# Site-to-Site Variability in Quantitative MRI: MINDSET's Scanner Onboarding and Traveling Phantom Method

## Scientific Relevance and Impact of Harmonization

Harmonizing MRI data across scanners will significantly improve MRI-based diagnosis and longitudinal tracking in progressive degenerative disease. Quantitative neuroimaging biomarkers of the same person can differ by up to 15% across MRI scanners.[1] This variability is problematic for longitudinal studies, multi-center clinical trials, and patient-centered clinical practice. Current harmonization techniques, such as ComBat[2], offer a statistical solution based on mixed effect regression to mitigate site effects by aligning data distributions across sites to adjust for scanner-related variability. However, application of ComBat is limited by reliance on assumptions of linearity and parametric distributions and the inability to harmonize observations on new scanners.[2] There is also the danger of removing genuine biological variability, as has been demonstrated in DTI data.[3] Moreover, when a new scanner is introduced, a purely statistical harmonization model may require collection of additional site-specific subject data before that scanner can be incorporated. This limits scalability in distributed clinical networks, where scanner onboarding must occur before large volumes of patient data are available.[9] Statistical methods are insufficient as minor preprocessing differences can yield inconsistent results, emphasizing the need for standardized protocols. Deep learning-based approaches, such as DeepResBat, have demonstrated superior performance in reducing batch effects while maintaining biological signal fidelity,[4] yet they have not been validated in clinical applications and they rely on supervised learning, which means that introduction of new scanners requires additional model training. The AI models in this area are rapidly evolving,[5] and it remains to be seen whether any existing method, statistical or based on deep learning, can successfully remove scanner effects without also removing diagnostic information.[6]

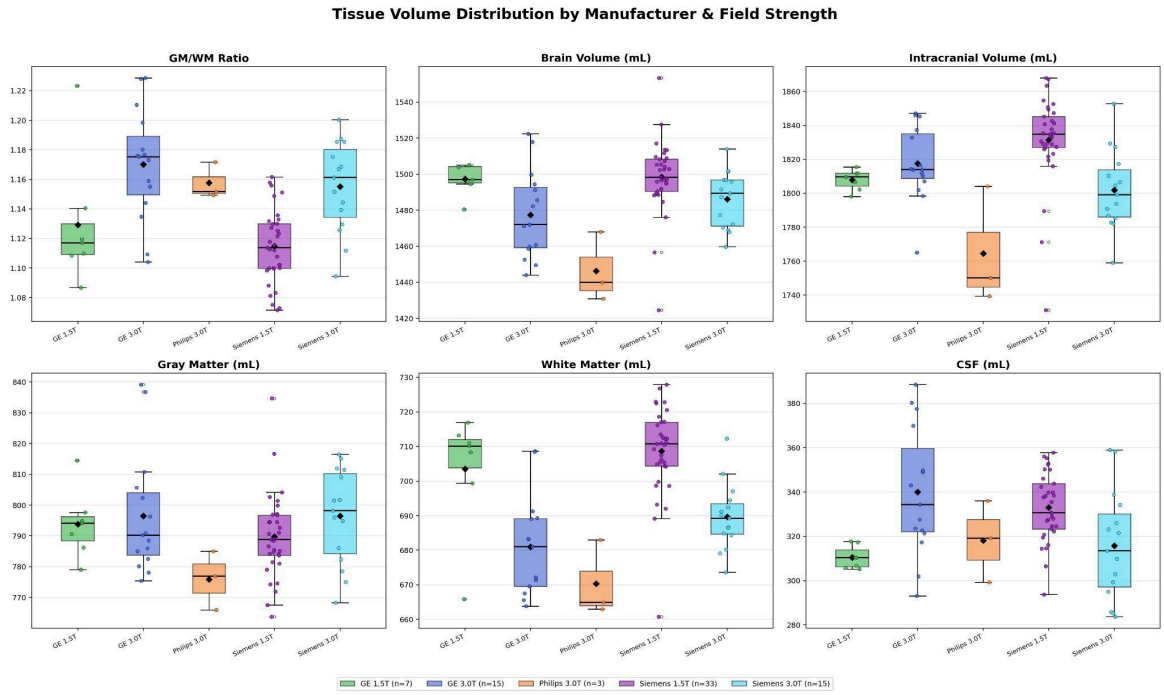


**Figure 1**. Distribution of segmented tissue volumes derived from T1-weighted MRI across 73 scans [ GE 1.5T (n = 7), GE 3.0T (n = 15), Philips 3.0T (n = 3), Siemens 1.5T (n = 33), and Siemens 3.0T (n = 15)] of a single travelling phantom subject acquired at multiple imaging facilities, stratified by scanner manufacturer and magnetic field strength. Six metrics are shown: gray-matter to white-matter ratio (GM/WM Ratio), total brain volume (BV), intracranial volume (ICV), gray-matter volume (GM), white-matter volume (WM), and cerebrospinal-fluid volume (CSF). All volumetric measures are reported in milliliters. Box plots display the median (horizontal line), interquartile range (box), and 1.5 x IQR whiskers; the group mean is indicated by a black diamond. Individual subject values are overlaid as colored points. Because all data were acquired on the same individual, the observed dispersion within each group reflects between-scanner variability rather than between-subject biological variation.

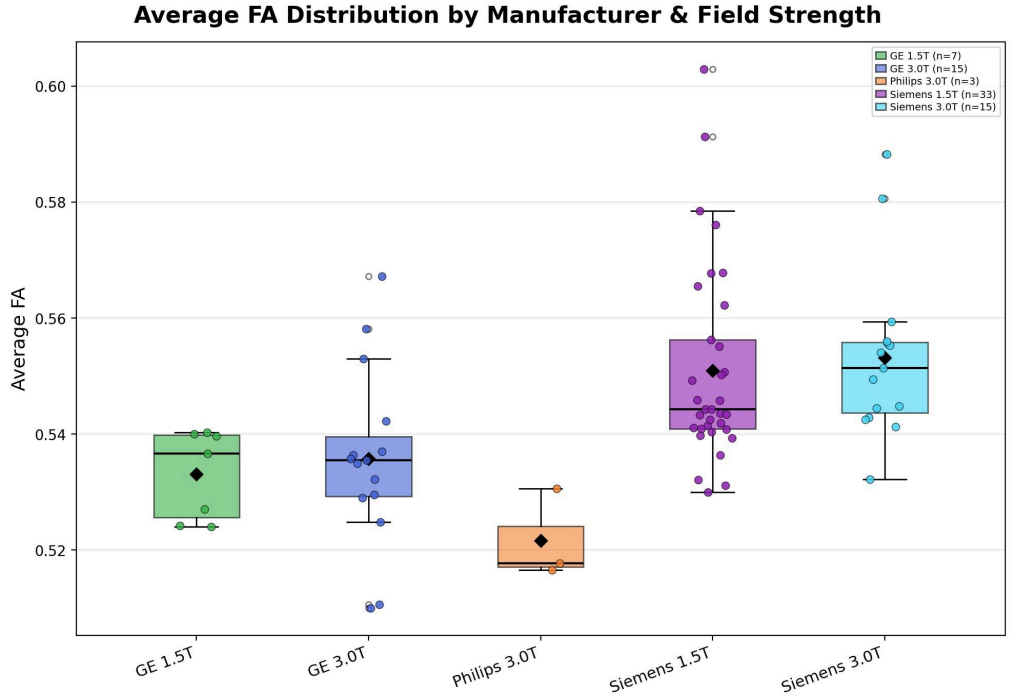


**Figure 2**. Distribution of white-matter mean fractional anisotropy (FA) across 73 scans [GE 1.5T (n = 7), GE 3.0T(n = 15), Philips 3.0T (n = 3), Siemens 1.5T (n = 33), and Siemens 3.0T (n = 15)] of a single travelling phantom subject acquired at multiple imaging facilities,stratified by scanner manufacturer and magnetic field strength. Box plots show the median (horizontal line), interquartile range (box), and 1.5 x IQR whiskers; the group mean is indicated by a black diamond. Individual subject values are overlaid as colored points.

Human phantoms scanned on multiple scanners is the current gold standard, even though it is logistically challenging and expensive. Physical phantoms help standardize imaging but can't capture true biological variability and are also logistically challenging to deploy across sites.[7] The relative contribution of scanner-based variation can be substantial, depending on what other variables are being studied. Scanner effects exceeded even the well-known effects of age in a recent study with 45 scanners, representing 40-60% of total variability.[7] MINDSET's analysis (Figures 1, 2) has revealed inter-scanner variability in both volumetric and fractional anisotropy (FA) metrics when imaging the same subject across multiple (N=73) clinical-grade MRI platforms. This data was collected on the same individual across 73 scanners across manufacturers, Siemens, GE, Philips and field strengths 1.5T and 3T. There is a clear need for harmonization approaches that can separate meaningful biological signals from scanner-driven variability, enabling valid comparisons across imaging sites.

**To account for this variability, MINDSET uses a traveling human subject harmonization approach in its quantitative reporting (QLuminate, Quantify).**

**High-Level Workflow at MINDSET**

MINDSET implements a standardized, multi-step scanner onboarding process to ensure data quality and consistency. MINDSET first onboards each new scanner with a structured intake, focused on confirming scanner readiness and ensuring protocol alignment. This step establishes consistent expectations, reduces variation introduced by setup differences, and creates a stable foundation for evaluation. It also supports smoother operational execution by aligning data transfer and onboarding requirements upfront. Before human phantom calibration, scans undergo a quality gating step designed to quickly identify issues that could affect quantitative reliability. This includes visual assessment to detect major artifacts, automated checks for protocol compliance, and objective quality metrics that help confirm scan integrity. By screening early, onboarding becomes faster and more predictable, while avoiding downstream effort on scans that are not suitable for calibration. Traveling human subject evaluation serves as the backbone of cross-site confidence by providing a controlled reference for assessing scanner performance. This method supports consistent evaluation across sites and helps isolate scanner-driven variability before a site is approved for deployment. For first-time evaluation of a new traveling phantom, any incidental findings renders the scan unacceptable and results in rejection from downstream calibration steps, ensuring the integrity of calibration

datasets and preventing confounded interpretation. Following quality gating, calibration readiness is assessed through ROI stability checks against established reference distributions. This allows MINDSET to confirm that scaling outputs remain consistent and suitable for downstream use. This design improves operational efficiency while maintaining strict standards for acceptance. Once acceptance criteria are met, sites are approved to enter production workflows with confidence that quantitative outputs will be reliable and comparable.

**Conclusion**

While retrospective harmonization methods such as ComBat can be useful when sufficient balanced site data are available, they are less practical for newly onboarded scanners with limited representative subject data. MINDSET's traveling human subject onboarding framework addresses this gap by prospectively assessing the scanner and supports scalable deployment across distributed imaging networks while reducing one of the major historical limitations of quantitative volumetric and diffusion imaging: scanner-dependent variability. The goal of MINDSET's framework is to make scanner readiness measurable, documented, and controlled before quantitative MRI outputs are used in practice.

**References**


1. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *Neuroimage* 208, (2020).
2. Fortin, J. P. *et al.* Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage* 167, 104–120 (2018).
3. Cetin-Karayumak, S. *et al.* Exploring the limits of ComBat method for multi-site diffusion MRI harmonization. *bioRxiv*(2020).
4. An, L. *et al.* DeepResBat: Deep residual batch harmonization accounting for covariate distribution differences. *Med Image Anal* 99, 103354 (2025).
5. Abbasi, S. *et al.* Deep learning for the harmonization of structural MRI scans: a survey. *BioMedical Engineering Online* vol. 23 (2024).
6. Gebre, R. K. *et al.* Cross–scanner harmonization methods for structural MRI may need further work: A comparison study. *Neuroimage* 269, (2023).
7. Garcia-Dias, R. *et al.* Neuroharmony: A new tool for harmonizing volumetric MRI data from unseen scanners. *Neuroimage* 220, (2020).
8. E.M. Palacios, A.J. Martin, M.A. Boss, F. Ezekiel, Y.S. Chang, E.L. Yuh, M.J. Vassar, D.M. Schnyer, C.L. MacDonald, K.L. Crawford, A. Irimia, A.W. Toga, P. Mukherjee Toward Precision and Reproducibility of Diffusion Tensor Imaging: A Multicenter Diffusion Phantom and Traveling Volunteer Study. *American Journal of Neuroradiology,* 38 (3) 537-545; (2017)
9. Jodoin, PM., Edde, M., Girard, G. et al. Challenges and best practices when using ComBAT to harmonize diffusion MRI data. *Sci Rep* 15, 41508 (2025).