Harmonizing multicenter quantitative imaging data: sources of variability, statistical solutions, and practical workflows in CT and MRI

Nurmakhan Zholshybek[1*] , Lazzat Bastarbekova[2]

[1]Department of Medicine, School of Medicine, Nazarbayev University, Astana 010000, Kazakhstan
[2]Radiology Unit, Heart Center, University Medical Center, Astana 010000, Kazakhstan

*Correspondence: Nurmakhan Zholshybek, Department of Medicine, School of Medicine, Nazarbayev University, Astana 010000, Kazakhstan; nurmakhan.zholshybek@nu.edu.kz

Abstract

Multicenter imaging studies are increasingly critical in epidemiology, yet variability across scanners, acquisition protocols, and reconstruction algorithms introduces systematic biases that threaten reproducibility and comparability of quantitative biomarkers. This paper reviews the major sources of heterogeneity in MRI, CT, and PET-CT data, highlighting their impact on epidemiologic inference, including misclassification, reduced statistical power, and compromised generalizability. We outline harmonization strategies spanning pre-acquisition standardization, phantom-based calibration, post-acquisition intensity normalization, and advanced statistical and machine learning methods such as ComBat and domain adaptation. Illustrative examples from MRI flow quantification and radiomic feature extraction demonstrate how harmonization can mitigate site effects and enable robust large-scale analyses.

Keywords

biomarkers, imaging, multicenter studies, image processing, reproducibility

Introduction

The harmonization of data is becoming increasingly important in imaging research. In particular, imaging data is affected by technical variability between scanners, which complicates comparisons across imaging sites, different scanners, and time points [1]. This issue impacts widely used cross-sectional modalities like magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), and computed tomography (CT), as well as derived measurements such as region of interest (ROI) volumes, regional analysis of volumes examined in normalized space maps, cortical thickness estimates, and connectome matrices. To enhance statistical power when aggregating data from multiple sources, post-processing harmonization techniques are essential for reducing unwanted variability [2].

A single-source dataset introduces the potential for institutional biases, which may affect the generalizability of the model. To address the problem, large-scale, multi-center studies were utilized, but the integration of data from diverse sources is essential for achieving robust and generalizable findings. Harmonizing and disseminating data across sites enables researchers to capture variability across populations, imaging systems, and clinical practices, thereby strengthening the study's validity. However, this process requires a well-designed infrastructure capable of acquiring, processing, and sharing data from multiple modalities while aligning with the workflows of all participating centers [3].

Sources of variability in quantitative imaging

The reliability of MRI-derived measurements of human cerebral cortical thickness was investigated by Han et al. [4], with a focus on the effects of field strength, scanner upgrades, and manufacturer differences. They found that the average variability was 0.15 mm for cross-scanner comparisons (Siemens/GE) and 0.17 mm for cross-field strength comparisons (1.5 T/3 T). Measurements across field strengths showed a slight bias, with cortical thickness appearing greater at 3 T [4]. In the variability assessment of volumetric data combined from five different scanners (2 General Electric Signa, 2 Siemens Symphony, and a Philips Gyroscan) at five different sites by repeating the scans of five volunteers at each of the sites using T1-only acquisitions, the two Siemens scanners exhibited a characteristic bias, overestimating white matter and underestimating gray matter compared to the other scanners. This bias, however, was not apparent when multimodal data were used [5]. The results indicated that the greatest compatibility between scanners is achieved when using equipment from the same manufacturer and maintaining image acquisition parameters as similar as possible.

Many factors can influence quantitative measurements during image acquisition. For example, the ability to select among various acquisition parameters and establish optimized protocols contributes to the diversification of positron emission tomography-CT (PET-CT) imaging techniques. Modern scanners incorporate numerous components that may affect quantitative accuracy, including 3D acquisition schemes, scintillators with intrinsic radioactivity, iterative reconstruction algorithms, CT-based attenuation correction, and scatter correction models that rely on multiple assumptions [6]. Moreover, the injected dose of [^{18}F]Fluorodeoxyglucose may range from 300 to 700 megabecquerels, depending on scanner-specific characteristics such as the PET-CT detector material and acquisition mode [7]. These factors in acquisition may lead to research and clinical outcomes. To better understand and control such variability, long half-life PET-CT calibration phantoms are used to compare quantitative measurements across scanners, acquisition protocols, and processing methods by eliminating patient-related factors. Studies using such phantoms have demonstrated that PET-CT measurements exhibit both variance and size-dependent bias influenced by object dimensions, ROI definition, scan duration, acquisition mode, and reconstruction parameters, with appreciable biases reported even for relatively large (37 mm) objects [8, 9].

Snaith et al. [10] reported substantial variation in pelvis radiography techniques, with corresponding implications for clinical decision-making. Calls for standardization of pelvis radiographic studies have been made, and some authors have even proposed specific acquisition protocols [11]. However, there is no evidence that these protocols have been implemented in diagnostic imaging centers [10].

Reconstruction algorithms differ across manufacturers and software platforms. An et al. [12] compared two of the latest 3D modeling software packages, Syngo and Mimics, for accuracy and computational efficiency. Using CT scan images in DICOM (Digital Imaging and Communications in Medicine) format, they evaluated segmentation accuracy, anatomical measurements, cost, and computational time as benchmarks. The authors reported that Mimics outperformed Syngo in terms of semi-automated segmentation and equipment cost, whereas Syngo demonstrated superior computational efficiency [13]. Another key challenge in the reconstruction algorithm process is the potential for intensity bias in the slice data, often caused by anatomical motion relative to the imaging coils. Consequently, slices capturing the same anatomical region at different times may display varying sensitivity. Such bias field inconsistencies can introduce artifacts into the final 3D reconstruction, affecting both the clinical interpretation of critical tissue boundaries and the automated analysis of the data [14].

Beyond scanners, acquisition protocols, and processing methods, variability can also arise from the human element. Carapella et al. [15] demonstrated that standardized training of operators performing manual post-processing of cardiac MRI T1 maps improved consistency in the quantification of T1 biomarkers by reducing subjective bias. Training led to more accurate estimation of mean left ventricular myocardial T1 values and wall thickness, reduced variability in these measurements, and decreased discrepancies relative to reference standards. Moreover, patient positioning performed by technologists has also been shown to significantly influence both radiation dose and image quality in CT [16].

Lange et al. [17] aimed to assess the inter-study reproducibility of cardiac MRI cine image-based hemodynamic forces (HDF) measurements and to explore the current capabilities and limitations of this emerging deformation imaging technique. They concluded that inter-study variability could be improved through further software optimization, and emphasized the need for additional validation studies to support the broader clinical adoption of cardiac MRI-based HDF analysis [17]. Respectively, the study of 25 athletes' cardiac MRI was analyzed by two independent observers and then re-analyzed by the same observer one week apart using HDF parameters derived from feature-tracking cardiac MRI showed a low inter- and intra-observer variability [18].

Strategies for harmonization

Pre-acquisition standardization

Consistency in acquisition parameters is fundamental for reducing variability across imaging sites. Harmonizing protocols involves aligning scanner settings such as tube voltage, current, slice thickness, field of view, and reconstruction kernels in CT, as well as echo time, repetition time, flip angle, and voxel size in MRI [19]. By defining and adhering to consensus protocols, multicenter studies can minimize inter-scanner variability and ensure that images are comparable across different institutions. Whenever possible, published consensus recommendations (The Quantitative Imaging Biomarkers Alliance, the American College of Radiology, and the European Association of Nuclear Medicine guidelines) should be followed to improve reproducibility and facilitate cross-study integration. In PET-CT, for instance, dedicated guidelines have been established to ensure greater consistency in recovery coefficients and the standardized uptake value measurements across different scanners [20].

Phantoms provide a reliable means to calibrate and benchmark scanner performance across centers. Regular phantom scans allow for the assessment of image quality parameters such as noise, resolution, contrast, and geometric fidelity. Using standardized, commercially available phantoms ensures comparability and enables the detection of systematic differences between scanners. Phantom-based harmonization can also help establish site-specific correction factors, thereby reducing bias in quantitative imaging biomarkers [21].

Routine quality assurance (QA) procedures are essential to maintain scanner stability over time. QA programs typically include daily, weekly, and monthly checks of scanner hardware and software, monitoring of calibration drifts, and verification of image quality metrics. In multicenter studies, establishing a centralized QA framework ensures that deviations are detected early and that corrective actions are taken promptly. This not only supports protocol adherence but also builds confidence in the reliability of data across different clinical environments [22]. Vendor-provided QA monitoring is now standard, but early ultra-high-field MRI required adapting 3 T QA procedures to address stronger magnet-gradient interactions. The introduction of higher-channel radiofrequency transmission and parallel transmission further expanded QA needs, leading to additional monitoring of phase differences, signal reflection, and coupling between radiofrequency elements [23].

Post-acquisition techniques

Once images have been acquired, variations in intensity distributions across scanners and protocols may still compromise comparability. Intensity normalization methods, including histogram matching, z-score normalization, and bias field correction, are used to standardize image intensities while preserving

underlying tissue contrasts. These approaches are particularly important in MRI, where scanner-dependent scaling differences can significantly influence quantitative metrics and subsequent analyses [24].

Differences in voxel dimensions, orientations, and slice thicknesses across imaging sites necessitate resampling and reformatting procedures. Spatial harmonization ensures that images share a common resolution and geometry, thereby facilitating multi-site pooling and analysis. Interpolation methods must be applied carefully to avoid introducing artifacts or bias, especially when quantitative biomarkers depend on spatial fidelity. Registration to standardized anatomical templates may also be employed to align data across patients and centers [25].

For advanced quantitative imaging, particularly radiomics, harmonization extends to the level of feature extraction. Variations in segmentation protocols, image preprocessing steps, and feature calculation algorithms can significantly impact feature reproducibility. Adoption of standardized feature definitions, such as those proposed by the Image Biomarker Standardization Initiative, helps ensure consistency across studies [26]. Additionally, harmonization methods such as ComBat can be applied to reduce site-specific variability in extracted features while preserving biologically relevant signals. Radiomic features are often significantly influenced by CT acquisition and reconstruction parameters, which can compromise their reproducibility. However, selecting a smaller subset of more robust features, combined with study-specific correction factors, can substantially enhance clustering reproducibility, for instance, in the analysis of metastatic liver lesions [27].

Statistical and machine learning approaches

ComBat is a data-driven method, meaning that the transformations it uses to align data into a common space must be specifically estimated for each study that includes data from multiple centers or protocols. In a study evaluating whether a compensation method could correct radiomic feature variability arising from different CT protocols, the application of ComBat achieved 100% sensitivity and specificity (48 of 48 volumes of interest) and effectively eliminated scanner and protocol effects while preserving the underlying differences between texture patterns [28]. However, ComBat relies on specific assumptions, and violations of these assumptions can lead to suboptimal or even flawed harmonization [29]. Before applying ComBat, it is important to ensure that the populations being harmonized are as comparable as possible in terms of age range, demographic characteristics, sex distribution, covariate slopes, and health status. Failure to account for these factors may impair harmonization during model training and can lead to substantial errors when the model is applied to new data [30].

Fortin et al. [2] proposed the use and adaptation of five statistical harmonization methods for DTI data: global scaling, functional normalization, RAVEL, Surrogate Variable Analysis, and ComBat, with unharmonized data referred to as "raw." Their findings demonstrated that ComBat effectively retains biological variability while eliminating unwanted site-related variation, increasing the number of voxels demonstrating site-effect reduction from 481 to 5,658 for fractional anisotropy maps, and from 23,136 to 32,203 for mean diffusivity maps [2].

Phantom-based calibration is used to determine scanner-specific acquisition and reconstruction protocols. The close agreement in contrast recovery coefficient measurements between phantom and subject data in the study of Panetta et al. [21] indicates that harmonization strategies established in phantom studies translate effectively to patient images. However, the quantitative consistency between different scanners, as reflected by the root mean squared percent difference, varies depending on the metric used for harmonization [21].

Advancements in style transfer techniques could help address variability in scanner acquisition and reconstruction parameters at the image level. Style transfer is a computer vision method that takes two images, one representing the content and the other providing the reference style, and blends them to create an output that retains the core features of the content image while adopting the artistic style of the reference. In cases where a radiomics model is unavailable for a new scanner or protocol, style transfer could be used to transform images from the new machine, making them appear as if they were captured by an existing scanner [25].

Illustrative example from echocardiography and MRI

Salustri et al. [31] demonstrated that HDF parameters could serve as a step toward standardization across clinical studies and are currently applicable to routinely acquired echocardiographic or cardiac MRI, regardless of equipment brand. Existing evidence highlights the clinical value of HDF in the early detection and monitoring of cardiomyopathy and heart failure, in assessing patients with dyssynchrony, and in evaluating the athlete's heart. Moreover, the authors note that the area under the curve (AUC) can be derived from either the HDF or hemodynamic power (HDP) curves. When computed from the HDF curve, the AUC reflects an impulse, representing a change in momentum, and when normalized by the time interval, it yields the normalized AUC (nAUC), while when calculated from the HDP curve, it corresponds to hemodynamic work [31].

Recommended workflow for harmonized imaging epidemiology

Data harmonization can be conducted either retrospectively or prospectively (Figure 1). In both approaches, the first step for researchers is to identify the variables to be harmonized. This decision should be guided by the overarching goal of the harmonization effort, whether it is theory-driven (aimed at testing specific relationships among selected variables) or data-driven (focused on exploring relationships across a broader set of variables). Additionally, the availability of data and the acceptable degree of harmonization must also be taken into account [32].
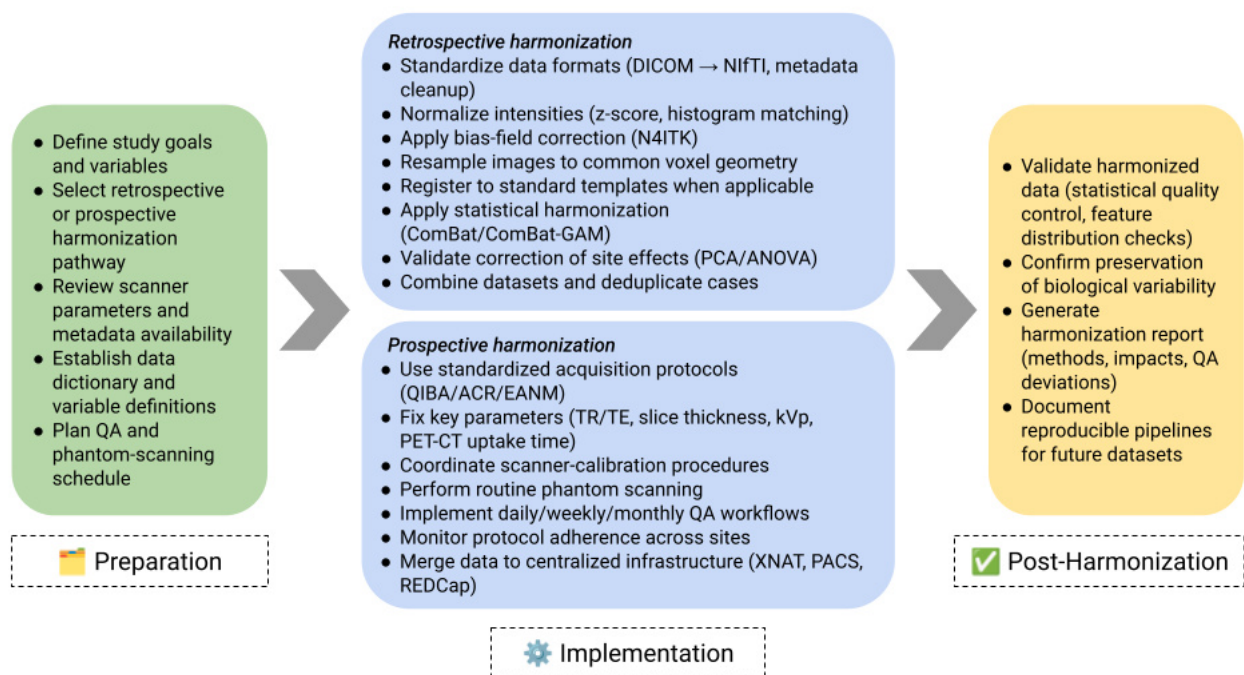

Figure 1. Comprehensive workflow for harmonizing multicenter quantitative imaging datasets. The process comprises three stages: (1) Preparation, in which study goals, variables, and harmonization pathways are defined; (2) Implementation, which includes retrospective and prospective harmonization procedures such as data standardization, intensity normalization, bias-field correction, spatial resampling, ComBat harmonization, protocol coordination, phantom scanning, and QA monitoring; and (3) Post-Harmonization, which focuses on data validation, assessment of preserved biological signal, and documentation of the harmonization process. This framework provides actionable methodological steps to support reproducible multicenter imaging research. QA: quality assurance; DICOM: Digital Imaging and Communications in Medicine; PET-CT: positron emission tomography-computed tomography.

Discussion

Although standardization efforts have been in place for a long time and may need to be further strengthened and expanded to better include radiomics, their ability to reduce variations in radiomic feature distributions across sites is still limited [33], and the preservation of fine anatomical detail and

clinically relevant predictive information is essential in medical imaging, and the downstream impacts of harmonization must be evaluated with caution [34]. The main reason is the continuing diversity of scanner models, proprietary reconstruction algorithms, and post-processing tools used in different clinical centers [33]. In particular, several key challenges were identified in developing deep learning models using multi-site structural brain MRI datasets. These challenges can be grouped into four main categories: (1) difficulty in locating relevant literature, (2) limited access to suitable datasets, (3) a widespread lack of annotation in large datasets, and (4) the need to navigate the trade-off between data harmonization and domain adaptation strategies [35].

Conclusion

The harmonization of multicenter imaging data is indispensable for advancing epidemiologic research. Technical heterogeneity introduced by scanner manufacturers, acquisition protocols, reconstruction algorithms, and operator-dependent factors significantly compromises the reproducibility and generalizability of imaging biomarkers. By systematically addressing these challenges through pre-acquisition standardization, phantom-based calibration, post-acquisition normalization, and statistical or machine learning methods, researchers can substantially reduce site-related variability while preserving biologically meaningful signals. The integration of harmonization workflows into study design not only strengthens causal inference and statistical power but also facilitates collaboration across institutions and populations.

Looking ahead, future directions should emphasize the development of international standards, the incorporation of radiomics and deep learning into harmonization pipelines, and the adoption of federated learning frameworks that allow data sharing without compromising privacy. These efforts will expand the reach of imaging epidemiology, enabling robust and reproducible insights into population health.

Abbreviations

AUC: area under the curve

CT: computed tomography

DTI: diffusion tensor imaging

HDF: hemodynamic forces

HDP: hemodynamic power

MRI: magnetic resonance imaging

PET-CT: positron emission tomography-computed tomography

QA: quality assurance

ROI: region of interest

Declarations

Author contributions

NZ: Conceptualization, Writing—original draft, Writing—review & editing. LB: Supervision, Writing—review & editing. Both authors read and approved the submitted version.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Ethical approval

Not applicable.

Publisher's note

Open Exploration maintains a neutral stance on jurisdictional claims in published institutional affiliations and maps. All opinions expressed in this article are the personal views of the author(s) and do not represent the stance of the editorial team or the publisher.

References

1. Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B, Pacheco 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. 2009;46:177–92. [DOI] [PubMed] [PMC]
2. Fortin J, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site diffusion tensor imaging data. Neuroimage. 2017;161:149–70. [DOI] [PubMed] [PMC]
3. Das S, Zijdenbos AP, Harlap J, Vins D, Evans AC. LORIS: a web-based data management system for multi-center studies. Front Neuroinform. 2012;5:37. [DOI] [PubMed] [PMC]
4. Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage. 2006;32:180–94. [DOI] [PubMed]
5. Reig S, Sánchez-González J, Arango C, Castro J, González-Pinto A, Ortuño F, et al. Assessment of the increase in variability when combining volumetric data from different scanners. Hum Brain Mapp. 2009;30:355–68. [DOI] [PubMed] [PMC]
6. Lodge MA, Lesniak W, Gorin MA, Pienta KJ, Rowe SP, Pomper MG. Measurement of PET Quantitative Bias In Vivo. J Nucl Med. 2021;62:732–7. [DOI] [PubMed] [PMC]
7. Beyer T, Antoch G, Müller S, Egelhof T, Freudenberg LS, Debatin J, et al. Acquisition protocol considerations for combined PET/CT imaging. J Nucl Med. 2004;45:25S–35S. [PubMed]
8. Boellaard R, Oyen WJG, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging. 2008;35:2320–33. [DOI] [PubMed]
9. Doot RK, Scheuermann JS, Christian PE, Karp JS, Kinahan PE. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. Med Phys. 2010;37:6035–46. [DOI] [PubMed] [PMC]
10. Snaith B, Field L, Lewis EF, Flintham K. Variation in pelvic radiography practice: Why can we not standardise image acquisition techniques? Radiography (Lond). 2019;25:374–7. [DOI] [PubMed]

11. Polesello GC, Nakao TS, de Queiroz MC, Daniachi D, Ricioli W Jr, Guimarães RP, et al. PROPOSAL FOR STANDARDIZATION OF RADIOGRAPHIC STUDIES ON THE HIP AND PELVIS. Rev Bras Ortop. 2015;46: 634–42. [DOI] [PubMed] [PMC]

12. An G, Hong L, Zhou XB, Yang Q, Li MQ, Tang XY. Accuracy and efficiency of computer-aided anatomical analysis using 3D visualization software based on semi-automated and automated segmentations. Ann Anat. 2017;210:76–83. [DOI] [PubMed]

13. Khan U, Yasin A, Abid M, Shafi I, Khan SA. A Methodological Review of 3D Reconstruction Techniques in Tomographic Imaging. J Med Syst. 2018;42:190. [DOI] [PubMed]

14. Kim K, Habas PA, Rajagopalan V, Scott JA, Corbett-Detig JM, Rousseau F, et al. Bias field inconsistency correction of motion-scattered multislice MRI for improved 3D image reconstruction. IEEE Trans Med Imaging. 2011;30:1704–12. [DOI] [PubMed] [PMC]

15. Carapella V, Puchta H, Lukaschuk E, Marini C, Werys K, Neubauer S, et al. Standardized image post-processing of cardiovascular magnetic resonance T1-mapping reduces variability and improves accuracy and consistency in myocardial tissue characterization. Int J Cardiol. 2020;298:128–34. [DOI] [PubMed]

16. Al-Hayek Y, Zheng X, Hayre C, Spuur K. The influence of patient positioning on radiation dose in CT imaging: A narrative review. J Med Imaging Radiat Sci. 2022;53:737–47. [DOI] [PubMed]

17. Lange T, Backhaus SJ, Schulz A, Evertz R, Schneider P, Kowallick JT, et al. Inter-study reproducibility of cardiovascular magnetic resonance-derived hemodynamic force assessments. Sci Rep. 2024;14:634. [DOI] [PubMed] [PMC]

18. Ismailov T, Khamitova Z, Jumadilova D, Khissamutdinov N, Toktarbay B, Zholshybek N, et al. Reliability of left ventricular hemodynamic forces derived from feature-tracking cardiac magnetic resonance. PLoS One. 2024;19:e0306481. [DOI] [PubMed] [PMC]

19. Basics of Image Processing. In: Alberich-Bayarri Á, Bellvís-Bataller F, editors. In: The Facts and Challenges of Data Harmonization to Improve Radiomics Reproducibility. Springer Nature; 2024. [DOI]

20. Kaalep A, Sera T, Rijnsdorp S, Yaqub M, Talsma A, Lodge MA, et al. Feasibility of state of the art PET/CT systems performance harmonisation. Eur J Nucl Med Mol Imaging. 2018;45:1344–61. [DOI] [PubMed] [PMC]

21. Panetta JV, Daube-Witherspoon ME, Karp JS. Validation of phantom-based harmonization for patient harmonization. Med Phys. 2017;44:3534–44. [DOI] [PubMed] [PMC]

22. Gatidis S, Kart T, Fischer M, Winzeck S, Glocker B, Bai W, et al. Better Together: Data Harmonization and Cross-Study Analysis of Abdominal MRI Data From UK Biobank and the German National Cohort. Invest Radiol. 2023;58:346–54. [DOI] [PubMed] [PMC]

23. Kraff O, May MW. Multi-center QA of ultrahigh-field systems. MAGMA. 2025;38:519–32. [DOI] [PubMed] [PMC]

24. Seoni S, Shahini A, Meiburger KM, Marzola F, Rotunno G, Acharya UR, et al. All you need is data preparation: A systematic review of image harmonization techniques in Multi-center/device studies for medical support systems. Comput Methods Programs Biomed. 2024;250:108200. [DOI] [PubMed]

25. Mali SA, Ibrahim A, Woodruff HC, Andrearczyk V, Müller H, Primakov S, et al. Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods. J Pers Med. 2021;11:842. [DOI] [PubMed] [PMC]

26. Whybra P, Zwanenburg A, Andrearczyk V, Schaer R, Apte AP, Ayotte A, et al. The image biomarker standardization initiative: standardized convolutional filters for reproducible radiomics and enhanced clinical insights. Radiology. 2024;310:e231319. [DOI] [PubMed] [PMC]

27. Meyer M, Ronald J, Vernuccio F, Nelson RC, Ramirez-Giraldo JC, Solomon J, et al. Reproducibility of CT Radiomic Features within the Same Patient: Influence of Radiation Dose and CT Reconstruction Settings. Radiology. 2019;293:583–91. [DOI] [PubMed] [PMC]

28. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. Radiology. 2019;291:53–9. [DOI] [PubMed]

29. Horng H, Singh A, Yousefi B, Cohen EA, Haghighi B, Katz S, et al. Generalized ComBat harmonization methods for radiomic features with multi-modal distributions and multiple batch effects. Sci Rep. 2022;12:4493. [DOI] [PubMed] [PMC]

30. Jodoin PM, Edde M, Girard G, Dumais F, Theaud G, Dumont M, et al. Challenges and best practices when using ComBAT to harmonize diffusion MRI data. Sci Rep. 2025;15:41508. [DOI] [PubMed] [PMC]

31. Salustri A, Tonti G, Zhankorazova A, Zholshybek N, Toktarbay B, Khamitova Z, et al. Left ventricular hemodynamic forces: gaining insight into left ventricular function. Explor Cardiol. 2025;3:101257. [DOI]

32. Cheng C, Messerschmidt L, Bravo I, Waldbauer M, Bhavikatti R, Schenk C, et al. A General Primer for Data Harmonization. Sci Data. 2024;11:152. [DOI] [PubMed] [PMC]

33. Papadimitroulas P, Brocki L, Chung NC, Marchadour W, Vermet F, Gaubert L, et al. Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data harmonization. Phys Med. 2021;83:108–21. [DOI] [PubMed]

34. Bashyam VM, Doshi J, Erus G, Srinivasan D, Abdulkadir A, Singh A, et al. Deep Generative Medical Image Harmonization for Improving Cross-Site Generalization in Deep Learning Predictors. J Magn Reson Imaging. 2022;55:908–16. [DOI] [PubMed] [PMC]

35. Bento M, Fantini I, Park J, Rittner L, Frayne R. Deep Learning in Large and Multi-Site Structural Brain MR Imaging Datasets. Front Neuroinform. 2022;15:805669. [DOI] [PubMed] [PMC]