,0.79]	0.77 [0.74,0.81]	0.77 [0.75,0.81]	0.05
Pearson's r	T_1w	0.93 [0.87,0.95]	0.90 [0.86,0.93]	0.88 [0.81,0.92]	0.89 [0.82,0.93]	0.89 [0.88,0.92]	0.05
	T_2w	**0.94** [0.90,0.96]	0.93 [0.89,0.95]	0.92 [0.86,0.94]	0.92 [0.87,0.95]	0.88 [0.82,0.92]	0.06
Lin's CCC	T_1w	0.60 [0.45,0.72]	0.52 [0.26,0.75]	0.53 [0.21,0.75]	0.61 [0.32,0.82]	0.59 [0.23,0.80]	0.09
	T_2w	**0.89** [0.83,0.93]	0.84 [0.58,0.92]	0.77 [0.65,0.88]	0.79 [0.64,0.88]	0.76 [0.54,0.85]	0.13
Md $Dice_{cor}$	T_1w	0.61 [0.53,0.70]	0.59 [0.52,0.69]	0.61 [0.54,0.66]	0.61 [0.53,0.69]	0.56 [0.48,0.66]	0.05
	T_2w	**0.74** [0.68,0.79]	0.65 [0.57,0.72]	0.62 [0.52,0.71]	0.67 [0.60,0.73]	0.59 [0.54,0.66]	0.15

Differences across scan resolutions were broadly consistent between T_2w and T_1w scans and were related to differences in voxel resolution and/or acquisition time. MRR scans, with the highest effective resolution, showed the highest performance across most measures, followed by high performance on (T_1w) ICC and (T_2w) CCC for ISO scans directly acquired at (lower) isotropic resolutions, with shorter thick-slice AXI, COR and SAG scans showing lower performance. Differences between the latter three acquisitions, as well as ISO scans on measures of Dice spatial overlap, were generally relatively minor. Differences across all five resolutions were more pronounced for cross-field correspondence than for 64 mT reliability statistics, and most prominent within measures of exact agreement across scanner field strengths (i.e., Lin's CCC and median $Dice_{cor}$), where MRR scans showed highest gains in performance relative to other resolutions (Table 1).

Furthermore, T_2w scans showed moderately higher performance than T_1w scans, possibly due to a slightly higher voxel resolution (e.g., T_2w MRR scans have an effective resolution of 1.5 mm^3 isotropic, compared to 1.6 mm^3 isotropic for T_1w MRR scans) as well as potential differences in partial volume artifacts at the time of data acquisition (see Discussion for details). Differences in performance across contrasts were again more prominent for measures of agreement across field strengths (Lin's CCC and median $Dice_{cor}$), where T_2w scans showed highest gains in performance compared to T_1w scans (e.g. with median performance gains of T_2w MRR relative to T_1w MRR scans of $\Delta CCC = 0.29$, $\Delta Dice_{cor} = 0.13$, compared to $\Delta ICC = 0.01$, $\Delta Dice_{rel} = 0.02$ and $\Delta r = 0.01$).

Measuring age-related changes in tissue volume

We wished to assess the extent to which 64 mT scans can be used to track changes in tissue volume, using linear models of both global and local volumes as a function of age, covarying for sex. Although our sample is relatively small and cross-sectional, the uniform distribution of participants across age (supplementary Figure S1) combined with marked changes in tissue volume across the lifespan (Bethlehem et al., 2022) result in clear age-related changes in both 3 T scans (Fig. 6A) and 64 mT scans (Fig. 6B). As expected, at the global level both field strengths show decreases in GM and WM volumes, and increases in CSF volumes, with similar effect sizes, albeit of reduced magnitude at 64 mT (including t-statistics for the effect of age, corresponding p-values and r^2 for the variance in tissue volume explained by age and sex; Fig. 6 left).

Local rates of change also showed excellent alignment across field strengths, with correspondence between local t-statistic maps of r = 0.90 and CCC = 0.74 (Fig. 6 right). Alignment was similar between local maps of the proportion of variance in tissue volumes explained by both age and sex (i.e., r^2), which showed r = 0.78 and CCC = 0.71 between field strengths and across regions.

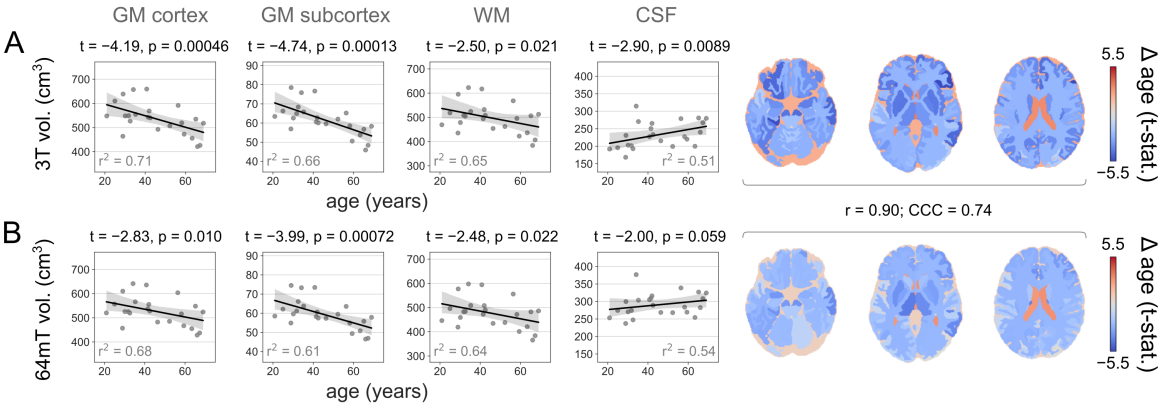

Figure 6: Changes in tissue volume from 3 T and 64 mT scans as a function of age. Rates of change of tissue volume as a function of age, in T_2w A) 3 T scans and B) 64 mT MRR scans, across both four aggregate tissue classes (left) and individual regions (right). Both scanners show similar changes (values correspond to Pearson's r and Lin's CCC). For results across contrasts and 64 mT resolutions, see supplementary Table S3.

For correspondence of regional age-related changes in tissue volume across contrasts and 64 mT resolutions, see supplementary Table S3.

Potential clinical applications

Finally, although our recruitment was restricted to neurologically healthy participants, we wished to explore the potential of clinical applications of ultra-low-field MRI in proof-of-concept analyses. For these analyses, we focused on T_2w MRR ultra-low-field scans, which previously showed highest reliability and correspondence to high-field imaging.

In Fig. 7A, we highlight the between-scanner reliability, and correspondence between 64 mT and 3 T, of hippocampal volume, a key region of interest for the diagnosis and monitoring of Alzheimer's Disease and other neurodegenerative processes. Reliability and correspondence statistics were derived as reported above. Similarity to other global and local tissue volumes, hippocampal volumes show high 64 mT test-retest reliability, and correspondence to 3 T (Fig. 7A).

Additionally, we investigated the potential of 64 mT scans to identify individuals who deviate from the norm (Marquand et al., 2019). We first corrected local tissue volumes for the effects of age and sex by retaining the residuals of linear models (described above), and subsequently computed non-parametric equivalents of the Z-score by robust-scaling corrected volumes, within each region, across participants (Leys et al., 2013). We then identified the participant and region showing the greatest deviation from the median in T_2w high-field scans, and visualized both (i) the corrected volume of this region from all three scanners, across participants, and (ii) the reliability and correspondence of all

regional deviations in this participant (Fig. 7B). The same participant was identified as the greatest outlier across scanners. Even after correcting regional volume for age and sex, regional deviations within this participant showed high reliability and correspondence across regions.

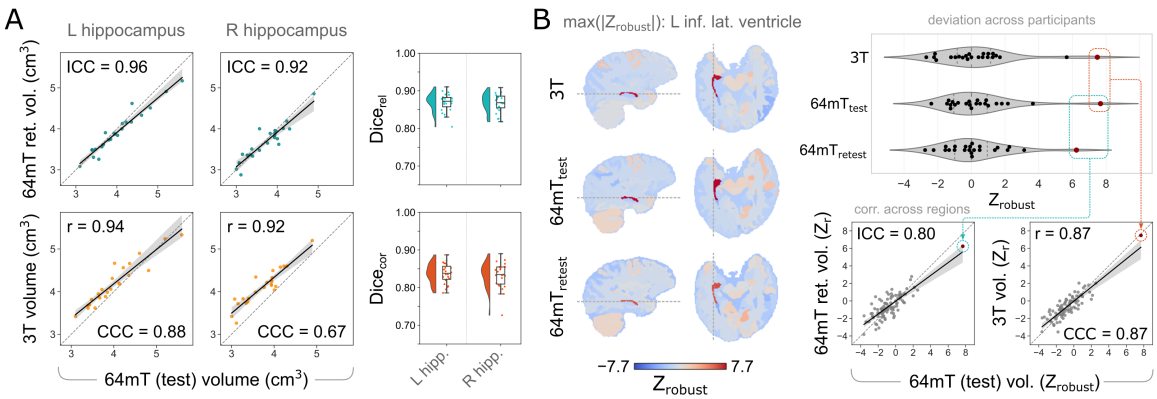

Figure 7: Clinical applications. A) Hippocampal volume and segmentation are relevant for the diagnosis and monitoring of Alzheimer's disease and other neurodegenerative processes, and show high 64 mT reliability (top) and correspondence to 3 T MRI (bottom), both in volume correlations (left) and Dice overlap of segmentations (right). B) Individual deviations of tissue volumes from the norm. Individual robust Z-scores of tissue volumes (corrected for age and sex) highlight the same outlier individual, and region, across scanners.

Discussion

We reported results from 23 healthy adult participants, scanned on two identical portable 64 mT ultra-low-field MRI scanners, and a 3 T high-field scanner. We segmented T_1w and T_2w scans acquired at a range of resolutions into 98 structures across 4 tissue types and estimated their volumes. We showed excellent test-retest reliability of morphometric measurements extracted from 64 mT scans, and high correspondence to 3 T MRI, using volume correlations and Dice overlap of both global tissue classes and individual regions. Reliability and correspondence were highest using T_2w MRR scans, obtained using multi-resolution registration of three orthogonal 64 mT acquisitions with thick slices but high in-plane resolution; however, these lower-resolution scans also showed excellent performance individually, as did scans directly acquired at (reduced) isotropic resolution. In additional analyses, we showcased the potential of ultra-low-field imaging for assessment of age-related changes in tissue volume, as well as potential clinical applications.

Determinants of reliability and correspondence

We reported higher between-scanner 64 mT reliability and correspondence to 3 T scans for morphometric measures extracted from T_2w ultra-low-field scans, than from T_1w scans, in this dataset. This may be due to several factors. First, the T_2w scans were of slightly higher resolution across acquisitions, including for example 1.5×1.5 mm^2 within the high-resolution plane for AXI, COR and SAG scans, compared to 1.6×1.6 mm^2 for corresponding T_1w scans; or a resulting effective resolution of 1.5 mm^3 isotropic for T_2w MRR scans, compared to 1.6 mm^3 isotropic for T_1w MRR scans (see supplementary Table S1 for details). Additionally, within the Hyperfine Swoop software version used at the time of acquisition (rc8.6.0), 64 mT T_1w sequence parameters resulted in non-zero CSF signal and partial-volume signal voids at tissue interfaces, possibly explaining lower correspondence to high-field imaging. Updated parameters in more recent software versions now yield T_1w scans with nulled CSF signal whilst retaining grey/white matter contrast, which may result in greater correspondence to high-field scans.

Both reliability and correspondence to 3 T were higher for 64 mT acquisitions with higher (effective) resolution; namely, highest for 1.5 mm^3 (and 1.6 mm^3) isotropic MRR scans, followed by isotropic

ISO scans and non-isotropic AXI / COR / SAG scans, all with identical voxel volumes (within contrasts). Differences between these lower-resolution scans were relatively minor, with a tendency for COR scans (with higher resolution within the coronal plane) to show higher performance. These results are in line with prior work showing that morphometric reliability is affected by voxel size and geometry, with larger voxels leading to lower morphometric reliability (Wonderlick et al., 2009; Haller et al., 2016) and accuracy (Mulder et al., 2019). While the SynthSeg+ algorithm used here for segmentation and volume estimation is designed to run on scans of any contrast and resolution (Billot et al., 2023), its performance remains partially dependent on variations in image properties.

We also found significant relationships of both reliability and correspondence statistics to region size (as estimated using high-field MRI), which can be summarized as: "the bigger (the region), the better (the agreement)". This aligns with prior research demonstrating a tendency for smaller (sub-)regions to have lower morphometric reliability (Morey et al., 2010; Whelan et al., 2016; Madan and Kensinger, 2017; Elliott et al., 2023).

Notably, prior work on the correspondence of 64 mT morphometric measurements to 3 T (Iglesias et al., 2023; Cooper et al., 2024) relied on the use of pre-trained deep-learning algorithms to enhance the quality of ultra-low-field scans (Iglesias et al., 2021). Here we show excellent reliability and correspondence across fields without prior application of image quality transfer (IQT). This eliminates the potential for "hallucinations" by neural networks, which may subtly alter anatomy while super-resolving images (Zhao et al., 2020; Lin et al., 2023), particularly if the input image deviates from the training dataset; for example, if applying a super-resolution model trained on healthy adults to paediatric and/or clinical scans (Iglesias et al., 2021, 2023; Baljer et al., 2024). Still, IQT may be useful for many applications, including qualitative visual assessment of scans for clinical purposes, or application of pre-processing tools which may require images of a specific contrast and/or resolution (e.g., "standard" 1 mm^3 isotropic T_1w scans; (Iglesias et al., 2021)). We also note that the Hyperfine Swoop 64 mT system includes a deep-learning processing step to reduce image blurring and noise that is performed on the scanner during image reconstruction (Krupinski et al., 2023); as such, this impacts all scans acquired using the Hyperfine Swoop ultra-low-field system.

Our results may help inform the choice of protocol for future 64 mT studies acquiring T_1w and/or T_2w scans. The specific choice of sequences is likely to depend on the question of interest, and time available to scan each participant; however, based on our results, for the purposes of segmentation and tissue volume estimation, it seems sensible to prioritize thick-slice T_2w scans acquired in three orthogonal orientations, which can be combined into a single isotropic scan using multi-resolution registration (Deoni et al., 2022). Beyond the analytic applications explored here, investigators can use the publicly available data (*upon publication*) to validate and optimize alternative or custom analytic pipelines, to maximize the reliability and/or cross-field correspondence of desired imaging phenotypes.

Ultra-low-field MR imaging

Analysis of MR images for morphometric measurements relies on the ability to discern anatomical structures based on the MR image contrast. In brain imaging, white matter, gray matter, and CSF can be clearly distinguished as their T_1, T_2, and T_2* relaxation times are different. These differences are driven by the underlying tissue microstructure, e.g., myelin, iron, and water content. However, these relaxation times change with Larmor frequency thus making the tissue contrast, i.e., the basis for our morphometric measurements, dependent on the field strength of the MR system.

The field strength dependence of relaxation times is complicated, non-linear, and depends on numerous factors, but it generally holds that T_1 is shorter at lower field strength (Rooney et al., 2007), T_2 only shows small changes, if any (Stanisz et al., 2005; Jordanova et al., 2023), and T_2* increases (Peters et al., 2007). Since the relaxation times, which produce the tissue contrast, do not change in the same way with field strength, different pulse sequence parameters are required at high and low field strengths (see supplementary Table S1), making it non-trivial to produce identical image contrast. This could impact morphometric measurements if brain regions are heterogeneous in their tissue composition (e.g., Fukunaga et al., 2010).

Another factor to consider is the signal-to-noise ratio (SNR) which is inherently tied to the field strength, resulting in greatly reduced SNR at 64 mT compared to 3 T (Marques et al., 2019). To partly compensate for this, ultra-low-field MR systems use large voxel size (here, $1.5\times1.5\times5$ mm^3

compared to 1×1×1 mm^3 for T_2w 64 mT and 3 T product sequences), and more signal averaging, resulting in longer acquisition times. When participants lie in the scanner for a longer period of time, the risk of motion artifacts also increases, an issue that could be particularly relevant for older and paediatric populations (Madan, 2018; Padormo et al., 2023).

The ultra-low-field MRI system used in this study, Hyperfine Swoop, is one of the few commercially available systems in the rapidly growing space of ultra-low-field MRI. With reduced costs and more lenient siting requirements, ultra-low-field systems are being developed by research groups around the world with various designs for both brain (O'Reilly et al., 2019; McDaniel et al., 2019; He et al., 2020; Liu et al., 2021; Cooley et al., 2021; Guallart-Naval et al., 2022) and whole-body imaging (Zhao et al., 2024). A unique aspect of ultra-low-field MR systems is the ability to make them portable and able to operate in a normal room, by the bedside in a clinical setting (Mazurek et al., 2021; Balaji et al., 2024), in resource-constrained environments (Chetcuti et al., 2022; Abate et al., 2024) or even in an ambulance during active transport (Roberts et al., 2023). However, this also means that the environment can change from scan to scan, including both temperature and electromagnetic interference. The latter is, in particular, an issue that is unique to portable MR systems and a growing field of research. The fact that the scanner may be located in a non-controlled environment should be considered as a reproducibility aspect in future studies.

Limitations and further work

We only investigated reliability and correspondence of tissue volumes and underlying segmentations, obtained using a single pre-processing algorithm (SynthSeg+; Billot et al., 2023). A much wider range of pre-processing and analysis approaches could be applied to systematically explore their impact on the reliability and correspondence of derived imaging phenotypes (e.g., Heinen et al., 2016; Sederevičius et al., 2021). However, raw images used in this study are publicly available (*upon publication*), and we envisage that they will constitute a useful resource for researchers interested in developing and validating other pre-processing and analysis approaches for 64 mT MRI.

Additionally, our data and conclusions are limited to healthy adult participants, and additionally tempered by a modest sample size. Acquisition and open sharing or similar paired high- and ultra-low-field scans from paediatric and/or clinical populations, as well as similar healthy adult scans, will help accelerate development and validation of analytic approaches for ultra-low-field imaging. For example, larger datasets could help establish "confidence intervals" and re-calibration values for ultra-low-field morphometric estimates, relative to high-field counterparts. Still, exact correspondence to 3 T reference measurements may not be required for many applications of 64 mT morphometry, including for example longitudinal assessment of neurodevelopmental and/or clinical change within individuals, where measures at new time-points can be referenced to prior scans.

We note that our data are generally of very high quality, thanks in part to a focus on healthy adult participants who can lie still in the scanner; correspondence is likely to be lower within paediatric, elderly and/or clinical populations, who are likely to exhibit increased head motion (Madan, 2018; Padormo et al., 2023). Additionally, we used image processing tools developed for healthy adult participants (i.e., SynthSeg+; Billot et al., 2023), which may not perform as well on alternative populations. However, these limitations may be offset by developments in ultra-low-field imaging hardware and software, as well as software used for image enhancement and pre-processing.

Our results indicate greater ultra-low-field reliability and correspondence to high-field MRI of tissue volumes, than of the underlying regional segmentations. This is unsurprising, as local errors in the labeling of individual voxels are compounded when quantifying the overlap of segmentation labels, but may cancel out when calculating an aggregate measure of volume. For example, a structure may be assigned the same number of voxels in two scans, but these voxels will not necessarily overlap between scans (Kahhale et al., 2023). Going forward, for a more systematic assessment of the spatial overlap of segmentations, additional measures may be useful (Reinke et al., 2024). We also note that 3 T segmentations and corresponding volumes were used as reference, although these may show imperfections as well. While all segmentations were visually quality-controlled, further work on the correspondence of morphometric measures across field strengths would benefit from paired high- and ultra-low-field data with manually-labeled, "gold-standard" segmentations (e.g., Klein and Tourville, 2012).

The current study has only investigated anatomical imaging sequences as these are currently the most widely used at ultra-low field. The low SNR makes it difficult to utilize techniques which require rapid imaging, such as perfusion or functional imaging. Some image contrasts are difficult to achieve simply due to the MR physics. BOLD contrast, as used in functional MR, currently rely on contrast change due to susceptibility changes in oxygenated blood, an effect which is vanishingly small at ultra-low field. Functional imaging studies at ultra-low field strength would therefore have to consider alternative methods of probing brain function. We note that the Hyperfine Swoop 64 mT scanner can currently be used to acquire limited diffusion MRI data, including a single-orientation diffusion-weighted volume and an apparent diffusion coefficient (ADC) map; reliability and of ultra-low-field diffusion MRI, and correspondence to 3 T, is an avenue for future work.

Another aspect for wider applicability is the anatomical coverage of ultra-low-field systems. Many are currently built for a specific region, such as the brain (McDaniel et al., 2019; O'Reilly et al., 2019; He et al., 2020; Cooley et al., 2021; Liu et al., 2021; Kimberly et al., 2023) or extremities (Guallart-Naval et al., 2022), as this can make the system more affordable and may result in optimal performance for a given application. However, recent advancements have shown the possibility of performing whole body imaging on a 50 mT MR system (Zhao et al., 2024).

Conclusion

Volume estimates from portable ultra-low-field MRI scans show excellent test-retest reliability and correspondence to high-field counterparts. Reliability and correspondence to 3 T were highest for T_2w 64 mT scans reconstructed using MRR. Our results pave the way for clinical applications of ultra-low-field imaging morphometry, such as modeling of individual deviations from the norm, across development and disease. In particular, the excellent between-scanner reliability of 64 mT morphometry and high correspondence to 3 T augurs well for the pooling of ultra-low-field MRI data across sites to help address issues of brain health at scale across the globe.

Methods

Design

In total, 23 healthy adult participants were recruited. To obtain a sample balanced for age and sex, a minimum of 2 male and 2 female participants were recruited from each of the following 5 age-strata: 20-29 years, 30-39 years, 40-49 years, 50-59 years and 60-69 years. Basic demographics including sex, and year and month of birth, were collected. Exclusion criteria were contraindications to MRI (including claustrophobia), major neurological or psychiatric conditions, or other health issues known to affect the brain. Screening was conducted using the standard MRI safety questionnaire and through self-report of medical history. Participants provided informed written consent for all aspects of the study. The study was approved by the King's College London human Research Ethics Committee (Reference: HR/DP-21/22-26481).

MRI data acquisition

MRI data for each participant were acquired during two scanning sessions, held no more than 36 days apart. For a distribution of time durations between the two sessions, see supplementary Figure S2. During the first session, at the Centre for Neuroimaging Sciences (CNS; Institute of Psychiatry, Psychology and Neuroscience, King's College London), participants were scanned for 45 minutes on a GE SIGNA Premier 3 T high-field scanner with a 48-channel head coil (GE Healthcare, Waukesha, WI), followed immediately by a 1-hour protocol on a Hyperfine Swoop 64 mT ultra-low-field scanner (Hyperfine Inc., Guildford, CT). During the second session, at the Evelina Newborn Imaging Centre (ENIC; St. Thomas's Hospital, Guys and St. Thomas's NHS Foundation Trust), participants were scanned for 1 hour on a second Hyperfine Swoop system, identical to the first, and running the same software version (rc8.6.0). In all three scanners, Pearltec inflatable pads were used to minimize head motion (Pearl Technology AG, Switzerland).

The 3 T high-field protocol included T_1w (MPRAGE) and T_2w scans, acquired at $1\times1\times1$ mm^3 resolution. 64 mT protocols included both T_1w and T_2w scans, acquired using non-isotropic product

sequences (T_1w: 1.6×1.6×5 mm^3; T_2w: 1.5×1.5×5 mm^3) with high resolution in the axial, sagittal or coronal plane, respectively referred to as AXI, SAG, and COR. Additionally, a research agreement with Hyperfine Inc. enabled acquisition of a custom isotropic sequence (T_1w: 2.3×2.3×2.3 mm^3, T_2w: 2.2×2.2×2.2 mm^3), referred to as ISO. All scans were acquired on both scanners, except T_1w ISO scans, which were only acquired during the first scanning session at the CNS (the second scanning session at ENIC instead included an axial T_2-FLAIR scan of similar duration, not included in these analyses). For further details regarding the sequences, including acquisition times, see supplementary Table S1 and main Fig. 1A.

Pre-processing

We used multi-resolution registration (Deoni et al., 2022) to iteratively co-register and average the three orthogonal non-isotropic 64 mT scans (AXI, COR, SAG) into a single "MRR" scan, of a higher effective resolution (T_1w: 1.6×1.6×1.6 mm^3, T_2w: 1.5×1.5×1.5 mm^3). This was implemented using the *antsMultivariateTemplateConstruction2.sh* function from ANTs (Avants et al., 2009) (Fig. 1B).

We then used SynthSeg+ (Billot et al., 2023), including both "robust" segmentation and cortical parcellation, resulting in 98 regional labels for each scan, and corresponding volume estimates. We additionally grouped most labels into four categories by tissue class, including cortical grey matter (GM_{cort}), subcortical grey matter (GM_{subc}), white matter (WM) and cerebrospinal fluid (CSF). For details of the names of individual SynthSeg+ structures and their assignment to tissue classes, see supplementary Table S4.

To assess the between-scanner test-retest reliability of 64 mT segmentations and their correspondence to 3 T scans at the voxel level, we rigid-registered each 64 mT scan to the 3 T scan of the corresponding contrast, using FSL FLIRT (Jenkinson and Smith, 2001), and subsequently applied the transform to segmentation outputs (with nearest-neighbour interpolation). This resulted in each participant's test-retest 64 mT segmentations also being co-registered.

We visually assessed the quality of all individual scans, segmentations and registrations. Raw scans and segmentations were visually of high quality; however, some rigid registrations of 64 mT to 3 T scans appeared imperfect (see supplementary Table S5). In the main text we report results from all scans; for a sensitivity analysis quantifying Dice overlap of segmentations (affected by registration quality) in a subset of scans with high-quality registrations, see supplementary Table S6.

Analysis

Most analyses were repeated using both 4 "global" aggregate tissue volumes and corresponding segmentation masks, as well as 98 "local" SynthSeg+ structures, on both T_1w and T_2w contrasts, across all 64 mT resolutions. Correspondence of 64 mT to 3 T data was assessed using 64 mT scans acquired at the CNS, on the same day as (and immediately after) 3 T scans.

We assessed between-scanner test-retest reliability of volume estimates across participants using the intraclass correlation coefficient ICC(3,1) (Chen et al., 2018), referred to as ICC. We assessed correspondence of volumes from 64 mT and 3 T scans using both Pearson's r, to quantify linear correlation; and Lin's concordance correlation coefficient (CCC), to quantify exact agreement (or alignment with the $x = y$ identity line). We used the Dice overlap of segmented structures within participants to quantify the reliability of 64 mT segmentations ($Dice_{rel}$) and their correspondence to 3 T segmentations ($Dice_{cor}$). We also repeated Dice overlap analyses only within subjects with high-quality rigid registrations of 64 mT to 3 T scans (and consequently also 64 mT scans to each other); see supplementary Tables S5 and S6.

To quantify the extent to which 64 mT scans tend to under- or over-estimate tissue volumes relative to 3 T, we quantified the percentage difference in volumes relative to high-field data (supplementary Table S2).

To assess whether larger and more reliable structures show higher correspondence to 3 T scans, we quantified Spearman's rank correlations between local maps of 64 mT reliability, correspondence to 3 T, and median regional volume from 3 T scans. This analysis was limited to 64 mT T_2w MRR scans.

We also used linear models to estimate changes in tissue volume as a function of age, covarying for sex, using both 3 T and 64 mT scans. We quantified the relationship across field strengths between

regional rates of change (t-statistics) and variance explained by age and sex (r^2) using Pearson's r and Lin's CCC (supplementary Table S3).

Finally, we extracted residuals from these linear models, to obtain individual tissue volumes adjusted for effects of participant age and sex (for T_2w 3 T scans and T_2w 64 mT MRR scans). We then computed robust-scaled adjusted volumes within each region and across participants, to obtain non-parametric equivalents of the Z-score (Leys et al., 2013), a measure of individual deviation from the norm. We then identified the participant and region showing the greatest deviation from the median in T_2w high-field scans, and visualized the corrected volume of this region from all three scanners, across participants. Additionally, we quantified the reliability and correspondence of all regional deviations in this participant, using ICC, Pearson's r and Lin's CCC.

Data availability

All raw imaging data used in this manuscript will be made available *upon publication.*

Code availability

All code used for pre-processing and statistical analysis will be made available *upon publication.*

Author contributions

F.V., C.B., R.L., S.C.R.W. designed the study. F.P., T.C.W., D.J.W., F.D.A., E.L. implemented MRI sequences and provided MRI technical support. F.V., C.B., P.C., T.A. acquired the data. F.V., N.J.B., L.B. pre-processed the data. F.V., U.B. performed quality control of the data. F.V. performed analyses. T.C.B., A.V., S.C.L.D., R.J.M. provided conceptual support and advice. F.V., C.B., E.L. wrote the manuscript. All authors discussed results and commented on the manuscript.

Acknowledgements

This study was supported by a National Institute for Health and Care Research (NIHR) Biomedical Research Centre Early Career Investigator Award to F.V., and by the Bill and Melinda Gates Foundation UNITY project [INV-032788; INV-047888]. T.A. receives funding from the MRC via a Senior Clinical Fellowship [MR/Y009665/1] and the MRC Centre for Neurodevelopmental Disorders, King's College London [MR/N026063/1]. A.V.V. is funded by the National Institute for Health and Care Research as NIHR Clinical Lecturer and supported by the NIHR Maudsley Biomedical Research Centre and NIHR HealthTech Research Centre in Brain Health at King's College London and South London and Maudsley NHS Foundation Trust. S.C.R.W. is funded by the NIHR Biomedical Research and HealthTech Centres at the South London & Maudsley Hospital.

References

Abate, F., A. Adu-Amankwah, K. Ae-Ngibise, F. Agbokey, V. Agyemang, C. Agyemang, [...], S. Deoni, and S. Williams. (2024). UNITY: A Low-Field Magnetic Resonance Neuroimaging Initiative to Characterize Neurodevelopment in Low and Middle-Income Settings. *Developmental Cognitive Neuroscience*: 101397. https://doi.org/10.1016/j.dcn.2024.101397 .

Arnold, T.C., C.W. Freeman, B. Litt, and J.M. Stein. (2023). Low-field MRI: Clinical promise and challenges. *Journal of Magnetic Resonance Imaging 57*(1): 25–44. https://doi.org/10.1002/jmri.28408 .

Arnold, T.C., D. Tu, S.V. Okar, G. Nair, S. By, K.D. Kawatra, T.E. Robert-Fitzgerald, L.M. Desiderio, M.K. Schindler, R.T. Shinohara, D.S. Reich, and J.M. Stein. (2022). Sensitivity of portable low-field magnetic resonance imaging for multiple sclerosis lesions. *NeuroImage: Clinical* 35: 103101. https://doi.org/10.1016/j.nicl.2022.103101 .

Avants, B.B., N. Tustison, and H. Johnson. (2009). Advanced Normalization Tools (ANTS). *Insight j 2*(365): 1–35 .

Balaji, S., N. Wiley, M.E. Poorman, and S.H. Kolind. (2024). Low-field MRI for use in neurological diseases. *Current Opinion in Neurology 37*(4): 381. https://doi.org/10.1097/WCO.0000000000001282 .

Baljer, L., Y. Zhang, N.J. Bourke, K.A. Donald, L.E. Bradford, J.E. Ringshaw, S.R. Williams, S.C. Deoni, S.C. Williams, K.S.S. Team, F. Váša, and R.J. Moran. (2024). Ultra-low-field paediatric MRI in low- and middle-income countries: super-resolution using a multi-orientation U-Net.

Bethlehem, R.A.I., J. Seidlitz, S.R. White, J.W. Vogel, K.M. Anderson, C. Adamson, [...], E.T. Bullmore, and A.F. Alexander-Bloch. (2022). Brain charts for the human lifespan. *Nature 604*(7906): 525–533. https://doi.org/10.1038/s41586-022-04554-y .

Billot, B., C. Magdamo, Y. Cheng, S.E. Arnold, S. Das, and J.E. Iglesias. (2023). Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets. *Proceedings of the National Academy of Sciences 120*(9): e2216399120. https://doi.org/10.1073/pnas.2216399120 .

Chen, G., P.A. Taylor, S.P. Haller, K. Kircanski, J. Stoddard, D.S. Pine, E. Leibenluft, M.A. Brotman, and R.W. Cox. (2018). Intraclass correlation: Improved modeling approaches and applications for neuroimaging. *Human Brain Mapping 39*(3): 1187–1206. https://doi.org/10.1002/hbm.23909 .

Chetcuti, K., C. Chilingulo, M.S. Goyal, L. Vidal, N.F. O'Brien, D.G. Postels, K.B. Seydel, and T.E. Taylor. (2022). Implementation of a Low-Field Portable MRI Scanner in a Resource-Constrained Environment: Our Experience in Malawi. *American Journal of Neuroradiology 43*(5): 670–674. https://doi.org/10.3174/ajnr.A7494 .

Cooley, C.Z., P.C. McDaniel, J.P. Stockmann, S.A. Srinivas, S.F. Cauley, M. Śliwiak, C.R. Sappo, C.F. Vaughn, B. Guerin, M.S. Rosen, M.H. Lev, and L.L. Wald. (2021). A portable scanner for magnetic resonance imaging of the brain. *Nature Biomedical Engineering 5*(3): 229–239. https://doi.org/10.1038/s41551-020-00641-5 .

Cooper, R., R.A. Hayes, M. Corcoran, K.N. Sheth, T.C. Arnold, J.M. Stein, D.C. Glahn, and M. Jalbrzikowski. (2024). Bridging the gap: improving correspondence between low-field and high-field magnetic resonance images in young people. *Frontiers in Neurology* 15: 1339223. https://doi.org/10.3389/fneur.2024.1339223 .

Deoni, S.C.L., J. O'Muircheartaigh, E. Ljungberg, M. Huentelman, and S.C.R. Williams. (2022). Simultaneous high-resolution T2 -weighted imaging and quantitative T2 mapping at low magnetic field strengths using a multiple TE and multi-orientation acquisition approach. *Magnetic Resonance in Medicine 88*(3): 1273–1281. https://doi.org/10.1002/mrm.29273 .

Elliott, M.L., L.C. Hanford, A. Hamadeh, T. Hilbert, T. Kober, B.C. Dickerson, R.W. Mair, M.C. Eldaief, and R.L. Buckner. (2023). Brain morphometry in older adults with and without dementia using extremely rapid structural scans. *NeuroImage* 276: 120173. https://doi.org/10.1016/j.neuroimage.2023.120173 .

Fukunaga, M., T.Q. Li, P. van Gelderen, J.A. de Zwart, K. Shmueli, B. Yao, J. Lee, D. Maric, M.A. Aronova, G. Zhang, R.D. Leapman, J.F. Schenck, H. Merkle, and J.H. Duyn. (2010). Layer-specific variation of iron content in cerebral cortex as a source of MRI contrast. *Proceedings of the National Academy of Sciences 107*(8): 3834–3839. https://doi.org/10.1073/pnas.0911177107 .

Guallart-Naval, T., J.M. Algarín, R. Pellicer-Guridi, F. Galve, Y. Vives-Gilabert, R. Bosch, E. Pallás, J.M. González, J.P. Rigla, P. Martínez, F.J. Lloris, J. Borreguero, A. Marcos-Perucho, V. Negnevitsky, L. Martí-Bonmatí, A. Ríos, J.M. Benlloch, and J. Alonso. (2022). Portable magnetic

resonance imaging of patients indoors, outdoors and at home. *Scientific Reports 12*(1): 13147. https://doi.org/10.1038/s41598-022-17472-w .

Haller, S., P. Falkovskiy, R. Meuli, J.P. Thiran, G. Krueger, K.O. Lovblad, T. Kober, A. Roche, and B. Marechal. (2016). Basic MR sequence parameters systematically bias automated brain volume estimation. *Neuroradiology 58*(11): 1153–1160. https://doi.org/10.1007/s00234-016-1737-3 .

He, Y., W. He, L. Tan, F. Chen, F. Meng, H. Feng, and Z. Xu. (2020). Use of 2.1 MHz MRI scanner for brain imaging and its preliminary results in stroke. *Journal of Magnetic Resonance* 319: 106829. https://doi.org/10.1016/j.jmr.2020.106829 .

Heinen, R., W.H. Bouvy, A.M. Mendrik, M.A. Viergever, G.J. Biessels, and J.d. Bresser. (2016). Robustness of Automated Methods for Brain Volume Measurements across Different MRI Field Strengths. *PLOS ONE 11*(10): e0165719. https://doi.org/10.1371/journal.pone.0165719 .

Iglesias, J.E., B. Billot, Y. Balbastre, A. Tabari, J. Conklin, R. Gilberto González, D.C. Alexander, P. Golland, B.L. Edlow, and B. Fischl. (2021). Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. *NeuroImage* 237: 118206. https://doi.org/10.1016/j.neuroimage.2021.118206 .

Iglesias, J.E., R. Schleicher, S. Laguna, B. Billot, P. Schaefer, B. McKaig, J.N. Goldstein, K.N. Sheth, M.S. Rosen, and W.T. Kimberly. (2023). Quantitative Brain Morphometry of Portable Low-Field-Strength MRI Using Super-Resolution Machine Learning. *Radiology 306*(3): e220522. https://doi.org/10.1148/radiol.220522 .

Jenkinson, M. and S. Smith. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis 5*(2): 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6 .

Jordanova, K.V., M.N. Martin, S.E. Ogier, M.E. Poorman, and K.E. Keenan. (2023). In vivo quantitative MRI: T1 and T2 measurements of the human brain at 0.064 T. *Magnetic Resonance Materials in Physics, Biology and Medicine 36*(3): 487–498. https://doi.org/10.1007/s10334-023-01095-x .

Kahhale, I., N.J. Buser, C.R. Madan, and J.L. Hanson. (2023). Quantifying numerical and spatial reliability of hippocampal and amygdala subdivisions in FreeSurfer. *Brain Informatics 10*(1): 9. https://doi.org/10.1186/s40708-023-00189-5 .

Kimberly, W.T., A.J. Sorby-Adams, A.G. Webb, E.X. Wu, R. Beekman, R. Bowry, S.J. Schiff, A. de Havenon, F.X. Shen, G. Sze, P. Schaefer, J.E. Iglesias, M.S. Rosen, and K.N. Sheth. (2023). Brain imaging with portable low-field MRI. *Nature Reviews Bioengineering 1*(9): 617–630. https://doi.org/10.1038/s44222-023-00086-w .

Klein, A. and J. Tourville. (2012). 101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol. *Frontiers in Neuroscience* 6. https://doi.org/10.3389/fnins.2012.00171 .

Krupinski, E.A., D. Harris, L.R. Arlinghaus, J. Schlemper, and M. Sofka. (2023). Image quality assessment of advanced reconstruction algorithm for point-of-care MRI scanner. *Journal of Medical Imaging 10*(S1): S11913. https://doi.org/10.1117/1.JMI.10.S1.S11913 .

Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology 49*(4): 764–766. https://doi.org/10.1016/j.jesp.2013.03.013 .

Lin, H., M. Figini, F. D'Arco, G. Ogbole, R. Tanno, S.B. Blumberg, L. Ronan, B.J. Brown, D.W. Carmichael, I. Lagunju, J.H. Cross, D. Fernandez-Reyes, and D.C. Alexander. (2023). Low-field magnetic resonance image enhancement via stochastic image quality transfer. *Medical Image Analysis* 87: 102807. https://doi.org/10.1016/j.media.2023.102807 .

Liu, Y., A.T.L. Leong, Y. Zhao, L. Xiao, H.K.F. Mak, A.C.O. Tsang, G.K.K. Lau, G.K.K. Leung, and E.X. Wu. (2021). A low-cost and shielding-free ultra-low-field brain MRI scanner. *Nature Communications 12*(1): 7238. https://doi.org/10.1038/s41467-021-27317-1 .

Madan, C.R. (2018). Age differences in head motion and estimates of cortical morphology. *PeerJ* 6: e5176. https://doi.org/10.7717/peerj.5176 .

Madan, C.R. and E.A. Kensinger. (2017). Test–retest reliability of brain morphology estimates. *Brain Informatics 4*(2): 107–121. https://doi.org/10.1007/s40708-016-0060-4 .

Marquand, A.F., S.M. Kia, M. Zabihi, T. Wolfers, J.K. Buitelaar, and C.F. Beckmann. (2019). Conceptualizing mental disorders as deviations from normative functioning. *Molecular Psychiatry 24*(10): 1415–1424. https://doi.org/10.1038/s41380-019-0441-1 .

Marques, J.P., F.F. Simonis, and A.G. Webb. (2019). Low-field MRI: An MR physics perspective. *Journal of Magnetic Resonance Imaging 49*(6): 1528–1542. https://doi.org/10.1002/jmri.26637 .

Mazurek, M.H., B.A. Cahn, M.M. Yuen, A.M. Prabhat, I.R. Chavva, J.T. Shah, [...], W.T. Kimberly, and K.N. Sheth. (2021). Portable, bedside, low-field magnetic resonance imaging for evaluation of intracerebral hemorrhage. *Nature Communications 12*(1): 5119. https://doi.org/10.1038/s41467-021-25441-6 .

McDaniel, P.C., C.Z. Cooley, J.P. Stockmann, and L.L. Wald. (2019). The MR Cap: A single-sided MRI system designed for potential point-of-care limited field-of-view brain imaging. *Magnetic Resonance in Medicine 82*(5): 1946–1960. https://doi.org/10.1002/mrm.27861 .

Morey, R.A., E.S. Selgrade, H.R. Wagner II, S.A. Huettel, L. Wang, and G. McCarthy. (2010). Scan–rescan reliability of subcortical brain volumes derived from automated segmentation. *Human Brain Mapping 31*(11): 1751–1762. https://doi.org/10.1002/hbm.20973 .

Mulder, M.J., M.C. Keuken, P.L. Bazin, A. Alkemade, and B.U. Forstmann. (2019). Size and shape matter: The impact of voxel geometry on the identification of small nuclei. *PLOS ONE 14*(4): e0215382. https://doi.org/10.1371/journal.pone.0215382 .

O'Reilly, T., W.M. Teeuwisse, and A.G. Webb. (2019). Three-dimensional MRI in a homogenous 27cm diameter bore Halbach array magnet. *Journal of Magnetic Resonance* 307: 106578. https://doi.org/10.1016/j.jmr.2019.106578 .

Padormo, F., P. Cawley, L. Dillon, E. Hughes, J. Almalbis, J. Robinson, A. Maggioni, M.D.L.F. Botella, D. Cromb, A. Price, L. Arlinghaus, J. Pitts, T. Luo, D. Zhang, S.C.L. Deoni, S. Williams, S. Malik, J. OMuircheartaigh, S.J. Counsell, M. Rutherford, T. Arichi, A.D. Edwards, and J.V. Hajnal. (2023). In vivo T mapping of neonatal brain tissue at 64 mT. *Magnetic Resonance in Medicine 89*(3): 1016–1025. https://doi.org/10.1002/mrm.29509 .

Peters, A.M., M.J. Brookes, F.G. Hoogenraad, P.A. Gowland, S.T. Francis, P.G. Morris, and R. Bowtell. (2007). $T2^*$ measurements in human brain at 1.5, 3 and 7 T. *Magnetic Resonance Imaging 25*(6): 748–753. https://doi.org/10.1016/j.mri.2007.02.014 .

Reinke, A., M.D. Tizabi, M. Baumgartner, M. Eisenmann, D. Heckmann-Nötzel, A.E. Kavur, [...], and L. Maier-Hein. (2024). Understanding metric-related pitfalls in image analysis validation. *Nature Methods 21*(2): 182–194. https://doi.org/10.1038/s41592-023-02150-0 .

Roberts, D.R., T. McGeorge, D. Abrams, R. Hewitt, D. LeBlanc, W. Dennis, M. Rosenberg, S.A. Kasab, C. Holmstedt, M.V. Spampinato, S. Torres-Rosado, R. Ancrum, M. Haschker, and J. Harvey. (2023). Mobile point-of-care MRI demonstration of a normal volunteer in a telemedicine-equipped ambulance. *Journal of Stroke and Cerebrovascular Diseases 32*(10). https://doi.org/10.1016/j.jstrokecerebrovasdis.2023.107301 .

Rooney, W.D., G. Johnson, X. Li, E.R. Cohen, S.G. Kim, K. Ugurbil, and C.S. Springer Jr. (2007). Magnetic field and tissue dependencies of human brain longitudinal 1H2O relaxation in vivo. *Magnetic Resonance in Medicine 57*(2): 308–318. https://doi.org/10.1002/mrm.21122 .

Sederevičius, D., D. Vidal-Piñeiro, Sørensen, K. van Leemput, J.E. Iglesias, A.V. Dalca, D.N. Greve, B. Fischl, A. Bjørnerud, K.B. Walhovd, and A.M. Fjell. (2021). Reliability and sensitivity of two whole-brain segmentation approaches included in FreeSurfer – ASEG and SAMSEG. *NeuroImage* 237: 118113. https://doi.org/10.1016/j.neuroimage.2021.118113 .

Stanisz, G.J., E.E. Odrobina, J. Pun, M. Escaravage, S.J. Graham, M.J. Bronskill, and R.M. Henkelman. (2005). T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magnetic Resonance in Medicine 54*(3): 507–512. https://doi.org/10.1002/mrm.20605 .

Whelan, C.D., D.P. Hibar, L.S. van Velzen, A.S. Zannas, T. Carrillo-Roa, K. McMahon, G. Prasad, S. Kelly, J. Faskowitz, G. deZubiracay, J.E. Iglesias, T.G.M. van Erp, T. Frodl, N.G. Martin, M.J. Wright, N. Jahanshad, L. Schmaal, P.G. Sämann, and P.M. Thompson. (2016). Heritability and reliability of automatically segmented human hippocampal formation subregions. *NeuroImage* 128: 125–137. https://doi.org/10.1016/j.neuroimage.2015.12.039 .

Wonderlick, J.S., D.A. Ziegler, P. Hosseini-Varnamkhasti, J.J. Locascio, A. Bakkour, A. van der Kouwe, C. Triantafyllou, S. Corkin, and B.C. Dickerson. (2009). Reliability of MRI-derived cortical and subcortical morphometric measures: Effects of pulse sequence, voxel geometry, and parallel imaging. *NeuroImage 44*(4): 1324–1333. https://doi.org/10.1016/j.neuroimage.2008.10.037 .

Zhao, M., X. Liu, H. Liu, and K.K.L. Wong. (2020). Super-resolution of cardiac magnetic resonance images using Laplacian Pyramid based on Generative Adversarial Networks. *Computerized Medical Imaging and Graphics* 80: 101698. https://doi.org/10.1016/j.compmedimag.2020.101698 .

Zhao, Y., Y. Ding, V. Lau, C. Man, S. Su, L. Xiao, A.T.L. Leong, and E.X. Wu. (2024). Whole-body magnetic resonance imaging at 0.05 Tesla. *Science 384*(6696): eadm7168. https://doi.org/10.1126/science.adm7168 .