Managing the effect of magnetic resonance imaging pulse sequence on radiomic feature reproducibility in the study of brain metastases [version 1; peer review: awaiting peer review]

Drew Mitchell¹, Samantha Buszek², Benjamin Tran², Maguy Farhat², Jodi Goldman², Lily Erickson¹², Brandon Curl², Dima Suki³, Sherise D. Ferguson³, Ho-Ling Liu¹, Suprateek Kundu⁴, Caroline Chung²

¹Department of Imaging Physics, University of Texas MD Anderson Cancer Center, Houston, Texas, 77030, USA
²Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas, 77030, USA
³Department of Neurosurgery, University of Texas MD Anderson Cancer Center, Houston, Texas, 77030, USA
⁴Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, 77030, USA

Abstract
Background: Despite the promise of radiomics studies, their limited reproducibility has hindered meaningful clinical translation. Variability in study designs as well as image acquisition and processing contribute to unreproducible radiomic results. This work's purpose was to (i) quantitatively compare variability of radiomic features extracted from 2-D spin echo (SE) and 3-D spoiled gradient echo (SPGR) T1-weighted post-contrast magnetic resonance (MR) images of brain metastases acquired within the same patient in a single imaging session, and (ii) provide a framework to inform data acquisition for reproducible radiomics studies.

Methods: A retrospective cohort of 29 patients with pathologically-confirmed brain metastases and contrast-enhanced T1-weighted MR images acquired using 2-D SE and 3-D SPGR sequences within one exam was identified. Metastases were segmented twice by different physicians using semi-automated methods. Radiomic features were extracted using PyRadiomics for 264 preprocessing variable combinations. Lin's concordance correlation coefficient (CCC) was computed between features extracted from images acquired by both pulse sequences and different tumor segmentations.

Results: We provided general recommendations to improve MR-based radiomic feature reproducibility by clustering and identifying low-concordance features and processing variables. Median CCC between 2-D SE and 3-D SPGR (measuring feature agreement between pulse sequences) was greater for fixed bin count intensity.
discretization (0.76 versus 0.63) and specific high-concordance features (0.74 versus 0.53). Applying all recommendations improved median CCC from 0.51 to 0.79. Median CCC between contours (measuring feature sensitivity to inter-observer variability) was higher for 2-D SE (0.93 versus 0.86) but improved to 0.93 for 3-D SPGR after low-concordance feature exclusion.

**Conclusions**: The following recommendations are proposed to improve reproducibility: 1) Fixed bin count intensity discretization for all studies, 2) for studies with 2-D and 3-D datasets, excluding high-variability features from downstream analyses, 3) when segmentation is manual or semi-automated, using only 2-D SE images or excluding features susceptible to segmentation variability.

**Keywords**
Magnetic resonance imaging, radiomics, reproducibility, brain metastases

---

This article is included in the **Oncology** gateway.

This article is included in the **Radiomics** collection.
Introduction

It is estimated that 20% of cancer patients develop brain metastases, and prognosis following metastasis to the brain is generally poor. 1 Surgical resection, whole-brain radiation therapy, and stereotactic radiosurgery are the most prevalent treatments and are necessary to extend survival, preserve neurologic function, and provide palliative care. 2 Magnetic resonance imaging (MRI) plays a critical role in the diagnosis and treatment of brain metastases. Several different pulse sequences are routinely employed in the detection and monitoring of brain metastases, and their effectiveness for these purposes has been the subject of several studies. 3-26 In particular, post-contrast T1-weighted images can be acquired in a number of ways, commonly including 2-D conventional spin echo (SE) and 3-D spoiled gradient echo (SPGR). Generally, 2-D SE acquisitions provide images with better lesion conspicuity, whereas SPGR can be acquired more easily with thin slices in 3-D, allowing better detection of small metastases. 3 A recent consensus recommendation for imaging brain metastases includes post-contrast T1-weighted acquisitions with both of these pulse sequences. 3

Despite the valuable role of MRI in imaging brain metastases, some clinical questions that determine course of treatment, such as differentiation between tumor progression and radiation necrosis or determination of metastatic tumor type, are difficult or impossible to answer by evaluating MR images with the human eye. Radiomics aims to completely characterize data contained in an image or region of interest (ROI) by using dozens to hundreds of different radiomic features that each capture some characteristic of the image or ROI. Radiomics confers several potential advantages: radiomic features may capture image characteristics that are nearly or completely invisible to the human eye, features can quantitatively capture image characteristics that otherwise require qualitative evaluation, and radiomics can be employed in automatic tools to augment clinical decision making. One of the principal applications of radiomic features is to train machine learning classifiers that can assist clinical decisions. 6 Several studies demonstrate the feasibility of such tools in cases such as patient outcomes in non-small cell lung cancer. 10

There are several important applications of radiomics for brain metastases in particular. One such application is training machine learning classifiers to distinguish between tumor progression and radiation necrosis, particularly in patients who have undergone Gamma Knife radiosurgery. 11-13 One study used radiomic features as a prognostic factor to predict effectiveness of Gamma Knife radiosurgery in brain metastases. 11 Another important application exists in cases where brain metastases are detected before diagnosing the primary cancer. In these cases machine learning classifiers have been trained on radiomic feature data to predict metastatic tumor type. 15,16

While radiomics studies have already yielded exciting findings and these tools have many promising applications, there are concerns about the reproducibility of radiomics studies due to the range of possible study designs and the inherent variability in radiomic features as a function of imaging modality, acquisition parameters, scanner, image preprocessing methods, and feature definitions. 17 Several studies have addressed these concerns in computed tomography (CT) 18,19 and positron emission tomography (PET). The Image Biomarker Standardization Initiative (IBSI) has worked to standardize radiomic features across several imaging modalities, including MRI, CT, and PET 20 and has also published a manual in which consensus-based recommendations and guidelines for radiomics are presented, as well as a general radiomics image processing scheme. 21 A recent review of repeatability and reproducibility studies of radiomic features in cancer patients found good representation of CT and PET images, however, relatively few investigated MR-based radiomic features. Furthermore, repeatability and reproducibility investigations have been limited to a small number of cancer types, with types such as non-small cell lung cancer (NSCLC) being dominant in the literature. 22 None addressed brain metastases. The need for standardization of MR-based radiomic features in particular is well understood,23 and there have been several suggestions for bringing uniformity to radiomics study workflows.24 It is critical to understand the variability of MR-based radiomic features in order to identify which will be reliable for downstream applications, such as machine learning classifiers. As recently as 2016, a review by Yip and Aerts found no investigations of MR-based radiomic feature repeatability, highlighting the need for such work. 25

In the intervening time, several studies have begun to address important questions about the variability of MR-based radiomic features. In a mathematical phantom, Ford et al. investigated MR-based radiomic feature variability as a function of pulse sequence parameter selection, 26 and Bologna et al. studied feature robustness to various acquisition parameters in a digital phantom. 27 In MRI phantoms, Mayerhofer et al. investigated how radiomic feature variability responded to variations in acquisition parameters, including the number of acquisitions, repetition time, echo time, sampling bandwidth, and spatial resolution. 28 Rai et al. performed a multicenter evaluation of MRI-based radiomic feature reproducibility in phantoms, 29 Wong et al. studied longitudinal acquisition repeatability of features on an American College of Radiology MRI phantom, 30 and Lee et al. used a test-retest scheme to quantify repeatability of radiomic features in an MRI radiomics phantom while varying acquisition parameters across multiple scanners. 31 Finally, in human images, one recent test-retest study investigated repeatability of MR-based radiomic features in glioblastoma across several preprocessing approaches, 32 one examined the sensitivity of radiomic features to inter-observer variability...
in apparent diffusion coefficient maps in cervical cancer,\textsuperscript{33} and another compared intensity normalization and discretization methods for gadolinium-enhanced T1-weighted and T2-weighted fluid-attenuated inversion recovery series in glioma patients.\textsuperscript{31} Recent MRI studies have also been performed in patients with Alzheimer’s,\textsuperscript{35} multiple sclerosis,\textsuperscript{35} lacrimal gland tumors,\textsuperscript{36} breast lesions,\textsuperscript{36} and glioblastoma multiform,\textsuperscript{37,38} as well as healthy volunteers.\textsuperscript{35,39–41} Many preprocessing variables have been studied, including contrast weighting,\textsuperscript{42} resolution,\textsuperscript{40} acceleration,\textsuperscript{40,41} gray level discretization,\textsuperscript{36,38} statistical normalization,\textsuperscript{35,39} and bias field correction,\textsuperscript{37} all of which affect the repeatability and reproducibility of extracted radiomic features.

Many studies of MR-based radiomic features have been test-retest repeatability studies,\textsuperscript{30,38–41} which typically minimize dataset heterogeneity in order to exclusively isolate the intra-scanner variability of radiomic features. However, several previous studies have pointed out the need for balance between dataset homogeneity, which results in low noise in order not to mask any radiomic signature, and heterogeneity, which offers increased generalizability for application to real-world datasets.\textsuperscript{42–44} Previous studies have addressed some practical concessions that must be made to real-world dataset quality. Many concluded that one of the best options to standardize results from multi-center studies or retrospective cases is to perform preprocessing prior to feature extraction and determine the stable features.\textsuperscript{27,42} Given the prevalence of this pragmatic approach, best practices must be determined for radiomics studies in these real-world conditions as well as in studies with highly controlled datasets.

The purpose of this work was to both (i) quantitatively compare the variability of radiomic features extracted from 2-D SE and 3-D SPGR MR images of brain metastases acquired within the same patient in a single imaging session to determine the impact of image acquisition on the identified radiomic features, and (ii) to provide a framework to use these results to inform data acquisition and processing to improve the reliability and reproducibility of radiomics studies. Consensus recommendations for acquiring post-contrast T1-weighted images using both pulse sequences are relatively recent, and many imaging protocols and existing datasets include only one post-contrast T1-weighted acquisition. With this unique dataset, it is critical to understand the variability of radiomic features extracted from both acquisition types, especially for retrospective studies where heterogeneous imaging data is often unavoidable. Furthermore, this information impacts the design of imaging protocols for future studies and the selection of appropriate radiomic data where data from both acquisitions are available. We consider this study to be innovative because it addresses a key aspect of reproducibility in MRI radiomics studies: the sensitivity of features to the input imaging data acquisition parameters. To our knowledge, no previous studies have compared variability of radiomics features derived from different common MR pulse sequences in human imaging data. Based on this data, we provide general recommendations for the design of reproducible radiomics studies, and the extended data can be used to guide study design in more specific cases.

Methods

Dataset

The study protocol (PA17-0374_MOD0008) was approved by the MD Anderson Office of Human Subjects Protection. Ethical approval and consent from participants were waived by the committee. A retrospective cohort was identified consisting of 225 patients treated with Gamma Knife for brain metastasis who subsequently developed radiological/clinical progression and required surgical resection of the same lesion. MR imaging was performed on multiple 1.5T and 3T General Electric MR scanners prior to surgical resection, after which brain metastases were confirmed via pathology. Patients were excluded who did not have T1-weighted post-contrast images acquired by both 2-D SE and 3-D SPGR in a single exam, resulting in 29 patients. T1-weighted post-contrast images were acquired after application of gadobutrol (Gadavist, Bayer Healthcare) with a dose of 0.1 mmol/kg body weight. Within the same exam, T1-weighted post contrast images were acquired using 2-D SE after contrast injection, followed by 3-D SPGR. Standard in-line reconstruction provided by the vendor was used, and no acceleration was employed. Table 1 shows characteristics of these 29 cases, and Table 2 lists scanners and pulse sequence parameter ranges for the two T1-weighted post-contrast acquisitions.

Tumor volume segmentation

Tumor volumes were contoured by experienced physicians using the treatment planning system RayStation. In patients with multiple lesions, the contoured lesion was the same one ultimately surgically resected. The enhancing volume in both T1-weighted post-contrast images (2-D SE and 3-D SPGR) was contoured via semi-automated methods using tools available in RayStation and adjusted manually as necessary. The tumor volume on each image was contoured twice by different physicians in order to examine the variability introduced through segmentation. Contour similarity was evaluated using Dice similarity coefficient (DSC). Example images and tumor segmentations are shown in Figure 1.

Image preprocessing and radiomic feature extraction

In total, 29 patients with brain metastases were included in this study. Each patient had two images (2-D SE and 3-D SPGR), and each tumor volume was segmented by two different physicians, resulting in 58 images and 116 segmentations.
of the tumor volumes that were analyzed. Radiomic features were extracted using PyRadiomics, which adheres to the image biomarker standardization initiative (IBSI) definitions of features. For each combination of preprocessing variables, 105 radiomic features were extracted from several classes, including shape, first order, gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM), and neighboring gray tone difference matrix (NGTDM). These features are listed in Table 3 under their respective classes along with abbreviations for the full feature names used in figures for readability.

Several preprocessing variables were investigated, including spatial normalization, 2-D or 3-D radiomic feature extraction, intensity discretization, and image filters. Spatial normalization resolutions included 0.4297 mm, 1 mm, 3 mm, and 1 mm. Both 2-D and 3-D radiomic features were extracted, where 3-D features consider voxels from adjacent slices to be neighboring for purposes of feature computations, and 2-D features only consider neighboring voxels within the same slice. Intensity discretization performed on original images included both fixed bin count and fixed bin width methods. Fixed bin counts of 16, 32, 64, 128, and 256 and fixed bin widths of 16, 32, 64, 128, and 256 were considered. This resulted in a total of 264 preprocessing scenarios and 30,624 different feature extractions. In addition to original images, radiomic features were extracted from filtered images. Filters included

### Table 1. Characteristics of brain metastases cases treated with Gamma Knife that subsequently developed tumor progression, requiring surgical resection of the same lesion.

In total, 29 cases had T1-weighted post-contrast images acquired by both 2-D spin echo and 3-D spoiled gradient echo pulse sequences in the same exam prior to surgical resection.

| Characteristic                        | Cases (n = 29) |
|--------------------------------------|---------------|
| **Gender**                           |               |
| Male                                 | 7             |
| Female                               | 22            |
| **Race**                             |               |
| Caucasian                            | 17            |
| African American                     | 3             |
| Other                                | 9             |
| **Histology**                        |               |
| Breast                               | 12            |
| NSCLC                                | 9             |
| Sarcoma                              | 3             |
| Esophagus                            | 1             |
| Colorectal                           | 1             |
| Head and neck                        | 1             |
| RCC                                  | 1             |
| Cervical                             | 1             |
| **Age at Primary Diagnosis**         |               |
| Mean                                 | 53            |
| Standard Deviation                   | 10.9          |
| **Number of BM at first diagnosis of BM** |           |
| Mean                                 | 2             |
| Standard deviation                   | 1.8           |
| **Time to BM Diagnosis from Primary diagnosis (months)** |           |
| Mean                                 | 26            |
| Standard deviation                   | 18.8          |
| **Gamma knife volume (mm³)**         |               |
| Mean                                 | 3754          |
| Standard deviation                   | 3220          |
Laplacian of Gaussian (sigma = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0 mm) and wavelet (LLH, LHL, LHH, HLL, HLH, HHL, HHH, LLL). Intensity normalization was also performed using square, square root, logarithm, exponential, and gradient scales.

**Data analysis**

Lin’s concordance correlation coefficient (CCC) was used to determine concordance between radiomic feature values extracted from the two acquisition types. CCC is defined between two variables $X$ and $Y$ as

$$
\rho_c = \frac{2\rho \sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2},
$$

where $\rho$ is the correlation coefficient between $X$ and $Y$, $\sigma_x$ and $\sigma_y$ are the standard deviations of $X$ and $Y$, respectively, and $\mu_x$ and $\mu_y$ are the means of $X$ and $Y$, respectively.

CCC was computed between values of each radiomic feature extracted from the 2-D SE series and those extracted from the 3-D SPGR series for all 264 combinations of preprocessing variables. Kernel density estimates were performed to compare the distribution of CCC among all radiomic features between each combination of preprocessing methods.

### Table 2. Acquisition parameters for 2-D SE and 3-D SPGR pulse sequences.

| Number of exams | scanner | Field strength (T) | TR (ms), TE (ms) | Resolution, Slice thickness (mm), slice spacing (mm) | Flip angle (°) |
|-----------------|---------|-------------------|-----------------|------------------------------------------------|---------------|
| 22              | GE Signa HDxt (16) GE Discovery MR750w (3) GE Discovery MR750 (2) GE Signa PET/MR (1) | 3                | 2-D SE TR 518.7-916.7, TE 11-13 | 2-D SE FOV: 220 mm, 352 × 224 ST: 5.0, SS: 6.0-6.5 | 2-D SE 68-90 |
|                 |         |                   |                 | 3-D SPGR TR 5.7-7.2, TE 2.1-2.5 | 3-D SPGR FOV: 240-280 mm, 256 × 256-352 × 224 ST: 1.4-1.8, SS: 0.7-1.8 | 3-D SPGR 10-20 |
| 7               | GE Signa HDxt (5) GE Optima MR450w (2) | 1.5              | 2-D SE TR 418.7-600, TE 8-12 | 2-D SE FOV: 220 mm, 256 × 192 ST: 5.0, SS: 6.0-6.5 | 2-D SE 90 |
|                 |         |                   |                 | 3-D SPGR TR 6.9-10.4, TE 2.5-4.2 | 3-D SPGR FOV: 220-280 mm, 256 × 256 ST: 1.0-1.8, SS: 0.7-1.8 | 3-D SPGR 12-25 |

1Field of view.
2Slice thickness.
3Slice spacing.

**Figure 1.** T1-weighted post-contrast images acquired by 2-D spin echo (left) and 3-D spoiled gradient echo (right) with the two lesion segmentations shown in red and green.

Laplacian of Gaussian (sigma = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0 mm) and wavelet (LLH, LHL, LHH, HLL, HLH, HHL, HHH, LLL). Intensity normalization was also performed using square, square root, logarithm, exponential, and gradient scales.

**Data analysis**

Lin’s concordance correlation coefficient (CCC) was used to determine concordance between radiomic feature values extracted from the two acquisition types. CCC is defined between two variables $X$ and $Y$ as

$$
\rho_c = \frac{2\rho \sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2},
$$

where $\rho$ is the correlation coefficient between $X$ and $Y$, $\sigma_x$ and $\sigma_y$ are the standard deviations of $X$ and $Y$, respectively, and $\mu_x$ and $\mu_y$ are the means of $X$ and $Y$, respectively.

CCC was computed between values of each radiomic feature extracted from the 2-D SE series and those extracted from the 3-D SPGR series for all 264 combinations of preprocessing variables. Kernel density estimates were performed to compare the distribution of CCC among all radiomic features between each combination of preprocessing methods.
Table 3. Class, full name, and abbreviations for all extracted features.

| Shape features                      | Full name | Abbreviation | First order features | Full name | Abbreviation | Full name | Abbreviation |
|-------------------------------------|-----------|--------------|----------------------|-----------|--------------|-----------|--------------|
| Elongation                          | Shape_Elong | FO_10P       | Autocorrelation      | GLCM_AC   |              |
| Flatness                            | Shape_Flat | FO_90P       | Joint average        | GLCM_JA   |              |
| Least axis length                   | Shape_LAL  | FO_Energy    | Cluster prominence   | GLCM_CP   |              |
| Major axis length                   | Shape_MaxAL| FO_Entropy   | Cluster shade        | GLCM_CS   |              |
| Maximum 2D diameter column          | Shape_M2DDC | FO_IR        | Cluster tendency     | GLCM_CT   |              |
| Maximum 2D diameter row             | Shape_M2DDR | FO_Kurt     | Contrast             | GLCM_Corr |              |
| Maximum 2D diameter slice           | Shape_M2DDS | FO_Max      | Correlation           | GLCM_Corr |              |
| Maximum 3D diameter                 | Shape_M3DD  | FO_MAD       | Difference average   | GLCM_DA   |              |
| Mesh volume                         | Shape_MV   | FO_Mean      | Difference entropy   | GLCM_DE   |              |
| Minor axis length                   | Shape_MinAL | FO_Median    | Difference variance  | GLCM_DV   |              |
| Sphericity                          | Shape_Spher | FO_Min       | Joint energy         | GLCM_JEn  |              |
| Surface area                        | Shape_SA   | FO_Range     | Joint entropy        | GLCM_JEnt |              |
| Surface volume ratio                | Shape_SVR  | FO_RMAD      | Informational measure of correlation 1 | GLCM_IMC1 |              |
| Voxel volume                        | Shape_VV   | FO_RMS       | Informational measure of correlation 2 | GLCM_IMC2 |              |
|                                   |            | FO_Skew      | Inverse difference moment | GLCM_IDM  |              |
| Neighboring gray tone difference matrix features | Total energy | FO_TE       | Inverse difference moment normalized | GLCM_IDMN |              |
| Full name                           | Abbreviation | Uniformity | FO_Unif | Inverse difference | GLCM_ID |
| Busyness                            | NGTDM_Busy | FO_Var       | Inverse difference normalized | GLCM_IDN  |              |
| Coarseness                          | NGTDM_Coarse |          | Inverse variance     | GLCM_IV   |              |
| Complexity                          | NGTDM_Comp  | Maximum probability | GLCM_MP   |              |
| Contrast                            | NGTDM_Cont | Sum entropy  | GLCM_SE   |              |
| Strength                            | NGTDM_Str   | Sum squares  | GLCM_SS   |              |
| Gray level run length matrix features |            |             |          |              |              |
| Gray level non uniformity           | GLRLM_GLNU  |             | Dependence entropy   | GLDM_DE   |              |
| Gray level non uniformity normalized | GLRLM_GLNUN |             | Dependence non uniformity | GLDM_DNU  |              |
CCC was also computed between values of each radiomic feature extracted from the first segmentation of the tumor volume and those from the second segmentation of the same tumor volume. CCC was calculated for each radiomic feature, combination of preprocessing methods, and acquisition type. Kernel density estimates were again performed to compare the distribution of CCC among all radiomic features between each different acquisition type and combination of preprocessing methods.

Table 3. Continued

| Gray level run length matrix features | Gray level size zone matrix features | Gray level dependence matrix features |
|--------------------------------------|-------------------------------------|---------------------------------------|
| Full name                            | Abbreviation                        | Full name                              | Abbreviation                        | Full name                              | Abbreviation                        |
| Gray level variance                  | GLRLM_GLV                           | Gray level variance                    | GLSZM_GLV                           | Dependence non uniformity normalized   | GLDM_DNUN                           |
| High gray level run emphasis         | GLRLM_HGLRE                         | High gray level variance                | GLSZM_HGLZE                         | Dependence variance                    | GLDM_DV                             |
| Long run emphasis                    | GLRLM_LRE                           | Large area emphasis                    | GLSZM_LAE                           | Gray level non uniformity              | GLDM_DGLNU                           |
| Long run high gray level emphasis    | GLRLM_LRHGLE                        | Large area high gray level Emphasis    | GLSZM_LAHGLE                        | Gray level variance                    | GLDM_DGLV                           |
| Long run low gray level emphasis     | GLRLM_LRLGLE                        | Large area low gray level Emphasis     | GLSZM_LALGLE                        | High gray level emphasis               | GLDM_HGLE                           |
| Low gray level run emphasis          | GLRLM_LGLRE                         | Low gray level zone emphasis           | GLSZM_LGLZE                         | Large dependence emphasis              | GLDM_LDE                            |
| Run entropy                          | GLRLM_RE                            | Size zone non uniformity               | GLSZM_SZNU                          | Large dependence high gray level emphasis | GLDM_LDHGLE                       |
| Run length non uniformity            | GLRLM_RLNU                          | Size zone non uniformity normalized    | GLSZM_SZNUN                         | Large dependence low gray level emphasis | GLDM_LDLGLE                       |
| Run length non uniformity normalized | GLRLM_RLNUN                         | Small area emphasis                    | GLSZM_SAE                           | Low gray level emphasis                | GLDM_LGE                            |
| Run percentage                       | GLRLM_RP                            | Small area high gray level emphasis    | GLSZM_SAHGLE                        | Small dependence emphasis              | GLDM_SDE                            |
| Run variance                         | GLRLM_RV                            | Small area low gray level emphasis     | GLSZM_SALGLE                        | Small dependence high gray level emphasis | GLDM_SDHGLE                       |
| Short run emphasis                   | GLRLM_SRE                           | Zone entropy                           | GLSZM_ZE                           | Small dependence low gray level emphasis | GLDM_SDLGLE                       |
| Short run high gray level emphasis   | GLRLM_SRHGLE                        | Zone percentage                        | GLSZM_ZP                           |                                        |                                    |
| Short run low gray level emphasis    | GLRLM_SRLGLE                        | Zone variance                          | GLSZM_ZV                           |                                        |                                    |
General recommendations for radiomic study design were given based on patterns of concordance. Exclusion was suggested for filters or preprocessing variable combinations with consistently low concordance or highly variable concordance between 2-D SE and 3-D SPGR extractions. Low concordance suggests these features, filters, and preprocessing variables are likely unreliable in heterogeneous datasets where both 2-D SE and 3-D SPGR pulse sequences were used. Features with consistently low concordance between extractions from the different lesion segmentations were likewise recommended for exclusion in studies in which segmentation variability was present.

A Kolmogorov-Smirnov (K-S) test was performed between radiomic feature values computed from 2-D SE images and those from 3-D SPGR images to assess whether the two distributions were significantly different. K-S tests were repeated to test for distribution differences in each radiomic feature and each different combination of preprocessing variables. Levene’s test for equality of variances was also performed on the same data to determine whether significant differences in variability were present.

**Results**

How to interpret and use these results to design your MR acquisitions and analysis

Results are provided for two scenarios: (i) concordance between radiomic features extracted from 2-D SE and from 3-D SPGR acquisitions, which measures feature agreement between pulse sequences, and (ii) concordance between radiomic features extracted from contours drawn by two different observers on (iia) 2-D SE or (iib) 3-D SPGR images, which captures feature sensitivity to inter-observer variability. For each of these scenarios, there were 105 radiomic features and 264 combinations of preprocessing decisions to consider. Our recommendations for radiomic study design are provided based on broad patterns of concordance displayed in these features and processing combinations, a subset of which is displayed in the following figures. These recommendations are presented in the form of flowcharts, and the quantitative effects of each study design decision are summarized graphically.

The extended data can also inform more specific decisions in the design of MR acquisitions and analysis by using the concordance of radiomic features for pulse sequences, contouring methods, and preprocessing methods of interest. The complete set of results can be found divided into several figures in the extended data. Alternatively, these data can be viewed as single next-generation clustered heat maps (NG-CHM) via an interactive viewer at [https://www.ngchm.net](https://www.ngchm.net) using the provided NG-CHM files for CCC between features from 2-D SE and 3-D SPGR data, CCC between different lesion segmentations on 2-D SE data, and CCC between different lesion segmentations on 3-D SPGR data. The interactive NG-CHMs allow the convenient visualization of the results in their entirety in order to provide further information with which to support the design of radiomics studies.

**Radiomic features extracted from 2-D SE and 3-D SPGR images**

CCC was computed between radiomic feature values extracted from 2-D SE and 3-D SPGR images. A value of 1 indicates perfect concordance, -1 indicates perfect discordance, and 0 indicates complete absence of concordance. Figure 2 shows the clustered heat map of CCC values computed for 2-D radiomic features extracted from original images and various preprocessing combinations, including spatial normalization to voxel sizes of 0.4297 × 0.4297 × 5 mm (vs435), 1 × 1 × 5 mm (vs15), 3 × 3 × 3 mm (vs33), and 1 × 1 × 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256). Figure 3 contains the clustered CCC values for 3-D radiomic features in the same combinations. Table 4 lists features with consistently high concordance between values extracted from 2-D SE and 3-D SPGR images. Heat maps of CCC values computed for filtered images, including Laplacian of Gaussian (LoG) with sigma values ranging from 0.5 mm to 5 mm and wavelet (LLH, LHL, LHH, HLL, HLH, HHH, and LLL), as well as square, square root, logarithm, exponential, and gradient normalization, are included in the extended data. Kernel density estimates of CCC distributions are also computed and displayed in the extended data for all combinations of spatial normalization and intensity discretization in original images.

Two-sample K-S goodness of fit tests were performed between radiomic feature values extracted from 2-D SE images and those from 3-D SPGR images for each radiomic feature and combination of preprocessing methods. Figures 4 and 5 show clustered binary heat maps of 2-D and 3-D features, respectively, and combinations of preprocessing methods for which the K-S test resulted in p < 0.05, indicating rejection of the null hypothesis that the two sets of radiomic feature values were from the same distribution. Similarly, Levene’s test for equality of variances was performed between radiomic feature values extracted from 2-D SE images and those from 3-D SPGR images for each radiomic feature and combination of preprocessing methods. These results are included in the extended data.

By hierarchically clustering CCC for each feature and preprocessing variable, low-concordance combinations can be identified and recommended for exclusion in future radiomics studies employing both pulse sequences in the acquisition...
of T1-weighted post-contrast images in order to improve reproducibility. Figure 6 summarizes recommendations for radiomic feature extraction from T1-weighted post-contrast images acquired by mixed pulse sequences, and Figure 7 groups CCC values to show the effects of each recommendation. The median CCC between features extracted from

Figure 2. Hierarchically clustered heat map of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from original 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

Figure 3. Hierarchically clustered heat map of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from original 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).
images acquired by 2-D SE and 3-D SPGR for recommended combinations of features and preprocessing variables was 0.79, compared to a median CCC of 0.51 for those combinations that are not recommended.

Radiomic features extracted from different segmentations
DSC was used to measure the similarity of ROIs delineated by two different physicians. On 2-D SE images, mean DSC was 0.9072 and standard deviation was 0.0574. On 3-D SPGR images, mean DSC was 0.9067 and standard deviation was 0.0456. This suggests good overall agreement between segmentation, which is suitable for assessing feature sensitivity to inter-observer variability. CCC was also computed between radiomic feature values extracted from two different segmentations of the tumor volume on 2-D SE images and separately on 3-D SPGR images. Figure 8 shows the clustered heat map of CCC values computed for original 2-D SE images and differing preprocessing combinations, including

Table 4. Classes and features with consistently high concordance between 2-D SE and 3-D SPGR series.

| Shape features | GLCM features | GLRLM features | GLSZM features | GLDM features |
|----------------|---------------|----------------|----------------|---------------|
| Least axis length | Informational measure of correlation 1 | Long run length emphasis | Zone percentage | Dependence non-uniformity |
| Maximum 2-D diameter slice | Informational measure of correlation 2 | Run length non-uniformity normalized | Large dependence emphasis |
| Minor axis length | Inverse difference moment | Run percentage | Small dependence emphasis |
| Sphericity | Inverse difference moment normalized | Run variance |
| Surface area | Inverse difference | Short run emphasis |
| Surface volume ratio | Inverse difference normalized |
| | Inverse variance |
spatial normalization to varying voxel sizes, bin counts for relative intensity discretization, and bin widths for absolute intensity discretization. Figure 9 shows the same for 3-D features and original 3-D SPGR images. Table 5 lists the features with consistently low concordance between values extracted from the two segmentations on 3-D SPGR images. Results for 2-D SE and 3-D SPGR LoG- and wavelet-filtered images, as well as square, square root, logarithm, exponential, and gradient normalizations, are provided in extended data. Kernel density estimates of these CCC distributions from 2-D SE and 3-D SPGR images are computed and displayed in the extended data for all combinations of spatial normalization and intensity discretization in original images.

For each radiomic feature and combination of preprocessing methods, two-sample K-S goodness of fit tests were performed between radiomic feature values extracted from two different segmentations of the tumor volume on 2-D SE images and separately on 3-D SPGR images. Only 12 total combinations of features and preprocessing methods from both 2-D SE and 3-D SPGR images resulted in K-S tests with p < 0.05, indicating rejection of the null hypothesis that the two sets of radiomic feature values are from the same distribution. Detailed figures are shown in the extended data. Similarly, Levene’s test for equality of variances was performed between radiomic feature values extracted from the two sets of tumor volume segmentations for each radiomic feature, each combination of preprocessing methods, and both 2-D SE and 3-D SPGR images. Seventeen features that consistently met the Levene's test significance threshold (p < 0.05) between two different lesion segmentations on 3-D SPGR series are listed in Table 6. Detailed results are again provided in extended data.

Low-concordance combinations of features and preprocessing variables were again identified and recommended for exclusion from future studies by hierarchically clustering CCC values for all combinations. Figure 10 summarizes recommendations for radiomic feature extraction from T1-weighted post-contrast images based on variability introduced by lesion segmentation methods, and Figure 11 groups CCC values to show the effects of each recommendation made. The median CCC between features extracted from two different lesion segmentations was 0.93 for images acquired by 2-D SE and 0.86 for images acquired by 3-D SPGR. By applying additional recommendations to feature extraction from 3-D SPGR data, median CCC of recommended features and preprocessing variables was increased to 0.93.

Discussion
This study compared the resulting MR-based radiomic features between two common T1-weighted MR pulse sequences (2-D SE and 3-D SPGR) for brain metastases imaging that were acquired in the same patient during the same imaging study. It highlights the impact of image acquisition and processing on the identified radiomic features by investigating
feature variability as a function of pulse sequence, spatial normalization, 2-D or 3-D feature extraction, intensity discretization, image filters, and inter-observer variability in segmentation. Repeatability and reproducibility studies of radiomic features are necessary to assess the generalizability of radiomics applications. For example, machine learning classifiers trained on radiomic features may not be generalizable to other datasets if training is performed with features that perform poorly in repeatability or reproducibility metrics. Repeatability and reproducibility studies also contribute to the future design and standardization of high-quality, reliable radiomics studies. For these reasons, repeatability and reproducibility studies of radiomic features are increasingly necessary.

The images in this study were not acquired with uniform pulse sequence parameters. A previous study found that the sensitivity of second-order texture features to variation in acquisition parameters increases non-uniformly with spatial resolution across features. However, it also determined that variations in acquisitions such as number of acquisitions,

**Figure 6. Flowchart summarizing recommendations based on results from this work.** This flowchart assumes an existing dataset with post-contrast T1-weighted images and provides suggestions for selecting pre-processing and feature extraction parameters.
repetition time, echo time, and slice bandwidth did not significantly affect results of pattern discrimination above certain spatial resolution thresholds, and it identified GLCM features as the most robust to variability introduced in datasets with lower or heterogeneous spatial resolutions. Because many radiomic features are likely dependent on acquisition parameters, one review of feature repeatability and reproducibility studies recommends performing benchmarking studies on datasets with heterogeneous acquisition variables, such as slice thickness, acquisition protocols, multiple scanner manufacturers, and multiple institutions.

In these results, the fixed bin count intensity discretization method for original images generally displays greater CCC between 2-D and 3-D sequences than the fixed bin width method (Figures 2 and 3). This result reaffirms the widespread preference for relative discretization methods when processing MR images. Fixed bin methods introduce some normalization to images for which intensity units are arbitrary and contrast is important. Features with consistently high concordance between 2-D SE and 3-D SPGR series are listed in Table 4. Some features, such as large area emphasis, large area high gray level emphasis, and zone variance from the GLSZM class, demonstrated high variability in concordance between 2-D SE and 3-D SPGR series. For these features, concordance may be sensitive to preprocessing variables but generally may not be robust to variation in acquisition parameters or pulse sequence. In a mathematical phantom, Ford et al. determined that radiomic features varied significantly between SE and SPGR pulse sequences, which is supported by these results.

When considering features extracted from filtered images, the LoG filter, for all tested values of sigma, generally resulted in greater concordance between 2-D SE and 3-D SPGR sequences than other filters (Figures S-1 and S-2). In particular,
normalization to square, square root, exponential, and gradient scales resulted in very poor concordance between 2-D SE and 3-D SPGR features (Figures S-3 and S-4), so these are generally not recommended in heterogeneous datasets.

**Figure 8.** Hierarchically clustered heat map of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from two different lesion segmentations on original 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

**Figure 9.** Hierarchically clustered heat map of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from two different lesion segmentations on original 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).
For spatial normalization, kernel density estimates of CCC distributions among radiomic features (Figure S-12) show that 2-D radiomic features extracted at the non-isotropic resolutions \((0.4297 \times 0.4297 \times 5 \text{ mm} \text{ and } 1 \times 1 \times 5 \text{ mm})\) possess higher concordance as a whole between 2-D SE and 3-D SPGR features. 3-D radiomic features extracted at non-isotropic resolutions possess lower concordance as a whole, which is expected since this eliminates rotational invariance of features. For this reason, this combination is not recommended, and in-plane computation of radiomic features should be employed for non-isotropic spatial normalization. For both 2-D and 3-D radiomic feature extraction, the isotropic resolutions \((3 \times 3 \times 3 \text{ mm} \text{ and } 1 \times 1 \times 1 \text{ mm})\) showed lower concordance as a whole between 2-D SE and 3-D SPGR. This suggests a significant penalty to agreement between features from the two sequences as a result of interpolation between slices.

From the two-sample K-S tests, several features met the significance threshold \((p < 0.05)\) for being sampled from different distributions (Figures 4 and 5). This suggests that these features should be treated with caution, as they may not be robust to differences between the acquisition parameters and pulse sequences. Several of the same features met the significance threshold \((p < 0.05)\) in Levene’s test for equality of variances (Figures S-17 and S-18). Additional features displayed statistically significantly different variances between 2-D SE and 3-D SPGR features but did not reach significance for the K-S test. These features may be more stable for images acquired with one pulse sequence compared to the other.

Nearly all features and combinations of preprocessing methods demonstrated very high concordance between the two segmentations drawn by different physicians on the 2-D SE images (Figures 8, S-21, S-22, S-23, S-24, and S-25). This is encouraging as it indicates that almost all features were robust to inter-observer variability in segmentation of the tumor volume. Likewise, most features and combinations of preprocessing methods showed high concordance between segmentations on 3-D SE images as well (Figures 9 and S-26). However, several features yielded consistently low concordance between observers (Table 5). This indicates that these features may be poor choices for applications trained on data acquired with 3-D SPGR sequences, as they likely are not robust to inter-observer variability during segmentation. Similar features displayed low concordance for filtered 3-D SPGR images (Figures S-27, S-28, S-29, and S-30). Again, most features demonstrated very low concordance for square, square root, exponential, and gradient treatments.

| Table 5. Classes and features with consistently low concordance between two different lesion segmentations on 3-D SPGR series. |
|---|
| **First order features** | **GLCM features** | **GLRLM features** | **GLSZM features** | **GLDM features** |
| Entropy | Joint energy | Gray level non-uniformity normalized | Large area emphasis | Large dependence emphasis |
| Kurtosis | Maximum probability | Run variance | Large area high gray level emphasis | |
| Uniformity | Sum entropy | | Zone variance | |

| Table 6. Classes and features that consistently meet Levene’s test for equality of variances significance threshold \((p < 0.05)\) between two different lesion segmentations on 3-D SPGR series. |
|---|
| **First order features** | **GLRLM features** | **GLSZM features** | **GLDM features** |
| Kurtosis | Joint entropy | Large area emphasis | Gray level non-uniformity |
| Skewness | Gray level non-uniformity | Large area high gray level emphasis | Large dependence emphasis |
| Uniformity | Gray level non-uniformity normalized | Large area low gray level emphasis | |
| Long run emphasis | Zone variance | |
| Run length non-uniformity normalized | | |
| Run percentage | | |
| Run variance | | |
| Short run variance | | |

For spatial normalization, kernel density estimates of CCC distributions among radiomic features (Figure S-12) show that 2-D radiomic features extracted at the non-isotropic resolutions \((0.4297 \times 0.4297 \times 5 \text{ mm} \text{ and } 1 \times 1 \times 5 \text{ mm})\) possess higher concordance as a whole between 2-D SE and 3-D SPGR features. 3-D radiomic features extracted at non-isotropic resolutions possess lower concordance as a whole, which is expected since this eliminates rotational invariance of features. For this reason, this combination is not recommended, and in-plane computation of radiomic features should be employed for non-isotropic spatial normalization. For both 2-D and 3-D radiomic feature extraction, the isotropic resolutions \((3 \times 3 \times 3 \text{ mm} \text{ and } 1 \times 1 \times 1 \text{ mm})\) showed lower concordance as a whole between 2-D SE and 3-D SPGR. This suggests a significant penalty to agreement between features from the two sequences as a result of interpolation between slices.

From the two-sample K-S tests, several features met the significance threshold \((p < 0.05)\) for being sampled from different distributions (Figures 4 and 5). This suggests that these features should be treated with caution, as they may not be robust to differences between the acquisition parameters and pulse sequences. Several of the same features met the significance threshold \((p < 0.05)\) in Levene’s test for equality of variances (Figures S-17 and S-18). Additional features displayed statistically significantly different variances between 2-D SE and 3-D SPGR features but did not reach significance for the K-S test. These features may be more stable for images acquired with one pulse sequence compared to the other.

Nearly all features and combinations of preprocessing methods demonstrated very high concordance between the two segmentations drawn by different physicians on the 2-D SE images (Figures 8, S-21, S-22, S-23, S-24, and S-25). This is encouraging as it indicates that almost all features were robust to inter-observer variability in segmentation of the tumor volume. Likewise, most features and combinations of preprocessing methods showed high concordance between segmentations on 3-D SE images as well (Figures 9 and S-26). However, several features yielded consistently low concordance between observers (Table 5). This indicates that these features may be poor choices for applications trained on data acquired with 3-D SPGR sequences, as they likely are not robust to inter-observer variability during segmentation. Similar features displayed low concordance for filtered 3-D SPGR images (Figures S-27, S-28, S-29, and S-30). Again, most features demonstrated very low concordance for square, square root, exponential, and gradient treatments.
normalization scales, suggesting that these are not reliable methods for processing MR images for feature extraction. As a whole, fixed bin width methods resulted in greater concordance between the two segmentations for 3-D SPGR series (Figures 9, S-32, S-37, and S-38). This might be explained by the inclusion or exclusion of voxels at the segmentation boundaries having less impact on the bins into which interior voxels fall for fixed bin widths than for fixed bin counts. Still, it is unlikely to be advantageous to sacrifice the normalizing effect of fixed bin count methods for MR data for better robustness to inter-observer variability from segmentation. If fixed bin width intensity discretization is used, it is important to include an appropriate intensity normalization method in image pre-processing.34

For 2-D SE data, spatial normalization did not appear to have a significant impact on agreement between features extracted from the two segmentations (Figures S-33, S-35, and S-36). For 3-D SPGR data, only the 3 × 3 × 3-mm spatial normalization appeared to perform consistently poorly compared to the other resolutions tested (Figures S-34, S-37, and S-38). This resolution results in a significant amount of averaging of the original image data, and therefore loss of feature concordance is unsurprising.

For two-sample K-S tests between feature values from the two segmentations, only six features and preprocessing methods met the significance threshold (p < 0.05) for being sampled from different distributions for 2-D SE data (Figures S-39 and S-40), and six features and preprocessing methods met the significance threshold for 3-D SPGR data.

**Figure 10. Flowchart summarizing recommendations based on results from this work.** This flowchart assumes an imaging protocol with post-contrast T1-weighted images is being designed before data collection and provides suggestions for selecting pulse sequence and pre-processing and feature extraction parameters.
However, a few features from the 2-D SE data met the significance threshold ($p < 0.05$) for Levene’s test for equality of variances (Figures S-43 and S-44), including long run high gray level emphasis and long run low gray level emphasis from the GLRLM class and large area high gray level emphasis, large area low gray level emphasis, and small area low gray level emphasis from the GLSZM class. Several more features from the 3-D SPGR data consistently met this significance threshold (Figures S-45 and S-46). This suggests that features extracted from 2-D SE images are generally more robust to inter-observer variability during segmentation than those from 3-D SPGR images. This may result from better lesion conspicuity or higher signal-to-noise ratio (SNR) in 2-D SE acquisitions. The features listed above may suffer instability in 3-D SPGR images as a result variability introduced during segmentation of the tumor volume.

The results from this study are summarized into a series of recommendations that are shown as flowcharts in Figures 6 and 10. Figure 6 walks through decisions for an existing dataset with post-contrast T1-weighted images and provides suggestions for selecting pre-processing and feature extraction parameters. For T1-weighted post-contrast images acquired by both SE and SPGR sequences, fixed bin count intensity discretization is strongly recommended, non-isotropic spatial normalization with comparable in-plane resolution and 2-D radiomic feature extraction are potentially beneficial, and exclusion of low-concordance feature groups, such as certain wavelet filters (HLH, HHL, and HHH) and square root, logarithm, and exponential normalization, is strongly recommended. Figure 7 demonstrates the effect of each individual recommendation and the cumulative impact of all recommendations. The strong recommendations above resulted in substantial improvement in CCC, and the potential recommendations resulted in more modest improvement.
Figure 10 considers decisions involved in designing an imaging protocol with post-contrast T1-weighted images before data collection and provides suggestions for selecting pulse sequences. For any segmentation with significant variability, especially manual or semi-automated methods, 2-D SE acquisitions are strongly preferred to reduce sensitivity to inter-observer variability. Fixed bin count intensity discretization and exclusion of low-concordance feature groups, such as HHL and HHH wavelet filters and square root and logarithm normalization, are strongly recommended for 2-D SE data. If 3-D SPGR acquisitions are included, fixed bin width intensity discretization may potentially reduce feature sensitivity to segmentation variability but requires appropriate intensity normalization. Exclusion of low-concordance features in Table 5 results in a comparable distribution of CCC values to those obtained from segmentations on images from 2-D SE acquisitions, so this is strongly recommended. Figure 11 again compares the effects of each individual decision in this process and the cumulative effect of all recommendations if 3-D SPGR acquisitions are included.

Some limitations to this work warrant further study. Although the sample set was small, the goal of this study was to leverage a unique dataset to compare the effect of two pulse sequences within the same patient and same tumors imaged in the same scanner in a single imaging session on the resulting radiomic features. Given that consensus recommendations on brain metastases protocols that include both 2-D SE and 3-D SPGR T1+C acquisitions are relatively recent, we are unlikely to be able to assemble a larger dataset similar to this one retrospectively. Over 85% of patients from the original cohort were rejected because they did not have both T1-weighted post-contrast series in the same exam. Second, the scanner models and acquisition parameters were not uniform across patients included in this study. In typical test-retest studies, dataset heterogeneity is often considered to be a weakness. Because this study does not strictly compare identical measurements, but rather concordance of similar measurements, we believe increased generalizability better protects against dataset variable dependence in the results. Several previous studies have pointed out the need for balance between dataset homogeneity, which affords low noise in order to detect radiomic signatures, and heterogeneity, which offers against dataset variable dependence in the results. Several previous studies have pointed out the need for balance between dataset homogeneity, which affords low noise in order to detect radiomic signatures, and heterogeneity, which offers increased generalizability for application to real-world datasets. However, it is important to note potential drawbacks of dataset heterogeneity. Third, the main goal of this study was to study feature concordance between two different pulse sequences in brain metastasis imaging, but several additional variables affect reproducibility, such as intensity normalization, and would be useful to focus on in future work. These variables introduce additional considerations, such as white matter segmentation dependence, effects on image texture, and tumor size-dependent distortion. Finally, it would be useful to complete the picture of brain metastasis imaging protocols by investigating other commonly employed sequences, such as T2 FLAIR.

Conclusions
MR-based radiomic features that demonstrate high concordance between values extracted from images acquired with different pulse sequences may be more reliable and robust inputs to feature-based models that assist with clinical decision making. Similarly, those with high concordance between feature values extracted from different tumor volume segmentations may be more stable against variability introduced during segmentation. Fixed bin count intensity discretization demonstrated higher concordance between features extracted from 2-D SE and 3-D SPGR images, which agrees with common recommendations for MR-based radiomic feature extraction. Non-isotropic spatial normalization was found to have higher concordance between features extracted from 2-D SE and 3-D SPGR images. This study found that the 2-D SE pulse sequence was more robust to inter-observer variability in tumor volume segmentation than the 3-D SPGR pulse sequence. We use these results to provide comprehensive recommendations for preprocessing in future radiomics studies with heterogeneous imaging data.

Data availability
Underlying data
Figshare: Extended Data for Managing the Effect of Magnetic Resonance Imaging Pulse Sequence on Radiomic Feature Reproducibility in the Study of Brain Metastases, https://doi.org/10.6084/m9.figshare.c.6039128.v1.

This project contains the following underlying data:

- pulsesquences_ccc_ngchm.ngchm. (The complete set of results for CCC between features from 2-D SE and 3-D SPGR data. These data can be viewed as a single next-generation clustered heat map (NG-CHM) via an interactive viewer at https://www.ngchm.net. The interactive NG-CHMs allow the convenient visualization of the results in their entirety in order to provide further information with which to support the design of radiomics studies, e.g. selection of pulse sequences, contouring methods, and preprocessing methods of interest.)

- segmentation_2dse_ccc_ngchm.ngchm. (The complete set of results for CCC between different lesion segmentations on 2-D SE data. These data can be viewed as a single next-generation clustered heat map (NG-CHM) via an interactive viewer at https://www.ngchm.net. The interactive NG-CHMs allow the convenient
visualization of the results in their entirety in order to provide further information with which to support the
design of radiomics studies, e.g. selection of pulse sequences, contouring methods, and preprocessing methods
of interest.)

- segmentation_3dspgr_ccc_ngchm.ngchm. (The complete set of results for CCC between different lesion
segmentations on 3-D SPGR data. These data can be viewed as a single next-generation clustered heat map
(NG-CHM) via an interactive viewer at https://www.ngchm.net. The interactive NG-CHMs allow the conve-
nient visualization of the results in their entirety in order to provide further information with which to support the
design of radiomics studies, e.g. selection of pulse sequences, contouring methods, and preprocessing methods
of interest.)

Data that cannot be shared
The following data cannot be shared due to restrictions on data sharing in the IRB protocol. Individuals may contact the
corresponding author to apply for access to the data, which will be granted upon IRB approval.

- 29 T1-weighted post-contrast image series acquired by 2-D spin echo sequences.
- 29 T1-weighted post-contrast image series acquired by 3-D spoiled gradient echo sequences.

Extended data
Figshare: Extended Data for Managing the Effect of Magnetic Resonance Imaging Pulse Sequence on Radiomic Feature
Reproducibility in the Study of Brain Metastases. https://doi.org/10.6084/m9.figshare.c.6039128.v1

This project contains the following extended data:

- Figure S-1.tif. (Figure S - 1. Heat map of concordance correlation coefficient (CCC) between 2-D radiomic
feature values extracted from Laplacian of Gaussian (LoG) filtered 2-D SE and 3-D SPGR images. Preproces-
sing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm
(vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and LoG filter sigma values from 0.5 mm to 5.0 mm in
increments of 0.5 mm.)
- Figure S-2.tif. (Figure S - 2. Heat map of concordance correlation coefficient (CCC) between 3-D radiomic
feature values extracted from Laplacian of Gaussian (LoG) filtered 2-D SE and 3-D SPGR images. Preproces-
sing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm
(vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and LoG filter sigma values from 0.5 mm to 5.0 mm in
increments of 0.5 mm.)
- Figure S-3.tif. (Figure S - 3. Heat map of concordance correlation coefficient (CCC) between 2-D radiomic
feature values extracted from wavelet filtered and square normalized 2-D SE and 3-D SPGR images. Preproces-
sing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm
(vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and low- and high-pass wavelet filter combinations
(LLH, LHL, LHH, HLL, HLH, HHL, HHH, and LLL.).)
- Figure S-4.tif. (Figure S - 4. Heat map of concordance correlation coefficient (CCC) between 3-D radiomic
feature values extracted from wavelet filtered and square, square root, logarithm, exponential, and gradient
normalized 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel
sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and
low- and high-pass wavelet filter combinations (LLH, LHL, LHH, HLL, HLH, HHL, HHH, and LLL.).)
- Figure S-5.tif. (Figure S - 5. Heat map with clustering of concordance correlation coefficient (CCC) between
2-D radiomic feature values extracted from original 2-D SE and 3-D SPGR images. Preprocessing combinations
include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x
3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32),
64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16),
32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256). Dendrograms are displayed on both axes.)
Figure S-6.tif. (Figure S-6. Heat map with clustering of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from original 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256). Dendrograms are displayed on both axes.)

Figure S-7.tif. (Figure S-7. Heat map with clustering of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from Laplacian of Gaussian (LoG) filtered 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and LoG filter sigma values from 0.5 mm to 5.0 mm in increments of 0.5 mm. Dendrograms are displayed on both axes.)

Figure S-8.tif. (Figure S-8. Heat map with clustering of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from Laplacian of Gaussian (LoG) filtered 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and LoG filter sigma values from 0.5 mm to 5.0 mm in increments of 0.5 mm. Dendrograms are displayed on both axes.)

Figure S-9.tif. (Figure S-9. Heat map with clustering of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from wavelet filtered and square normalized 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and low- and high-pass wavelet filter combinations (LLH, LHL, LHH, HLL, HLH, HHL, HH, and LLL). Dendrograms are displayed on both axes.)

Figure S-10.tif. (Figure S-10. Heat map with clustering of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from wavelet filtered and square, square root, logarithm, exponential, and gradient normalized 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and low- and high-pass wavelet filter combinations (LLH, LHL, LHH, HLL, HLH, HHL, HH, and LLL). Dendrograms are displayed on both axes.)

Figure S-11.tif. (Figure S-11. Kernel density estimates (KDE) of concordance correlation coefficient (CCC) distributions between radiomic feature values extracted from 2-D SE and 3-D SPGR images. Subplots group together distributions computed from various intensity discretization methods, including fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

Figure S-12.tif. (Figure S-12. Kernel density estimates (KDE) of concordance correlation coefficient (CCC) distributions between radiomic feature values extracted from 2-D SE and 3-D SPGR images. Subplots group together distributions computed from 2-D and 3-D radiomic feature extraction with various spatial normalization methods, including voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and low- and high-pass wavelet filter combinations (LLH, LHL, LHH, HLL, HLH, HHL, HH, and LLL). Dendrograms are displayed on both axes.)

Figure S-13.tif. (Figure S-13. Number of radiomic features extracted from original images with concordance correlation coefficient (CCC) falling into various ranges. CCC is computed between radiomic features values extracted from 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and 2-D (f2d) or 3-D (f3d) feature extraction.)

Figure S-14.tif. (Figure S-14. Number of radiomic features extracted from original images with concordance correlation coefficient (CCC) falling into various ranges. CCC is computed between radiomic features values extracted from 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256), and 2-D (f2d) or 3-D (f3d) feature extraction.)
Figure S-15.tif. (Figure S - 15. Binary heat map with clustering of 2-D radiomic features and preprocessing combinations for which p < 0.05 from Kolmogorov-Smirnov test computed between radiomic feature values extracted from 2-D SE and 3-D SPGR images. The dendrogram grouping features is displayed. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

Figure S-16.tif. (Figure S - 16. Binary heat map with clustering of 3-D radiomic features and preprocessing combinations for which p < 0.05 from Kolmogorov-Smirnov test computed between radiomic feature values extracted from 2-D SE and 3-D SPGR images. The dendrogram grouping features is displayed. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

Figure S-17.tif. (Figure S - 17. Binary heat map of 2-D radiomic features and preprocessing combinations for which p < 0.05 from Levene's test for equality of variances computed between radiomic feature values extracted from 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

Figure S-18.tif. (Figure S - 18. Binary heat map of 3-D radiomic features and preprocessing combinations for which p < 0.05 from Levene's test for equality of variances computed between radiomic feature values extracted from 2-D SE and 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

Figure S-19.tif. (Figure S - 19. Binary heat map with clustering of 2-D radiomic features and preprocessing combinations for which p < 0.05 from Levene's test for equality of variances computed between radiomic feature values extracted from 2-D SE and 3-D SPGR images. The dendrogram grouping features is displayed. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

Figure S-20.tif. (Figure S - 20. Binary heat map with clustering of 3-D radiomic features and preprocessing combinations for which p < 0.05 from Levene's test for equality of variances computed between radiomic feature values extracted from 2-D SE and 3-D SPGR images. The dendrogram grouping features is displayed. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

Figure S-21.tif. (Figure S - 21. Heat map of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from two different lesion segmentations on original 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).

Figure S-22.tif. (Figure S - 22. Heat map of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from two different lesion segmentations on Laplacian of Gaussian (LoG) filtered 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and LoG filter sigma values from 0.5 mm to 5.0 mm in increments of 0.5 mm.)
• Figure S-23.tif. (Figure S - 23. Heat map of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from two different lesion segmentations on Laplacian of Gaussian (LoG) filtered 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and LoG filter sigma values from 0.5 mm to 5.0 mm in increments of 0.5 mm.)

• Figure S-24.tif. (Figure S - 24. Heat map of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from two different lesion segmentations on wavelet filtered and square normalized 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and low- and high-pass wavelet filter combinations (LLH, LHL, LHH, HLL, HLH, HHL, HHH, and HHH).)

• Figure S-25.tif. (Figure S - 25. Heat map of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from two different lesion segmentations on wavelet filtered and square, square root, logarithm, exponential, and gradient normalized 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), and low- and high-pass wavelet filter combinations (LLH, LHL, LHH, HLL, HLH, HHL, HHH, and HHH).)

• Figure S-26.tif. (Figure S - 26. Heat map of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from two different lesion segmentations on original 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

• Figure S-27.tif. (Figure S - 27. Heat map of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from two different lesion segmentations on Laplacian of Gaussian (LoG) filtered 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

• Figure S-28.tif. (Figure S - 28. Heat map of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from two different lesion segmentations on Laplacian of Gaussian (LoG) filtered 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

• Figure S-29.tif. (Figure S - 29. Heat map of concordance correlation coefficient (CCC) between 2-D radiomic feature values extracted from two different lesion segmentations on wavelet filtered and square, square root, logarithm, exponential, and gradient normalized 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

• Figure S-30.tif. (Figure S - 30. Heat map of concordance correlation coefficient (CCC) between 3-D radiomic feature values extracted from two different lesion segmentations on wavelet filtered and square, square root, logarithm, exponential, and gradient normalized 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

• Figure S-31.tif. (Figure S - 31. Kernel density estimates (KDE) of concordance correlation coefficient (CCC) distributions between radiomic feature values extracted from two different lesion segmentations on 2-D SE images. Subplots group together distributions computed from various intensity discretization methods, including fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)
• Figure S-32.tif. (Figure S - 32. Kernel density estimates (KDE) of concordance correlation coefficient (CCC) distributions between radiomic feature values extracted from two different lesion segmentations on 3-D SPGR images. Subplots group together distributions computed from various intensity discretization methods, including fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

• Figure S-33.tif. (Figure S - 33. Kernel density estimates (KDE) of concordance correlation coefficient (CCC) distributions between radiomic feature values extracted from two different lesion segmentations on 2-D SE images. Subplots group together distributions computed from 2-D and 3-D radiomic feature extraction with various spatial normalization methods, including voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11).)

• Figure S-34.tif. (Figure S - 34. Kernel density estimates (KDE) of concordance correlation coefficient (CCC) distributions between radiomic feature values extracted from two different lesion segmentations on 3-D SPGR images. Subplots group together distributions computed from 2-D and 3-D radiomic feature extraction with various spatial normalization methods, including voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11).)

• Figure S-35.tif. (Figure S - 35. Number of radiomic features extracted from original images with concordance correlation coefficient (CCC) falling into various ranges. CCC is computed between radiomic features values extracted from two different lesion segmentations on 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and 2-D (f2d) or 3-D (f3d) feature extraction.)

• Figure S-36.tif. (Figure S - 36. Number of radiomic features extracted from original images with concordance correlation coefficient (CCC) falling into various ranges. CCC is computed between radiomic features values extracted from two different lesion segmentations on 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256), and 2-D (f2d) or 3-D (f3d) feature extraction.)

• Figure S-37.tif. (Figure S - 37. Number of radiomic features extracted from original images with concordance correlation coefficient (CCC) falling into various ranges. CCC is computed between radiomic features values extracted from two different lesion segmentations on 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and 2-D (f2d) or 3-D (f3d) feature extraction.)

• Figure S-38.tif. (Figure S - 38. Number of radiomic features extracted from original images with concordance correlation coefficient (CCC) falling into various ranges. CCC is computed between radiomic features values extracted from two different lesion segmentations on 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256), and 2-D (f2d) or 3-D (f3d) feature extraction.)

• Figure S-39.tif. (Figure S - 39. Binary heat map of 2-D radiomic features and preprocessing combinations for which p < 0.05 from Kolmogorov-Smirnov test computed between radiomic feature values extracted from two different lesion segmentations on 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

• Figure S-40.tif. (Figure S - 40. Binary heat map of 3-D radiomic features and preprocessing combinations for which p < 0.05 from Kolmogorov-Smirnov test computed between radiomic feature values extracted from two different lesion segmentations on 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11).)
Figure S-41.tif. (Figure S - 41. Binary heat map of 2-D radiomic features and preprocessing combinations for which \( p < 0.05 \) from Kolmogorov-Smirnov test computed between radiomic feature values extracted from two different lesion segmentations on 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

Figure S-42.tif. (Figure S - 42. Binary heat map of 3-D radiomic features and preprocessing combinations for which \( p < 0.05 \) from Kolmogorov-Smirnov test computed between radiomic feature values extracted from two different lesion segmentations on 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

Figure S-43.tif. (Figure S - 43. Binary heat map of 2-D radiomic features and preprocessing combinations for which \( p < 0.05 \) from Levene's test for equality of variances computed between radiomic feature values extracted from two different lesion segmentations on 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

Figure S-44.tif. (Figure S - 44. Binary heat map of 3-D radiomic features and preprocessing combinations for which \( p < 0.05 \) from Levene's test for equality of variances computed between radiomic feature values extracted from two different lesion segmentations on 2-D SE images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

Figure S-45.tif. (Figure S - 45. Binary heat map of 2-D radiomic features and preprocessing combinations for which \( p < 0.05 \) from Levene's test for equality of variances computed between radiomic feature values extracted from two different lesion segmentations on 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

Figure S-46.tif. (Figure S - 46. Binary heat map of 3-D radiomic features and preprocessing combinations for which \( p < 0.05 \) from Levene's test for equality of variances computed between radiomic feature values extracted from two different lesion segmentations on 3-D SPGR images. Preprocessing combinations include spatial normalization to voxel sizes of 0.4297 x 0.4297 x 5 mm (vs435), 1 x 1 x 5 mm (vs15), 3 x 3 x 3 mm (vs33), and 1 x 1 x 1 mm (vs11), intensity discretization to fixed bin counts of 16 (bc16), 32 (bc32), 64 (bc64), 128 (bc128), and 256 (bc256), and intensity discretization to fixed bin width of 16 (bw16), 32 (bw32), 64 (bw64), 128 (bw128), and 256 (bw256).)

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Acknowledgements
The authors would like to thank David Jaffray, Ph.D. for constructive criticism of the manuscript.
38. Hoebel KV, Patel JB, Beers AL, et al.: Radiomics repeatability pitfalls in a scan-rescan mri study of glioblastoma. Radiol. Artif. Intell. 2021; 3(1): e190199. PubMed Abstract | Publisher Full Text

39. Pandey U, Saini J, Kumar M, et al.: Normative Baseline for Radiomics in Brain MRI: Evaluating the Robustness, Regional Variations, and Reproducibility on FLAIR Images. J. Magn. Reson. Imaging. 2021; 53(2): 394–407. PubMed Abstract | Publisher Full Text

40. Eck B, Chirra PV, Muchhala A, et al.: Prospective Evaluation of Repeatability and Robustness of Radiomic Descriptors in Healthy Brain Tissue Regions in vivo Across Systematic Variations in T2-Weighted Magnetic Resonance Imaging Acquisition Parameters. J. Magn. Reson. Imaging. 2021; 54(3): 1009–1021. PubMed Abstract | Publisher Full Text

41. Kim M, Jung SC, Park JE, et al.: Reproducibility of radiomic features in SENSE and compressed SENSE: impact of acceleration factors. Eur. Radiol. 2021; 31(9): 6457–6470. PubMed Abstract | Publisher Full Text

42. Shur JD, Doran SJ, Kumar S, et al.: Radiomics in oncology: A practical guide. Radiographics. 2021; 41(6): 1717–1732. PubMed Abstract | Publisher Full Text

43. Keek SA, Leijenaar RT, Jochems A, et al.: A review on radiomics and the future of theranostics for patient selection in precision medicine. Br. J. Radiol. 2018; 91(1091): 20170926. PubMed Abstract | Publisher Full Text

44. Kumar V, Gu Y, Basu S, et al.: Radiomics: The process and the challenges. Magn. Reson. Imaging. 2012; 30(9): 1234-1248. PubMed Abstract | Publisher Full Text

45. van Griethuysen JJM, Fedorov A, Parmar C, et al.: Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017; 77(21): e104–e107. PubMed Abstract | Publisher Full Text

46. Broom BM, Ryan MC, Studky M, et al.: Interactive Clustered Heat Map Builder: An easy web-based tool for creating sophisticated clustered heat maps. F1000Res. 2020; 8: 1750. PubMed Abstract | Publisher Full Text

47. Broom BM, Ryan MC, Brown RE, et al.: A galaxy implementation of next-generation clustered heatmaps for interactive exploration of molecular profiling data. Cancer Res. 2017; 77(21): e23–e26. PubMed Abstract | Publisher Full Text

48. Mitchell D: Extended Data for Managing the Effect of Magnetic Resonance Imaging Pulse Sequence on Radiomic Feature Reproducibility in the Study of Brain Metastases. figshare. [Dataset]. 2022. Publisher Full Text
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com