

SwiftSight–Brain Consistency Whitepaper

Consistency of automated brain volumetry across repeat scans,
scanners, field strengths, and acquisition protocols

Woojin Jung, Jay Suh, Seungwook Yang, Insoo Chung, Dowon Kim, Joonyoung Yang

1. Introduction

Brain volumetry — the quantitative measurement of regional and global brain structures from structural magnetic resonance imaging (MRI) — is increasingly used as a biomarker for the diagnosis and longitudinal monitoring of neurological disease, including Alzheimer’s disease, mild cognitive impairment, multiple sclerosis, temporal-lobe epilepsy, and traumatic brain injury [1]. For a volumetric measurement to be clinically actionable, it must be interpreted against an age- and sex-adjusted normative reference. Translating that requirement into routine practice calls for an automated solution that couples accurate structural measurement with a purpose-built normative reference of its own. SwiftSight-Brain — a deep-learning-based, US FDA 510(k)-cleared MR brain volumetry solution — is built on its own normative database framework, whose reference population and shifted-softplus percentile-modeling approach are described in the companion SwiftSight-Brain Normative Database Whitepaper [2].

A normative comparison, in turn, is only as trustworthy as the measurement feeding into it. If the same patient is scanned twice — on the same scanner minutes apart, on a different scanner, at a different field strength, or with a different acquisition protocol — the reported volumes should not change more than the morphology they are meant to track. Measurement variability arising from the imaging chain rather than from the patient’s physiology degrades the sensitivity of longitudinal monitoring and can produce shifts that mimic or mask true atrophy [3,

4]. Consistency, the reproducibility of a volumetric estimate under changing acquisition conditions, is therefore a prerequisite for clinical utility.

In practice, this is not merely a theoretical concern but an operational constraint that actively shapes how neuroimaging is scheduled and interpreted. When the imaging chain contributes variability that cannot be separated from true biological change, a clinician reading a follow-up scan cannot confidently attribute an apparent volume change to disease progression rather than to a change in imaging environment [5]. Consequently, to preserve the interpretability of serial comparisons, many imaging institutions adopt the pragmatic workaround of routing patients who require longitudinal monitoring back to the same MR scanner, preferably under the same acquisition protocol for every follow-up examination [6]. However, this same-scanner constraint is operationally fragile: it is frequently unachievable when a patient is referred across sites, when a scanner is upgraded or decommissioned mid-course, or when scheduling pressures force reassignment, and it inherently limits the ability to pool or compare volumetric measurements across institutions [7]. There is therefore an urgent need for a volumetry solution whose measurements remain consistent across scanners, field strengths, and acquisition protocols, so that longitudinal monitoring is driven by the biology of the patient rather than by the imaging environment that happened to be available.

This whitepaper illustrates and quantifies the measurement consistency of SwiftSight-Brain and benchmarks it against FastSurfer [8], a widely used deep-learning reimplement of the FreeSurfer volumetric pipeline [9, 10] used here as an established open-source reference. Consistency is characterized along four axes of variation a deployed volumetry tool encounters in practice: repeated acquisition on a fixed scanner (test-retest), a change in magnetic field strength (1.5T versus 3T), a change of scanner hardware (inter-scanner), and a change in in-plane spatial resolution. For each axis the guiding principle is variable isolation: subgroups are constructed so that, as far as physically possible, only the axis under study differs while other acquisition conditions are held constant. These properties are evaluated with two complementary datasets: a large retrospective real-world cohort, in which a single acquisition variable is isolated at a time (Sections 3.1–3.5 and 4.1–4.5), and a prospective traveling-subjects sub-study, in which the same individuals are imaged across many deliberately varied protocols on three scanner platforms (Sections 3.6 and 4.6).

2. How SwiftSight-Brain is designed for consistency

SwiftSight-Brain is engineered so that its volume measurements track anatomy, not acquisition. It is deployed as an integrated pipeline in which every image is first processed by SwiftMR — an AI-based image enhancement engine from AIRS Medical — and then segmented via SwiftSight-Brain. Consistency is a design goal built into both stages. Upstream, SwiftMR contributes to reducing the input image quality variability; downstream, the volumetry model is trained and structured to remain insensitive to the acquisition variability that remains. Each stage targets a source of acquisition variability that the empirical results in this whitepaper quantify.

Stage 1 — Input reconstruction (SwiftMR). Before any measurement, each image is processed by SwiftMR, a vendor-neutral deep-learning MR reconstruction built on a U-Net convolutional network that operates on the reconstructed DICOM image. By recovering the signal-to-noise ratio and improving spatial resolution, this upstream step narrows the variability the volumetry model must absorb — most directly the noise- and SNR-driven differences that accompany lower-field, accelerated, or higher-resolution protocols — before segmentation begins.

Because segmentation is itself sensitive to image noise, elevated noise can cause a segmentation to break down locally — dropping or mislabeling tissue that is truly present — so the error enters the measurement before any volume is computed. Reducing noise upstream therefore protects the segmentation itself, not merely the appearance of the image (Figure 1).

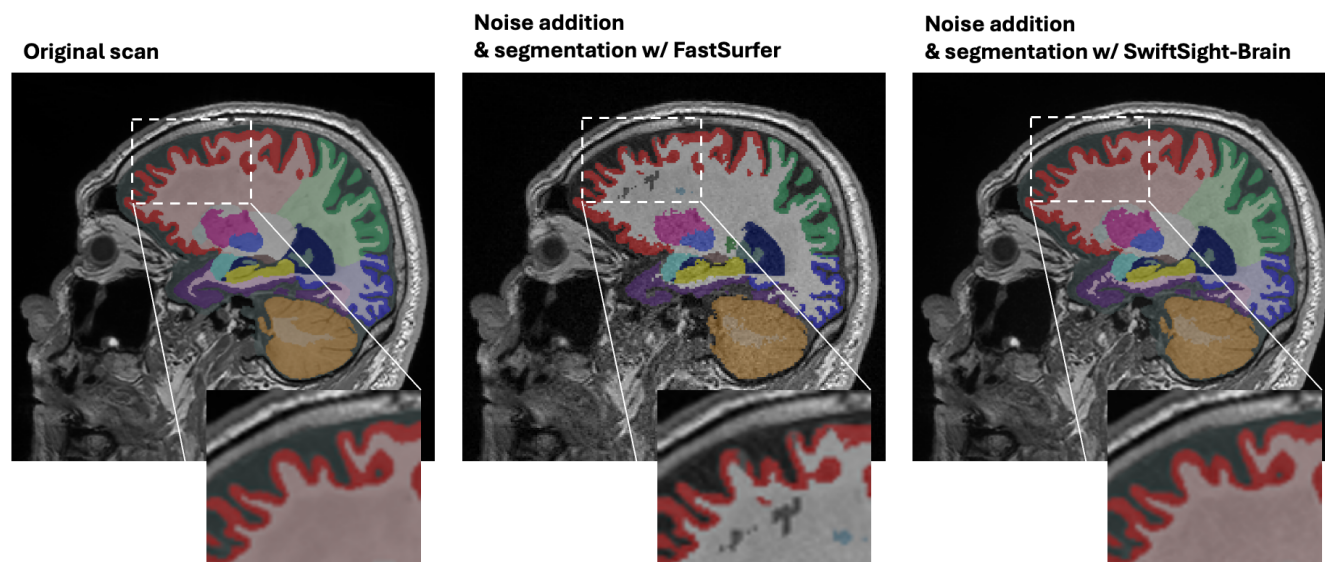


Figure 1. A single-case illustration of image noise effect on segmentation results. A representative slice from an original 3D T1 acquisition (left). After introducing simulated noise ($+1.5\sigma$), FastSurfer's segmentation developed focal dropouts in the cortical segmentation (center, inset), whereas applying SwiftMR reconstruction before SwiftSight-Brain segmentation keeps the mask intact at the same noise level (right).

Stage 2 — Reconstruction-invariant volumetry. On the SwiftMR-enhanced image, three complementary mechanisms keep the measurement stable across the acquisition variability that remains:

1. **Scan geometry standardization.** Every volume is reoriented and resampled to a single fixed, isotropic canonical geometry before segmentation, so in-plane resolution, slice thickness, and orientation are removed as variables and the model always operates on a common voxel grid. (This standardizes the sampling grid; it does not recover detail a thick-slice scan never captured — the appearance side is handled by mechanisms 2 and 3). This underlies the in-plane-resolution consistency in Section 4.5.
2. **Contrast, noise, and intensity invariance.** The model is trained with an intensity-domain augmentation regime — spanning contrast, brightness, structured intensity variations, and noise levels — so that its output is insensitive to the signal differences introduced by pulse-sequence timing (TE/TR/TI), and factors which affect SNR such as field strengths, etc. This underlies the cross-protocol robustness in the field-strength and traveling-subjects results.
3. **Acquisition diversity represented in training.** The model is trained on a large multi-site cohort (~28,000 volumes) spanning all three major MRI vendors (GE, Philips, Siemens) and both clinical field strengths (1.5T and 3T), so scanner, vendor, and field-strength variation is represented in-distribution and every acquisition is processed by the same model in the same way. This underlies the inter-scanner, inter-vendor, and field-strength consistencies.

These stages are mutually reinforcing: SwiftMR standardizes the input, the geometry step standardizes what can be standardized, and training instills invariance to the acquisition diversity that remains. Test-retest reproducibility — repeated scans of an unchanged subject on the same scanner — represents the measurement floor at which this combined robustness is most visible, since only sub-visible acquisition differences separate the repeats. It is this combination, rather than any single adjustment, that produces the consistency quantified here.

3. Methods

3.1 Volumetric measurements and comparator

Two brain MR image segmentation engines were applied to an identical set of 3D T1-weighted MRI data. Both report volumes for a selected, harmonized 19-structure atlas spanning global measures (whole brain, cortical grey matter, cerebral white matter), lobar grey matter (frontal, temporal, parietal, occipital, insula, cingulate), subcortical grey-matter nuclei (hippocampus, amygdala, thalamus, caudate, putamen, pallidum, accumbens), cerebellum and brain stem, and lateral ventricle. Both engines start from the identical acquired 3D T1-weighted series, and each is evaluated as it is used in clinical practice, so any difference in reproducibility reflects the complete measurement solution rather than differences in the acquired input. The evaluation cohort is predominantly normal-appearing but includes examinations from patients with a range of diagnosed neurological conditions, so the reported consistency reflects genuine morphological variability across subjects. SwiftSight-Brain reports peer-normative volume percentiles and relative structural atrophy against a normative database; it does not perform disease-specific classification or disease-stratified lookup, and no such stratification was applied here.

3.2 Consistency metric

Consistency is quantified with the coefficient of variation (CoV): the ratio of the standard deviation to the mean of repeated measurements of the same structure, expressed as a percentage. A lower CoV denotes a more reproducible measurement. Because the per-group CoV distribution is heavy right-tailed — most groups cluster near zero while a small number of segmentation-failure outliers reach very high values — the summary statistic across groups is a 20% trimmed mean, which is robust to those outliers (and agrees with the median to within 5%); comparisons are made with rank-based (non-parametric) tests.

3.3 Subgroup construction and variable control

Each consistency dimension is defined by a subgroup of examinations in which a single acquisition variable changes. Two conditions govern how repeated scans within a subject are aggregated. First, condition-mean aggregation: for each subject, all scans acquired under a given level of the target variable are averaged into one condition mean, and the CoV is computed across the condition means rather than across individual scans — so that within-condition repeat variability (e.g., 3T-to-3T noise) does not contaminate the between-condition estimate. Second, single-scanner isolation for the in-plane-resolution dimension: each group is restricted to a single subject on a single physical scanner so that only the target imaging parameter varies. For each of the 19 structures the per-group CoV values are summarized by their trimmed mean across groups, and a dimension-level summary is the mean across structures.

3.4 Quality control

All examinations carry a quality flag from upstream review. Prior to any CoV computation, every scan flagged as unsuitable — including segmentation failures, motion, unsuitable contrast, non-3D-T1, file corruption, and derived/averaged reconstructions — was excluded uniformly across all subgroups. Failed segmentations produce grossly outlying volumes (in the limit, a zero or near-zero structure) that would otherwise dominate a CoV.

3.5 Statistical testing

Because FastSurfer and SwiftSight-Brain are applied to the identical scans, the two engines are compared with a paired test. For each acquisition group (the subject-and-scanner unit for test-retest and resolution; the subject for field strength and inter-scanner), a per-structure CoV is computed for each engine, and a (group, structure) pair is retained only when both engines yield a valid CoV. For each structure, the difference is assessed with the two-sided Wilcoxon signed-rank test. The more reproducible engine is determined by the trimmed mean, and p-values are Holm-corrected across all individual structures within each dimension. As a supplementary reliability metric, the intraclass correlation coefficient (ICC; one-way random-effects, single measurement) was also computed per structure for each engine in the test-retest scenario.

3.6 Prospective traveling-subjects sub-study

To complement the retrospective cohort with prospectively acquired, deliberately controlled data, 3D T1 images from six subjects were acquired from GE, Philips, and Siemens scanner platforms while acquisition parameters were systematically varied — flip angle, repetition and inversion time, echo time, preparation/recovery timing, bandwidth, and reconstruction option — for a total of 256 acquisitions. FastSurfer and SwiftSight-Brain were applied to the identical image series, so any difference again reflects the complete measurement solution rather than differences in the acquired input. The comparison in this case was restricted to the nine structures for which volumetric measurements are most commonly reported: whole brain, cerebral cortex, cerebellum, lateral ventricles, the four lobar cortices, and the hippocampus. Unlike the retrospective design, which isolates one acquisition variable at a time, this sub-study allows many parameters to vary together, emulating the protocol heterogeneity encountered when a patient is imaged across sites and scanners.

For each subject, the CoV of every structure was computed across that subject's acquisitions (separately for each engine), and the engines were compared at the subject level ($n = 6$; two-sided Wilcoxon signed-rank). Also, acquisitions were classified into a routine tier (standard clinical 3D-T1: flip angle 7–15°, standard reconstruction including compressed sensing, and all sampled bandwidths; $n = 220$) and a stress tier (deliberately non-routine settings: out-of-phase reconstruction, non-standard k-space ordering, and out-of-range flip angles of 6° and 20°; $n = 36$) to assess models' behaviors in differently varying imaging parameter settings.

4. Results

4.1 Summary across consistency dimensions

For each consistency dimension, Table 1 summarizes the final subgroup size utilized during consistency analysis, along with the per-structure mean CoV (%) values for both segmentation engines and the statistical significance. On QC-filtered, paired data, SwiftSight-Brain is significantly more reproducible than FastSurfer across all four dimensions — test-retest, field strength, inter-scanner, and in-plane resolution.

Dimension	Subgroup size	FastSurfer mean CoV (%)	SwiftSight mean CoV (%)	Statistical significance
Test-retest (repeated scan under same conditions)	4,565	0.64	0.45	$p < 0.001$
Field strength (1.5T vs 3T)	189	2.92	1.67	$p < 0.001$
Inter-scanner (two different scanners within 30 days)	327	1.99	1.34	$p < 0.001$
In-plane resolution	1,019	0.61	0.49	$p < 0.001$

Table 1. Overview of the consistency analysis: sample size and overall consistency by acquisition-variation dimension. For each of the four consistency dimensions, the number of subjects, overall mean per-structure CoVs for FastSurfer and SwiftSight-Brain, and statistical significance are shown. SwiftSight-Brain displays overall lower mean CoV in every dimension.

4.2 Test-retest (repeat scans under same conditions)

The test-retest subgroup (4,565 subjects after quality control) establishes the consistency floor: the variability observed when nothing about the acquisition changes. On the identical groups, SwiftSight-Brain has the lower mean CoV in all 19 structures, and the difference is significant after Holm correction (atlas-wide mean 0.45% vs 0.64% for FastSurfer, a 30% reduction; whole brain 0.18% vs 0.22%; cerebral white matter 0.25% vs 0.60%; thalamus 0.32% vs 0.56%).

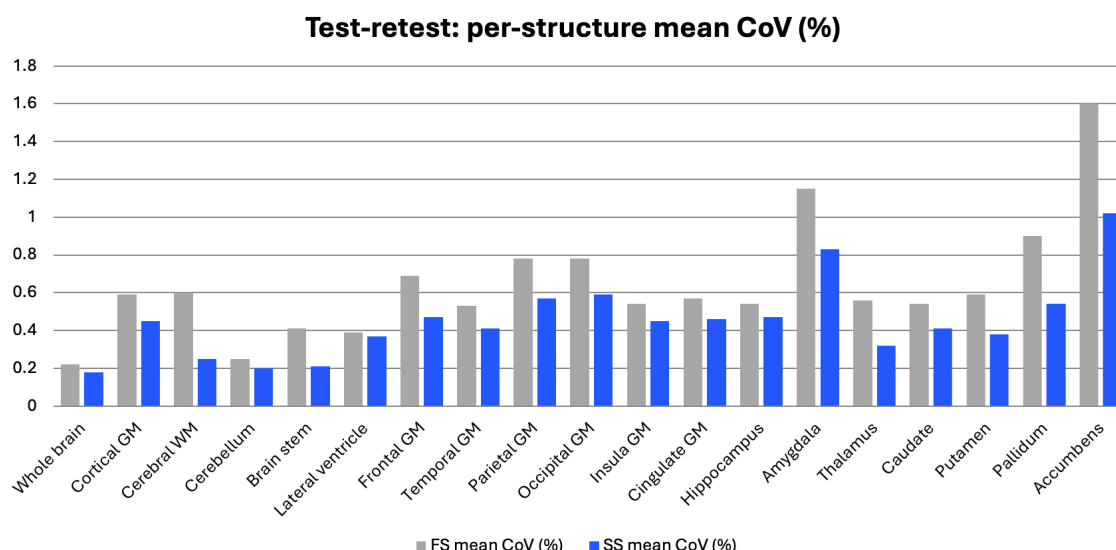


Figure 2. Test-retest consistency. Per-structure mean CoV across the 19-region atlas; FastSurfer (grey), SwiftSight-Brain (blue); lower = more reproducible. Same subject, same scanner and protocol, repeated within a 30-day window (n = 4,565). SwiftSight-Brain has a lower mean CoV in all regions, with atlas-wide mean at 0.64% (FastSurfer) and 0.45% (SwiftSight-Brain) (p < 0.001).

In the reliability metric conventionally reported for volumetry software, the same pattern holds: test-retest ICC was high for both engines and higher for SwiftSight-Brain in all 19 structures (mean 0.98 vs 0.96 for FastSurfer; SwiftSight-Brain range 0.97–1.00), at or above the upper end of values reported for commercial volumetry packages in the literature (typically ~0.83–0.98). Because those figures derive from different cohorts and study designs, the comparison is indicative rather than a controlled head-to-head.

4.3 Field-strength-associated variability (1.5T vs 3T)

A change in magnetic field strength cannot be studied in complete isolation, because field strength is an intrinsic property of the magnet: a scanner is either 1.5T or 3T, and no single scanner acquires the same examination at both field strengths. Any 1.5T-versus-3T comparison therefore also involves a change of scanner hardware, so the dimension is reported as field-strength-associated variability rather than a pure field-strength effect.

For each of 189 subjects scanned at both field strengths within a 30-day window, the scans at each field strength were averaged into a field-level mean before computing CoV. On the paired comparison, SwiftSight-Brain is significantly more consistent in 18 of 19 structures (atlas-wide mean 1.67% vs 2.92% for FastSurfer — a 43% reduction; Holm-corrected Wilcoxon), with the pallidum showing no significant difference; the largest gains were seen in cerebral white matter (1.74% vs 4.52%), the accumbens, and occipital grey matter.

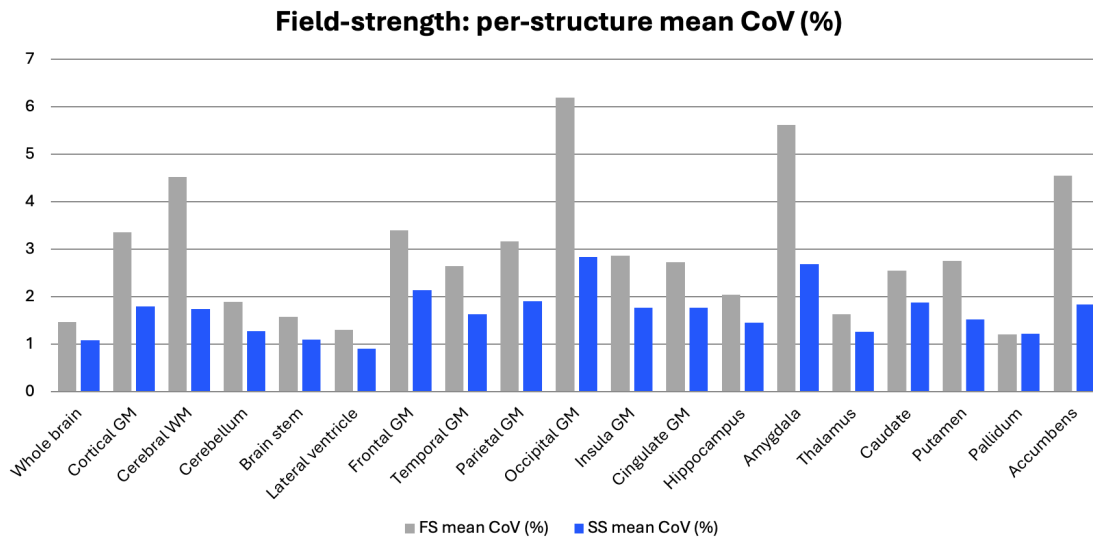


Figure 3. Consistency across field strength (1.5T vs 3T). Per-structure mean CoV; the same subjects imaged at both 1.5T and 3T within 30 days ($n = 189$), other conditions matched. SwiftSight-Brain is significantly more consistent in 18 of 19 regions, with the exception of the pallidum — where no significant difference was found. Atlas-wide mean at 2.92% for FastSurfer and 1.67% for SwiftSight-Brain ($p < 0.001$); the largest consistency gains are seen in cerebral white matter, the accumbens, and occipital grey matter.

4.4 Inter-scanner variability

The inter-scanner subgroup comprises 327 subjects, each imaged on two or more different scanner models within a 30-day window; this is among the harshest perturbations — vendor, model, and often protocol change together — so a scanner change introduces substantial genuine acquisition variability that affects both engines. Even so, SwiftSight-Brain is significantly more reproducible in 18 of 19 structures, led by cerebral white matter (1.14% vs 2.54%), the accumbens, and putamen. The two engines are equivalent for whole-brain volume, where both approach their consistency floor ($\sim 0.87\%$ CoV) and differ by only a few hundredths of a percentage point.

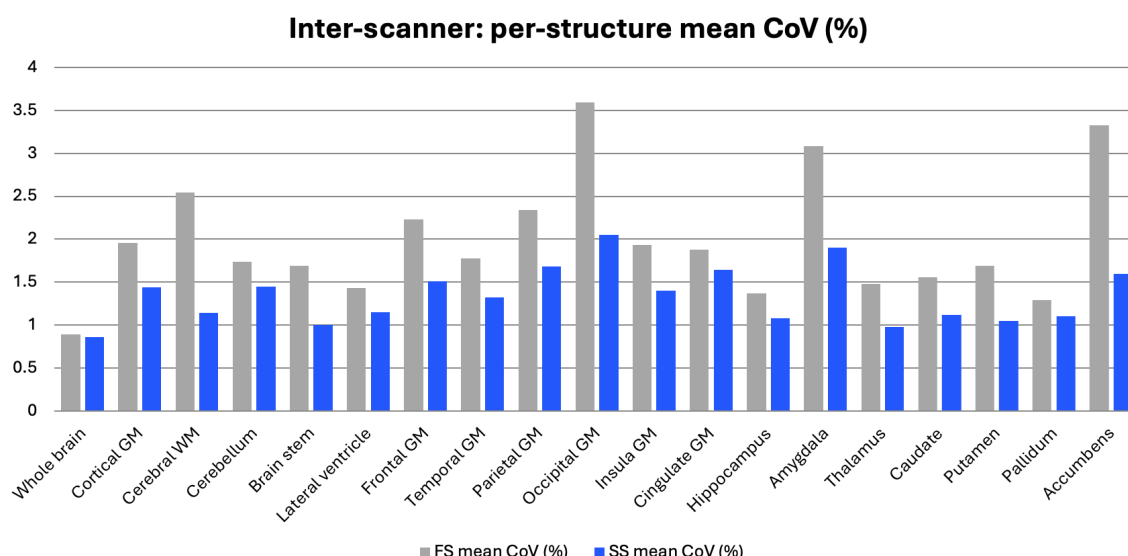


Figure 4. Inter-scanner consistency. Same subjects imaged on two or more distinct scanner models within 30 days ($n = 327$). SwiftSight-Brain is more consistent in 18 of 19 regions, with no significant difference seen for the whole brain CoV. Atlas-wide mean CoV decreases from 1.99% (FastSurfer) to 1.34% (SwiftSight-Brain) ($p < 0.001$).

4.5 In-plane resolution

The in-plane resolution subgroup (1,019 subjects) is restricted to scans of the same subject on the same scanner and identical slice thickness taken within the 30-day window, differing only in in-plane resolution, so that only the target geometric parameter varies. Both engines are stable along this dimension, with mean CoV below 1% in nearly all structures; even so, SwiftSight-Brain is significantly more reproducible in 14 of 19 structures.

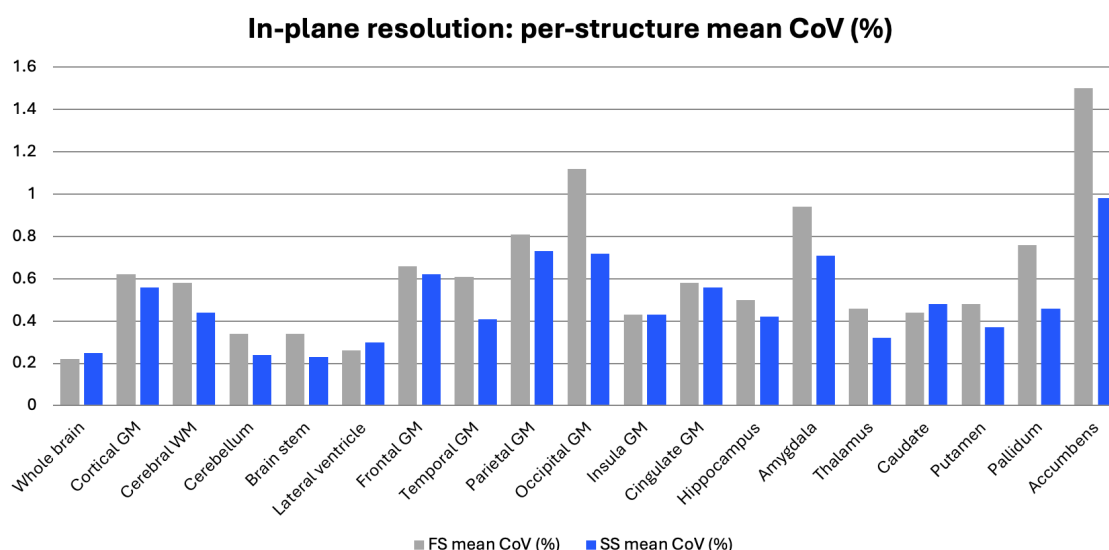


Figure 5. Consistency across different in-plane resolutions. Same subject on the same scanner and slice thickness, differing only in in-plane resolution, within 30 days ($n = 1,019$). Both engines are highly stable, while SwiftSight-Brain displays overall lower CoV in 14 of 19 regions, reducing the atlas-wide mean from 0.61% to 0.49% ($p < 0.001$).

Although both engines remain reproducible across in-plane resolution in aggregate, the structure-level consequence of a resolution change is clearest on a single scan. Figure 6 shows this failure mode directly: as in-plane (phase) resolution is reduced, adjacent but distinct structures such as the hippocampus and amygdala become difficult to distinguish using a conventional segmentation method, whereas SwiftSight-Brain is capable of preserving the boundary between them.

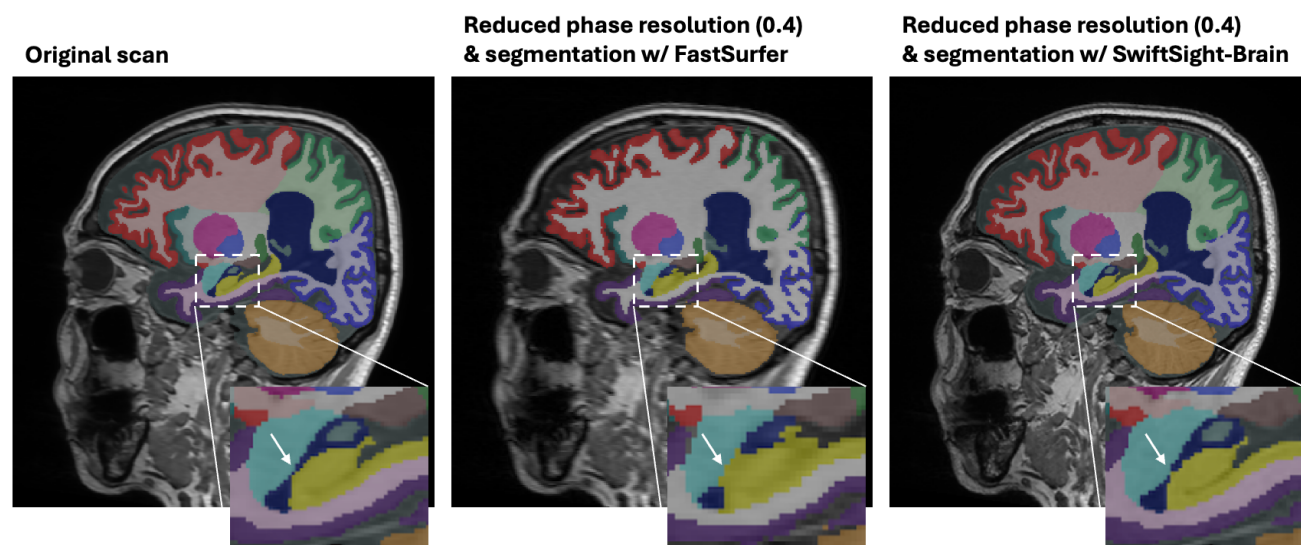


Figure 6. Illustration of segmentation boundary integrity at reduced in-plane resolution. Left: the hippocampus (yellow) and its neighbouring structures segmented on the original scan. Center: at reduced in-plane (phase) resolution, hippocampus boundary was overestimated with FastSurfer (inset, arrow). SwiftSight-Brain keeps the two structures separated at the same resolution (Right).

This retrospective analysis derives its strength from scale and isolation — thousands of routinely acquired examinations in which a single acquisition variable is permitted to change at a time — so that any measured variability can be cleanly attributed to that one source. However, this same variable-isolation strategy is also its principal constraint: in routine, multi-site practice, acquisition rarely changes one parameter at a time, and a patient re-imaged at a different institution typically encounters a different vendor, field strength, pulse sequence, and reconstruction all at once.

To address this real-world imaging practice, we next present a prospective sub-study utilizing data from traveling subjects imaged in varying scanners and protocols.

4.6 A prospective, cross-vendor, cross-protocol validation with traveling subjects

Across the routine clinical protocols, SwiftSight-Brain was more reproducible than FastSurfer in all nine structures and in every subject, roughly halving the atlas-wide mean CoV (2.5% vs 5.4%; per-subject SwiftSight-to-FastSurfer ratio 0.52, 95% CI 0.44–0.63; lower in 6 of 6 subjects, paired $p = 0.031$), including the hippocampus (2.9% vs 6.0%). The advantage was maintained under the stress tier (2.2% vs 5.1%, roughly 56% lower; 6 of 6 subjects), indicating robustness even to deliberately aggressive acquisition. Because these protocols were prospectively and deliberately varied, this sub-study directly probes the large contrast- and protocol-variation that the retrospective analysis could not.

Two further checks confirm the advantage is real rather than an artifact of the analysis. Holding the scanner fixed and varying only protocol within each vendor preserved the advantage (SwiftSight-Brain 1.9% vs FastSurfer 4.1%; lower in 16 of 16 subject-vendor groups, $p < 0.001$), ruling out pooling across scanners as the source. And because a within-tier CoV could in principle be deflated by a consistent directional shift, we measured how far each stress-tier value departed from the same subject's routine mean: SwiftSight-Brain's stress measurements stayed close to routine (mean absolute deviation $\approx 1.8\%$), whereas FastSurfer's departed roughly twice as far ($\approx 4\%$) — confirming the advantage reflects genuine stability, not a shared bias.

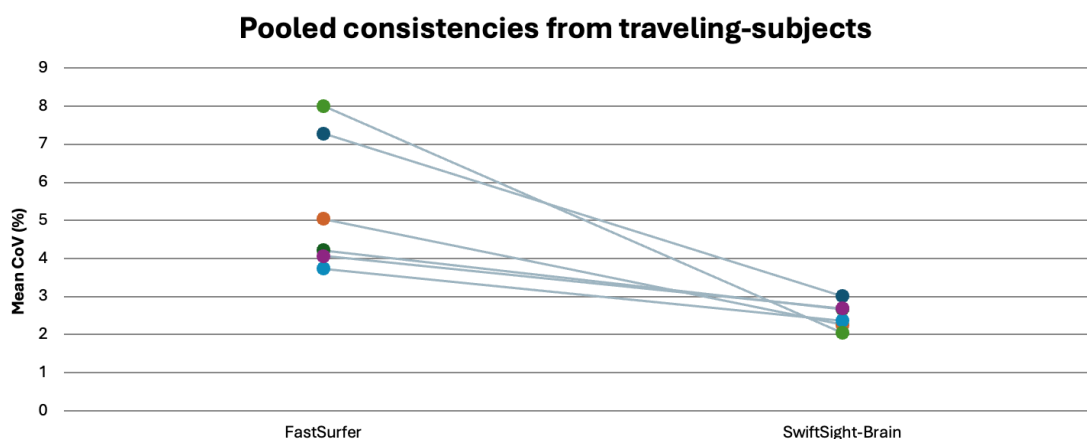


Figure 7. Cross-vendor, cross-protocol consistency from prospective traveling subjects. Six subjects were imaged across GE, Philips, and Siemens platforms under a wide range of deliberately varied acquisition protocols (28–51 per subject), with FastSurfer and SwiftSight-Brain run on identical series. Each line is one subject, connecting that subject's mean per-structure coefficient of variation (CoV) across all of their protocols from FastSurfer (left) to SwiftSight-Brain (right); every subject is more reproducible under SwiftSight-Brain — with the atlas-wide mean roughly halved (2.5% vs 5.4%). Values are a relative comparison across a deliberately sampled protocol range ($n = 6$ subjects; eight lobar and global structures plus the hippocampus).

5. Representative Cases: Visualizing Consistency

Consistency is a distributional property — it is defined across repeated acquisitions and therefore cannot be conveyed by a single image. The cases below use two complementary visual demonstrations: same-day scan-rescan on modern clinical scanners, in which the same subject is imaged several times within a single session, and long-term serial follow-up on a fixed scanner. In the same-day cases, because the brain is physiologically unchanged and the repeat images are visually near-identical, any difference in the reported volume is attributable to measurement instability of the segmentation method rather than to biology; in the longitudinal case, departures from a subject's underlying atrophy trajectory reflect the same measurement instability. Quality-failed acquisitions were excluded. Each case is shown in whichever form best exposes its instability: a plot of the measured volume across the repeated scans where the difference is spatially distributed, or MR slices with FastSurfer and SwiftSight-Brain segmentation overlays where it is focal enough to localize on a single image.

5.1 Cortical grey matter — same-day scan-rescan

A subject was imaged four times in one session on a Siemens Vision 1.5T system. FastSurfer reported scattered cortical grey-matter volumes of 336.8, 352.1, 350.1, and 332.1 mL over the four acquisitions with no single outlier scan (CoV 2.9%). SwiftSight-Brain reported 345.7, 346.2, 345.4, and 344.2 mL (CoV 0.25%), essentially invariant—an order of magnitude more reproducible. Because all four acquisitions are motion-clean and visually near-identical, the difference cannot be attributed to image quality: the conventional pipeline's grey/white cortical boundary is simply not reproduced from scan to scan, whereas SwiftSight-Brain returns effectively the same measurement each time.

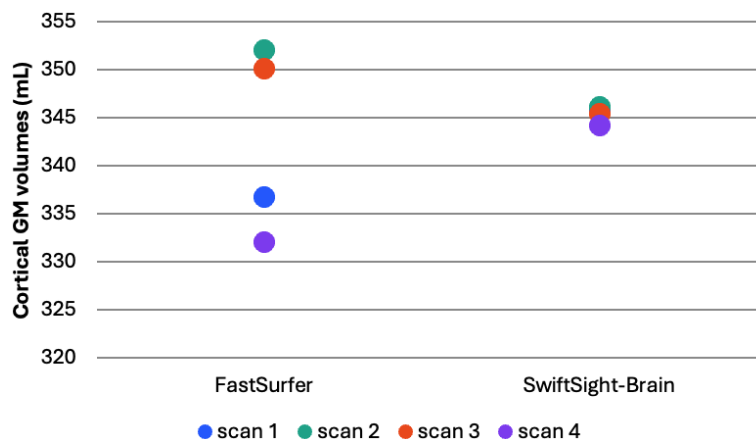


Figure 8. Cortical grey-matter volume across same-session repeats. FastSurfer reported scattered volumes of 336.8, 352.1, 350.1 and 332.1 mL (CoV 2.9%) with no single outlier scan, whereas SwiftSight-Brain returned 345.7, 346.2, 345.4 and 344.2 mL (CoV 0.25%) — an order of magnitude more reproducible. Each marker denotes one acquisition.

5.2 Hippocampus — same-day scan-rescan

A subject was imaged three times in one session on a Siemens TrioTim 3T system. FastSurfer's bilateral hippocampal volume was 6.04, 6.10, and 7.46 mL (CoV 12.3%), a ~23% swing on the third acquisition, while SwiftSight-Brain reported 5.65, 5.66, and 5.82 mL (CoV 1.7%).

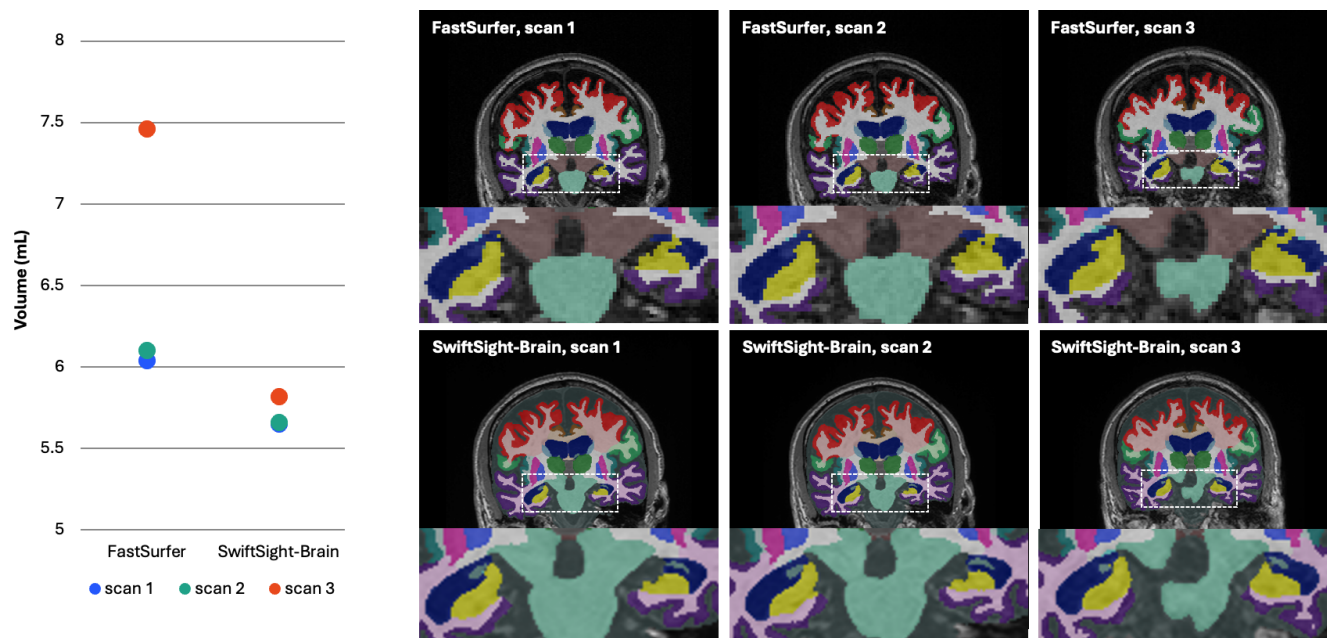


Figure 9. An illustration of scan-rescan differences in hippocampal volumes. Same subject was imaged three times within a single scan session (Siemens TrioTim 3T). FastSurfer's total hippocampal volume was 6.04 and 6.10 mL on the first two acquisitions but 7.46 mL on the third (CoV 12.3%) — a bilaterally symmetric ~23% over-segmentation (left +22.5%, right +23.3%) — whereas SwiftSight-Brain returned 5.65, 5.66 and 5.82 mL (CoV 1.7%). As the acquisitions are repeats of an unchanged brain, an excursion reflects segmentation instability, not true volume change.

Across these same-day cases the pattern is the same: on repeat acquisitions where no biological change is possible, FastSurfer's volume estimate is several-fold to an order of magnitude less reproducible than SwiftSight-Brain's, which stays under 2%. Because these measurements feed directly into age- and sex-adjusted percentile interpretation [11], the practical consequence is that a single SwiftSight-Brain measurement can be trusted to reflect the patient's anatomy, whereas a conventional estimate may need to be repeated or averaged to reach comparable confidence.

5.3 Longitudinal follow-up — hippocampus

Beyond same-day rescans, consistency also governs the stability of serial measurements used to track volume change over time [12]. A selected subject underwent repeated 3D T1-weighted imaging over approximately seven years on the same scanner. Because both engines analyzed identical images, the divergence between their hippocampal trajectories (Figure 10) reflects algorithmic consistency alone, where obvious outliers can be observed. Such errors

may inflate longitudinal measurement noise and can obscure or mimic genuine volume changes.

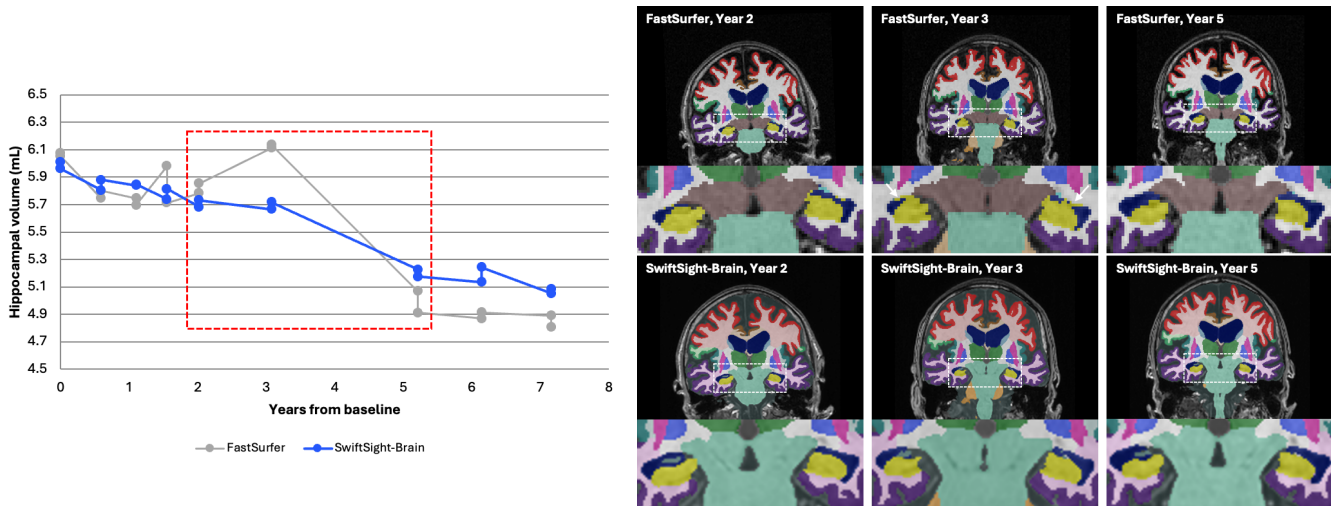


Figure 10. Longitudinal hippocampal volumetry in a single subject. One representative subject with mild cognitive impairment, imaged repeatedly over ~7 years on the same scanner (Siemens Symphony, 1.5T); FastSurfer (grey) and SwiftSight-Brain (blue) processed identical images. Both trace the same true atrophic decline (~0.1–0.2 mL/yr), but FastSurfer scatters more than threefold about that trend (residual CoV 4.3% vs 1.2%; off-trend excursions up to ~11% vs ~3%), driven by intermittent over-segmentation at the superior hippocampal margin (inset, arrows) that SwiftSight-Brain avoids.

6. Limitations

Several limitations apply to this work. The first concerns the analysis itself: because field-strength-associated variability is, by the physics of MRI, inseparable from the accompanying change of scanner hardware, a matched same-vendor analysis was employed to keep the two effects from being confounded.

A second set of limitations concerns the scope of the data. The analysis covers a selected 19-structure atlas in the adult population represented in the evaluation cohort, leaving pediatric coverage to future work, and the prospective traveling-subjects sub-study is narrower still — resting on six subjects and a nine-structure subset, with vendor coverage incomplete for two of them (though the conclusion is unchanged when they are excluded). Because that sub-study's coefficients of variation reflect the range of protocols deliberately sampled rather than a clinical-practice frequency distribution, they should be read as a relative comparison rather than an absolute specification. Among resolution factors, consistency was assessed across in-plane resolution; through-plane (slice-thickness) variation among 3D T1-weighted protocols was not separately isolated. Gross structural abnormalities such as large mass lesions, which violate the assumptions of normative comparison and fall outside the engine's intended use, were likewise not specifically evaluated.

Finally, the scope of this whitepaper is deliberately bounded: it characterizes measurement reproducibility (precision) rather than accuracy against a manually labeled reference or clinical-outcome validity. Segmentation accuracy was independently reviewed in the course of SwiftSight-Brain's FDA 510(k) clearance, and normative and clinical-interpretation performance is addressed in the companion SwiftSight-Brain Normative Database Whitepaper; the present work isolates the complementary question of consistency across acquisition conditions.

7. Conclusion

Measurement consistency is a precondition for trustworthy quantitative neuroimaging. Evaluated on an identical, quality-controlled set of examinations against FastSurfer using a paired, trimmed-mean analysis, SwiftSight-Brain is significantly more reproducible across all four dimensions studied — same-scanner test-retest, field strength, inter-scanner, and in-plane resolution — with the advantage extending across cortical, subcortical, and global structures. The single exception is whole-brain volume under a scanner change, where both engines already operate at their reproducibility floor and are effectively equivalent. Together with FDA 510(k)-cleared segmentation accuracy and the age- and sex-adjusted normative framework described in the companion whitepaper, this consistency supports the use of SwiftSight-Brain for cross-sectional percentile assessment and longitudinal monitoring, where stable measurement across scanners and protocols is essential to distinguishing true structural change from acquisition-driven noise. A prospective traveling-subjects sub-study, in which the same individuals were imaged across many deliberately varied protocols on three vendor platforms, independently corroborated this advantage — SwiftSight-Brain roughly halved cross-protocol measurement variability in both routine and deliberately aggressive acquisitions.

8. References

- [1] Jack, C.R., et al. Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimers Dement.* 7(4):474–485 (2011).
- [2] AIRS Medical. SwiftSight-Brain Normative Database Whitepaper (2026).
- [3] Han, X., et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *NeuroImage* 32(1):180–194 (2006).
- [4] Jovicich, J., et al. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage* 46(1):177–192 (2009).
- [5] Lee, H., et al. Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements. *NeuroImage* 184:555–565 (2019).
- [6] Huppertz, H.-J., et al. Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. *NeuroImage* 49(3):2216–2224 (2010).
- [7] Heinen, R., et al. Robustness of automated methods for brain volume measurements across different MRI field strengths. *PLoS ONE* 11(10):e0165719 (2016).
- [8] Henschel, L., et al. FastSurfer — A fast and accurate deep learning based neuroimaging pipeline. *NeuroImage* 219:117012 (2020).
- [9] Fischl, B. FreeSurfer. *NeuroImage* 62(2):774–781 (2012).
- [10] Fischl, B., et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33(3):341–355 (2002).
- [11] Whitwell, J.L., et al. Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *AJNR Am J Neuroradiol.* 22(8):1483–9 (2001).
- [12] Reuter, M., et al. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* 61(4):1402–1418 (2012).