Received: 19 September 2019 | Revised: 18 December 2019 | Accepted: 10 January 2020

DOI: 10.1002/mrm.28192

FULL PAPER

Magnetic Resonance in Medicine

A novel gamma GLM approach to MRI relaxometry comparisons

Rohan Kapre[1,2] | **Junhan Zhou**[3] | **Xinzhe Li**[1] | **Laurel Beckett**[2] | **Angelique Y. Louie**[1,3]

[1]Department of Biomedical Engineering, University of California, Davis, CA

[2]Biostatistics Graduate Group, University of California, Davis, CA

[3]Chemistry Graduate Group, University of California, Davis, CA

Correspondence
Angelique Y. Louie, UC Davis, Department of Biomedical Engineering, Genome and Biomedical Sciences Facility, 451 Health Sciences Dr, Davis, CA 95616.
Email: aylouie@ucdavis.edu

Funding information
National Heart, Lung, and Blood Institute, Grant/Award Number: 137387

Purpose: To demonstrate that constant coefficient of variation (CV), but nonconstant absolute variance in MRI relaxometry (T_1, T_2, R_1, R_2) data leads to erroneous conclusions based on standard linear models such as ordinary least squares (OLS). We propose a gamma generalized linear model identity link (GGLM-ID) framework that factors the inherent CV into parameter estimates. We first examined the effects on calculations of contrast agent relaxivity before broadening to other applications such as analysis of variance (ANOVA) and liver iron content (LIC).

Methods: Eight models including OLS and GGLM-ID were initially fit to data obtained on sulfated dextran iron oxide (SDIO) nanoparticles. Both a resampling simulation on the data as well as two separate Monte Carlo simulations (with and without concentration error) were performed to determine mean square error (MSE) and type I error rate. We then evaluated the performance of OLS/GGLM-ID on R_1 repeatability and LIC data sets.

Results: OLS had an MSE of 4–5× that of GGLM-ID as well as a type I error rate of 20–30%, whereas GGLM-ID was near the nominal 5% level in the relaxivity study. Only OLS found statistically significant effects of MRI facility on relaxivity in an R_1 repeatability study, but no significant differences were found in a resampling, whereas GGLM was more consistent. GGLM-ID was also superior to OLS for modeling LIC.

Conclusions: OLS leads to erroneous conclusions when analyzing MRI relaxometry data. GGLM-ID factors in the inherent CV of an MRI experiment, leading to more reproducible conclusions.

KEYWORDS

coefficient of variation, gamma GLM, MRI relaxometry, relaxivity, repeatability

[Correction added after online publication 15 June, 2020. The author has added a Data Availability Statement.]

1 | INTRODUCTION

MRI relaxometry is used increasingly as a method to quantify differences in MRI images.[1-3] Relaxometry involves mapping of relaxation times T_1, T_2, T_2^* (or alternatively, the relaxation rates R_1, R_2, R_2^* corresponding to their inverses) spatially across voxels in an MRI image. Compared to traditional MRI imaging based on T_1- or T_2-weighted signal intensity, MRI relaxometry is more robust for quantitative comparison as it is less sensitive to nuisance parameters such as coil positioning or scanner variabilities.[1,2] Among other applications, relaxometry has been used to detect brain abnormalities[1,4] and myocardial fibrosis[5] or evaluate metabolic imaging,[3] liver iron content (LIC),[6-9] and contrast agent (CA) properties.[10-14]

As MRI relaxometry becomes more widespread, there is an increasing need to develop an appropriate statistical framework to accurately quantify differences in relaxometric biomarkers, and correlate these differences to physiological phenomenon. Currently, most comparisons are performed using *t*-tests, analysis of variance (ANOVA), or ordinary least squares (OLS) regression[15-17] that all fall under the category of standard linear models.[15,18] In fact, both Student's *t*-test and ANOVA are special cases of OLS regression with categorical variables.[18,19] Beyond the linearity of the relationship, OLS has three key assumptions[18,19]:

1. The response variables (relaxation rate or time) must be independent and identically distributed.
2. The model residuals must follow a normal distribution.
3. The model residuals must all have a constant absolute variance (homoscedasticity).

Assumption 1 is automatically satisfied for relaxometry data from different scans and subjects. Otherwise, correlation can be accounted for through linear mixed models. Estimates from OLS are also somewhat robust to violations of assumption 2, especially for large sample sizes. However, assumption 3 is often violated in the context of MRI relaxometry data, where the absolute error in the relaxation time (T_1, T_2, or T_2^*) or rate (R_1, R_2, R_2^*) often tends to increase with the measured value itself (heteroscedasticity).[6,16] Much of the literature on reproducibility in MRI relaxometry is focused on measuring the coefficient of variation (CV),[6,20-22] which implies a constant relative error rather than absolute error. Various factors can influence the CV of T_i/R_i, ranging from the signal intensity fitting error, the exact pulse sequence used, the flip angle, concentration errors, as well as natural biological variability.[2,23]

Violations of assumption 3 ultimately end up leading to less precise estimates for coefficients in a linear model and more importantly, biased standard errors (SEs).[18] These biased SEs will result in invalid *P*-values for comparisons, leading to potentially erroneous conclusions.[24] Sometimes the assumptions of standard linear models are bypassed using nonparametric tests such as Wilcoxon rank-sum-test[3] or the Kruskal-Wallis test.[25] However, these two tests are limited in that they can only be applied to compare two populations or multiple levels within a single factor, respectively. Additionally, nonparametric methods are less powerful at detecting differences than parametric counterparts provided the data can be described adequately by a normal or other parametric distribution.[26,27]

Ideally, the best parametric method would assume constant CV rather than constant absolute error. One way to do so is the log transform, which is known to stabilize absolute errors that have a constant CV. This amounts to assuming that relaxometry data are lognormally distributed,[22] rather than normally distributed. In fact, the log transform has been used occasionally for both linear mixed models[4,5] and LIC calibration curves.[8,9,28] Recently, the gamma distribution has also been used to account for the constant CV in neuroimaging relaxometry data.[29] In the context of regression, the gamma distribution can be used through a generalized linear model (GLM) rather than OLS.[19,30,31] Although results will often be similar to a log-transform,[19,31,32] the gamma GLM (GGLM) framework is more flexible as it preserves the physical additivity of relaxation rates[15] and simple interpretation of the results on the original untransformed scale.

The simplest example of the linear model in MRI relaxometry occurs in experiments to determine relaxivity ($\mathrm{mM}^{-1}\ \mathrm{s}^{-1}$), a measure of the strength of a CA. A higher relaxivity implies that the CA is more effective at increasing the relaxation rate of protons surrounding its metal-ion sphere.[33-35] Using the Solomon–Bloemborgen theory of relaxation, it can be shown that relaxation rate is a linear function of CA concentration[33-35]:

$$R_i = R_{i,0} + r_i\,[CA] + \epsilon\ (i = 1,2) \tag{1}$$

$$R_i = T_i^{-1} \tag{2}$$

To determine relaxivity r_i and associated SE, investigators typically perform an OLS regression of R_i versus $[CA]$ and take the slope.[13] ϵ corresponds to the residual error term in the fit.[15,18,19] Relaxivity studies are prime examples of heteroscedasticity where the error bars tend to increase in size with the relaxation rate.[10-13] Currently, statistically suspect comparisons of relaxivities under different experimental conditions (e.g., field strength) or between CAs are regularly performed and displayed in bar charts or tables.[10-13] To date, there is no standard protocol for how to best design and analyze a relaxivity experiment.[36] Similar problems have been addressed previously in UV-Vis spectroscopy[37] and liquid chromatography-mass spectrometry (LC-MS)[38] through weighted least squares (WLS), which has also been applied on occasion to relaxivity.[12,39] Although WLS is a potential solution, the GGLM approach

proposed in this work is more flexible as it assigns weights empirically by an iterative reweighted least squares algorithm.[31] Therefore, it is more easily extended to MRI relaxometry experiments involving more complex situations other than relaxivity, such as developing quantitative imaging biomarkers,[22] where exact weights may be unknown. In this work, relaxivity measurements will be used as a starting basis for demonstrating the superiority of weighting, log-transform, and in particular, the GGLM framework over the current standard of OLS. Afterward we will move toward more general applications of GGLM in MRI relaxometry, such as ANOVA and the measurement of LIC.

2 | METHODS

2.1 | NMR relaxometry measurements

Sulfated dextran iron oxide (SDIO) nanoparticles were prepared as described in previous protocols published by our group[40,41] (Supporting Information Figure S1). Iron content was determined via inductively coupled plasma mass spectrometry. Ten solutions of SDIO in filtered deionized water (milliQ Advantage A10) with the following [Fe] concentrations were prepared: 0–0.08 mM spaced by 0.02 mM, and 0.1–0.5 mM spaced by 0.1 mM. T_2 measurements were acquired with the Carr-Purcell-Meiboom-Gill pulse sequence on a Bruker Minispec mq60 (1.41 T) at 37°C. Gain was set to 57 dB and the number of scans was set to four for each individual measurement. The entire procedure of preparing the ten solutions and obtaining T_2 measurements was repeated four times, for a total of 40 data points.

2.2 | Relaxivity model fitting

All statistical analysis was performed in R version 3.6.0.[42] Along with OLS, a variety of models were fit to the data. The linear models included WLS, GGLM-ID, GGLM-INV, and Thiel-Sen (TS) fit directly to Equation 1. The weights for WLS were defined to be $w = 1/R_2^2$, corresponding to the appropriate weight for a constant CV in R_2. GGLM-ID and GGLM-INV refer to gamma GLMs with the identity and inverse links, respectively. The identity link corresponds to applying the gamma distribution to $R_2 = 1/T_2$ values directly, whereas the inverse link applies the gamma distribution to T_2 values followed by inversion to obtain R_2.[19,30] TS is a nonparametric linear fitting method using the *mblm* package (version 0.12.1, for R)[43] involving the median of all possible slopes from pairs of points[44] at different concentrations and has been recommended once for relaxivity measurements.[45] The nonlinear fits included nonlinear least squares (NLS), nonlinear weighted least squares (NWLS), and log-transformed nonlinear least squares (LNLS). Both NLS and NWLS were fit to Equation 3 corresponding to inverse of Equation 1, whereas LNLS was fit to Equation 4 corresponding to the log transform of Equation 1. The weights for NWLS were set to $w = 1/T_2^2$. Note that NLS just as OLS does not account for heteroscedasticity and has been included only for comparison purposes, and not as a proposed model.

$$T_2 = \left(R_{2,0} + r_2\,[CA]\right)^{-1} + \epsilon \tag{3}$$

$$\log R_2 = \log\left(R_{2,0} + r_2\,[CA]\right) + \epsilon \tag{4}$$

The variance function was estimated via a Bayesian log-log linear regression of SD T_2 versus mean T_2[46] at each concentration. A Bayesian framework was used to obtain a distribution for the slope in the variance function regression with a normal prior of mean 1 and SD 0.3. This accounts for the prior knowledge that the slope will be near 1 because of an approximately constant CV. Furthermore, using a Bayesian framework allows us to obtain a distribution for the slope in the variance function that is not strictly possible in a purely frequentist approach.

2.3 | Restricted bootstrap resampling

To simulate batch to batch variability, one point at each concentration was randomly chosen and each of the models were fit to the data to obtain a relaxivity and associated SE. The process was repeated 100,000 times to obtain a histogram of possible relaxivities. These individual relaxivities and SE were compared to the full relaxivity obtained from fitting all 40 points to determine the type I error rate via a *t*-test. The SE estimate for each individual fit was also compared to the SD of the relaxivity histogram. Furthermore, the mean square error (MSE)[47] for each model was computed using the bias of the mean bootstrap relaxivity with respect to the full relaxivity and the SD of the relaxivity histogram:

$$\mathrm{Bias}(\hat{r}_2) = \mathrm{E}(\hat{r}_2) - r_2^{\mathrm{full}} \tag{5}$$

$$\mathrm{MSE}(\hat{r}_2) = \mathrm{Bias}^2(\hat{r}_2) + \mathrm{SD}^2(\hat{r}_2) \tag{6}$$

2.4 | Relaxivity Monte Carlo simulation

Two Monte Carlo simulations each with 100,000 runs were performed using the above concentrations assuming a true relaxivity of 100 $\mathrm{mM}^{-1}\,\mathrm{s}^{-1}$. A Gaussian error with a 2.5% CV was added to T_2 in the first simulation while the second simulation additionally included a Gaussian error with 10% CV in concentration. To account for deviations from perfect linear

proportionality in the error in T_2, the error was set to be constant CV of T_2^p, with P set to originate from the posterior normal distribution with slope value and associated uncertainty from the variance function determined earlier. As in the resampling, the MSE and type I error rates were computed.

2.5 | Data analysis: R_1 repeatability study ANOVA

To demonstrate an application of the GGLM-ID framework, we analyzed published relaxometry data at 7T from Waterton et al[20] on the repeatability/reproducibility of R_1 measurements in $NiCl_2$ phantoms across multiple MRI facilities. The concentrations of Ni^{2+} tested in the study were 0.50, 1.04, 2.02, 4.08, and 8.05 mM in agarose. The R_1 measurements were performed at five different facilities (A, C, E, F, G1) and three ROIs (O, X, Z) with saturation recovery using a rapid acquisition with relaxation enhancement (RARE) sequence. For this analysis, only the isocenter ROI "O" was considered (to ensure independence) and ANOVA was be performed using OLS, GGLM-ID, and LNLS frameworks with facility and concentration interaction terms considered. The −1, 0, 1 indicator variable coding with −1 denoting facility G1 was used, shown in Equations 7 and 8. Akaike Information Criterion (AIC) was used to compare the models in addition to an ANOVA F-test with respect to the null model containing only concentration and solvent (agarose) relaxation rate. Bonferroni's post-hoc test was used to evaluate statistically significant differences in relaxivity between the individual facilities using the *emmeans* package.[48]

$$R_1 = R_{1,0} + r_1\left[Ni^{2+}\right] + \left(\Delta R_{1,0,A} I_A + \cdots + \Delta R_{1,0,F} I_F\right) + \left(\Delta r_{1,A} I_A + \cdots + \Delta r_{1,F} I_F\right)\left[Ni^{2+}\right] \tag{7}$$

$$I_i = \begin{cases} 1: Facility = i \\ 0: Facility \neq i, G1 \\ -1: Facility = G1 \end{cases} \tag{8}$$

2.6 | R_2-LIC Monte Carlo simulations

One medical application of the GGLM-ID framework is the estimation of LIC via MRI relaxometry.[6,7] The original study done by St. Pierre et al used a spin-echo sequence and found a complex nonlinear relationship between R_2 and LIC[7,9]:

$$R_2 = 6.88 + 26.06 LIC^{0.701} - 0.438 LIC^{1.402} \tag{9}$$

$$LIC = \left(29.75 - \sqrt{900.7 - 2.283 R_2}\right)^{1.424} \tag{10}$$

The study noted significant heteroscedasticity with CV for R_2 given as 8%. The CV for LIC was piecewise-dependent: 19% for LIC < 4 mg/g and 9% for LIC > 9 mg/g.[7,8] Equation 10 and the above CVs were used to perform two simulations (10,000 runs) with lognormal error in LIC and with lognormal error in both R_2 and LIC. LIC values in between 4 and 9 mg/g were set to have a random CV uniformly distributed between 9% and 19%. The models considered are shown in Table 1. The true R_2 ranged from 10–275 s^{-1}, translating to a true LIC of 0.05–39.55 mg/g. For all models, the regression was defined with LIC on the y-axis and R_2 on the x-axis. The reasoning for this is that the LIC is more variable than the R_2 values from MRI, and the goal is to predict LIC and not R_2. For simplicity, we will only be examining polynomial models as the nonlinear fitting method was not reported in the original paper by St. Pierre et al.[7] The models were evaluated by examining the model with the minimum median RMS percent prediction error (MRMSPPE):

$$MRMSPPE = 100\% \times \sqrt{\text{med}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{LIC_{\text{true}} - LIC_{\text{pred}}}{LIC_{\text{true}}}\right)^2\right)} \tag{11}$$

The median helped to ensure that any outliers in the 10,000 runs did not affect the results. To further quantify the uncertainty in the MRMSPPE, we reported 95% percentile intervals. Equation 10 with the squared terms was used rather than absolute values, because these have better properties and more closely match the maximum likelihood estimate corresponding to a heteroscedastic Gaussian distribution for LIC. Finally, the OLS and GGLM-ID quadratic fits were plotted to an approximation of the R_2-LIC data obtained using PlotDigitizer.[49] To help quantify and visualize variability in the calibration curve, we also added the same lognormal error structure mentioned earlier for both R_2 and

TABLE 1 Models considered for the parametric R_2-LIC simulation based on the equation derived by St. Pierre et al[7] (Equations 10 and 11)

Model name	Functional form	Method	# of Parameters
Quadratic	$LIC = \beta_0 + \beta_1 R_2 + \beta_2 R_2^2$	OLS GGLM-ID	3
Double log linear	$\log LIC = \beta_0 + \beta_1 \log R_2$	OLS GGLM-LOG	2

The exact nonlinear fitting method performed by St. Pierre et al.[7] was is not given, but we used the equation to simulate the data for MRMSPPE calculation.

LIC directly to the approximated points. Following this, we generated 95% prediction intervals to depict variability in the calibration under different realizations of the same data set.

3 | RESULTS

The results of the SDIO relaxivity experiment are shown in Figure 1. As evidenced by the substantially lower AIC, Figure 1 suggests that the GGLM-ID framework provided a better fit to the data than OLS. Although adjusted R^2 does not exist for GLMs and is not the ideal metric to assess fit despite widespread use, a pseudo-adjusted R^2 was calculated

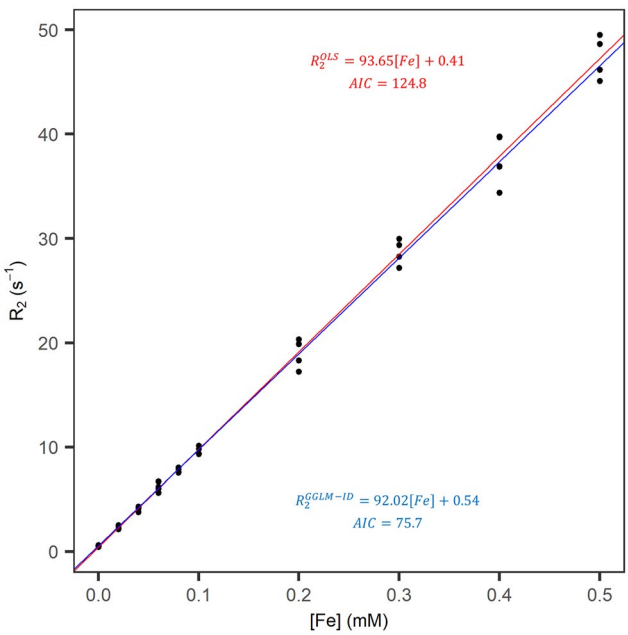

FIGURE 1 SDIO R_2 relaxivity fits using OLS and GGLM-ID. The lower AIC suggests that GGLM-ID provides a better fit. Adjusted R^2 for the OLS fit was 0.995, whereas the pseudo-adjusted R^2 for the GGLM-ID fit was 0.996. All eight models were fit to the full data but only OLS and GGLM-ID are shown in the figure

for the purposes of data reporting using the residual and null deviance and degrees of freedom.[19] Furthermore, the spread in R_2 increased with R_2 itself. To determine the dependence, we performed a log-log Bayesian linear regression of $\log(\mathrm{SD}(T_2))$ versus $\log(T_2)$ (Supporting Information Figure S2). The Bayesian linear regression for the variance function revealed that the slope has a 95% probability of being between 0.94 and 1.41. This indicates that $\mathrm{SD}(T_2)$ may not be exactly proportional to T_2 but to a power in T_2 between 0.94 and 1.41.

Because many relaxivity studies do only one trial, one point was resampled at each concentration and fit to all models to assess model performance with MSE and type I error rates. Results are shown in Table 2 and a histogram of the resampled relaxivities is shown in Figure 2. Histograms for the estimated SE and variance are shown in Supporting Information Figure S3.

As shown in Table 2, OLS had the third worst MSE and second worst type I error rate of 28.23%. The SE from the OLS fit underestimated the true SD. The OLS histogram (Figure 2) was asymmetric, suggesting that relaxivity from OLS was poorly behaved and not amenable to the usual comparisons that are widely performed. Fits that account for the heteroscedasticity because of constant CV all performed better, with WLS, NWLS, LNLS, GGLM-INV, and GGLM-ID being nearly equivalent. However, WLS was slightly biased in the resampling. TS did not appear to be viable despite being a nonparametric method without assumptions. Overall, the resampling simulations suggest that NWLS, LNLS, and both GGLMs perform much better than OLS as assessed by MSE and type I error (bolded).

To further investigate the optimal model, we performed Monte Carlo simulations with $\sigma_{T_2} = 0.025T_2^p$ and $\sigma_{[Fe]} = 0.10\,[Fe]$. Results are shown in Table 3. The first simulation (Table 3, condition a) was done accounting for only 2.5% error in T_2 whereas the second simulation (Table 3, condition b) additionally accounted for a 10% error in concentration. The inclusion of concentration error allowed us to determine which fitting method is best in a practical sense.

TABLE 2 Resampling results (100,000 iterations)

Method	Full r_2 ($\mathrm{mM^{-1}\,s^{-1}}$)	Resample $\overline{r_2}$ ($\pm$ Bias) ($\mathrm{mM^{-1}\,s^{-1}}$)	Mean predicted SE ($\mathrm{mM^{-1}\,s^{-1}}$)	Resample SD	$P(SE^{pred} < SD)$ (%)	Type I error rate (%)	MSE ($\mathrm{mM^{-2}\,s^{-2}}$)
OLS	93.655	93.664 (+0.009)	1.809	2.922	94.25	28.16	8.54
NLS	90.447	90.260 (−0.188)	1.696	4.890	100	49.35	23.95
WLS	91.750	91.556 (−0.194)	1.776	1.856	58.55	5.47	3.48
NWLS	92.172	92.284 (+0.112)	1.768	1.819	55.44	**5.34**	**3.32**
LNLS	91.950	91.922 (−0.028)	1.771	1.834	56.66	**5.39**	3.37
GGLM-INV	91.881	91.800 (−0.081)	1.776	1.841	56.87	**5.39**	**3.39**
GGLM-ID	92.021	92.043 (+0.022)	1.766	1.829	56.56	**5.38**	**3.34**
TS	92.076	92.830 (+0.754)	2.111	2.869	84.36	18.96	8.80

NWLS, LNLS, and GGLM-ID perform best as assessed by minimum MSE. NWLS, GGLM-INV, and GGLM-ID have the lowest type I error rates closest to the ideal 5%. OLS has very high type I error rate as the predicted SE nearly always underestimates the true SD in the column $P(SE^{pred} < SD)$.

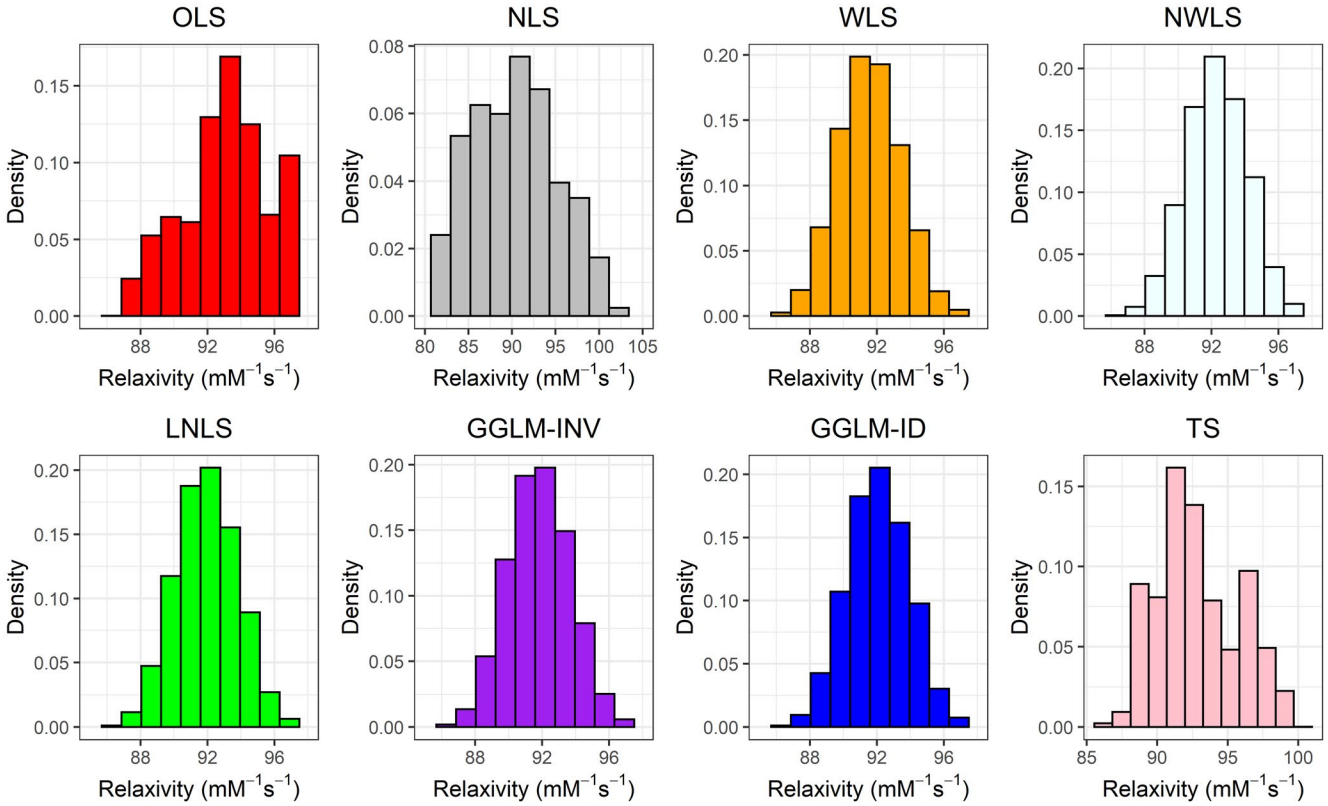

FIGURE 2 Histograms of the relaxivity resampling distributions from each method. OLS relaxivity is bimodal and asymmetric. WLS, NWLS, LNLS, GGLM-INV, and GGLM-ID relaxivities are all normally distributed. TS does not appear to be viable in practice

The Monte Carlo Simulation results in Table 3 confirmed that OLS is a suboptimal method to determine relaxivity and perform comparisons. The MSE for OLS relative to the bolded methods was roughly three times higher without concentration errors and four times higher with 10% concentration errors. Therefore, the estimate from OLS had poor precision/reproducibility. For making statistical comparisons between CAs, the type I error rate for OLS was roughly 22% and increased to 26% when concentration errors were considered. This indicates that the probability of wrongly concluding two relaxivities are statistically different is 20–30% with OLS at a nominal significance level of 5%.

Many of the proposed fits to account for heteroscedasticity were similar when no concentration error was considered. The type I errors were close to the nominal 5% level of significance. However, interestingly, concentration errors appeared to worsen the performance of WLS the most relative to NWLS, LNLS, and both GGLMs. This is in line with the resampling results from earlier. Furthermore, the Monte Carlo simulation with concentration error appeared to indicate that the identity link for GGLM performs better than the inverse link for practical biochemical situations.

To showcase an application of the GGLM-ID framework to more complex linear models in MRI relaxometry, we performed ANOVA with OLS, GGLM-ID, and LNLS to the R_1 repeatability and reproducibility study data at 7 T from Waterton et al.[20] A plot of R_1 versus concentration in each facility can be found in Supporting Information Figure S4 and clearly shows variance increasing with R_1. The −1, 0, 1 coding was used for each facility dummy variable, with facility G1 assigned −1. Results are shown in Table 4.

For the $NiCl_2$ phantom in the study, we found (for the isocenter alone) the mean relaxivity across all facilities to be 0.840 mM^{-1} s^{-1} using OLS, and 0.849 mM^{-1} s^{-1} using GGLM-ID/LNLS. ANOVA and minimum AIC both suggested that for GGLM-ID/LNLS, the model could be simplified to the standard relaxivity model without facility as a factor. Although this is not definitive proof that the MRI facility has no effect on relaxivity (not rejecting the null hypothesis does not prove the null hypothesis), the data analysis with GGLM-ID/LNLS did not support the conclusion that relaxivity differed across facilities. OLS, however, did suggest that relaxivity between facilities differed by more than chance alone. Bonferroni's post-hoc test revealed statistically significant differences in relaxivity (at familywise $\alpha = 0.05$) calculated by OLS ANOVA between facility A and each of facility C/F/G1 as well as facility E and F (Supporting Information Table S1). However, considering the increased type I error rate of OLS because of the violation of homoscedasticity, this conclusion is suspect and not necessarily replicable. To investigate this, we

TABLE 3 Results of relaxivity simulation (100,000 iterations)

Method	SD[a] ($mM^{-1}\,s^{-1}$)	Bias[a] ($mM^{-1}\,s^{-1}$)	MSE[a] ($mM^{-2}\,s^{-2}$)	*P* (type I error) (%)	SD[b] ($mM^{-1}\,s^{-1}$)	Bias[b] ($mM^{-1}\,s^{-1}$)	MSE[b] ($mM^{-2}\,s^{-2}$)	*P* (type I error) (%)
OLS	1.115	0.021	1.244	21.68	7.165	0.048	51.346	26.25
NLS	1.642	0.010	2.695	44.18	6.263	−0.495	39.473	34.97
WLS	0.639	−0.031	**0.410**	**4.35**	3.605	−1.730	15.988	6.71
NWLS	0.642	0.057	0.416	4.40	3.434	0.884	**12.572**	**5.00**
LNLS	0.640	0.013	**0.410**	**4.34**	3.450	−0.413	**12.076**	**5.01**
GGLM-INV	0.639	−0.002	**0.409**	**4.34**	3.486	−0.852	12.880	5.32
GGLM-ID	0.640	0.027	0.411	4.36	3.430	0.022	**11.768**	**4.90**
TS	0.738	−0.002	0.545	8.85	4.977	−0.010	24.768	10.13

The true relaxivity used to generate the T_2 data was $r_2^{true} = 100\ mM^{-1}\,s^{-1}$ with intercept $R_{2,0} = 0.5\ s^{-1}$. A normally distributed T_2 error where $\sigma_{T_2} = 0.0257\eta_2^b$ was added to the ground truth. The SD, bias, and MSE of the estimated relaxivity is shown along with the type I error probability for a two-tailed *t*-test against the true relaxivity. *P* also originated from a normal distribution with mean 1.18 and SD 0.12, based on the Bayesian linear fit of $\log(SD(T_2))$ versus $\log(T_2)$.

[a]No additional concentration errors, T_2 error only.

[b]With 10% concentration errors in addition to T_2 error. NWLS, LNLS, and both GGLMs have type I error rates robust to random concentration errors. GGLM-ID is slightly better under concentration errors than NWLS, LNLS, and GGLM-INV as assessed by MSE. The WLS relaxivity appears to acquire a slight bias under concentration error. OLS has very high type I error rates of 20–30%.

performed bootstrap (10,000 runs)-based ANOVA (with OLS/GGLM-ID) with resampling to generate a null distribution[50] (Supporting Information Table S2). No statistically significant differences between facilities were found in the bootstrap for either model, further suggesting that the OLS conclusions are not replicable.

Many of the coefficient values and effect sizes (Table 4 and Supporting Information Table S1) themselves were quite different from those calculated through the more accurate GGLM-ID/LNLS methods. Facility A and facility F, for example, have a relaxivity difference of 0.058 $mM^{-1}\,s^{-1}$ (7%) with OLS but only a 0.029 $mM^{-1}\,s^{-1}$ (3.4%) difference with GGLM-ID. Furthermore, the A–F relaxivity difference is marked as statistically significant ($P = 0.0002$) with OLS but not GGLM-ID ($P = 0.135$). Although the coefficients from OLS are theoretically unbiased in the long run, they are less efficient,[44] and a far greater sample size would be necessary to obtain representative effect sizes. The OLS coefficients are less precise than the GGLM-ID counterparts and may fail to be reproducible if the study were to be repeated.

Interestingly, a benefit of the GGLM-ID/LNLS frameworks is the ability to estimate the global CV for relaxation rate across all factors in the experiment directly through the summary output of the model in R. For GGLM-ID, $CV = \sqrt{\phi}$ where ϕ is the estimated dispersion parameter of the gamma distribution.[19] For LNLS, $CV = RSE$ where RSE is the residual SE of the LNLS fit. In the case of OLS, the RSE would be an estimate of the absolute error only if the assumptions were satisfied. Because LNLS uses a log-transform, the RSE estimates the absolute error in $\log(R_1)$, which back transforms to CV of R_1 on the original scale.[19,22] The estimated CV for this data set was 1.74% using GGLM-ID and 1.76% using LNLS, which roughly agree with each other as well as the isocenter-only CV of 1.56% reported by Waterton et al.[20]

So far, this work has focused on relatively simple scenarios in MRI relaxometry such as relaxivity and inter-facility variation within phantoms. To demonstrate a potential clinical application of GGLM-ID, we performed simulations based on the LIC calibration curve and CV values derived by St. Pierre et al.[7] Results are shown in Table 5.

According to Table 5, GGLM-ID quadratic fit performed the best of the models examined. The 95% confidence intervals (CIs) were much narrower, and the MRMSPPE was lower in both situations of no error in R_2 and 8% error in R_2. Although the MRMSPPE values appear high in Table 5, one should note the determination of LIC by biopsy itself is only accurate to 9–19%.[7,8] A simple OLS quadratic fit did not perform well and explains why a quadratic model has not been used. Through GGLM-ID, however, a quadratic relationship between LIC and R_2 is more reasonable. For the purposes of visualization, the quadratic OLS/GGLM-ID models were plotted to an approximation of the original data obtained via PlotDigitizer[49] in Figure 3.

TABLE 4 Linear model results for Waterton et al[20] data

Coefficient	OLS	GGLM-ID	LNLS
Intercept $R_{1,0}$ (s^{-1})	0.401 ± 0.0159***	0.381 ± 0.004***	0.381 ± 0.005***
Mean r_1 $(mM^{-1}s^{-1})$	0.840 ± 0.004***	0.849 ± 0.004***	0.849 ± 0.004***
$\Delta R_{1,0,A}$ (s^{-1})	0.056 ± 0.032	0.012 ± 0.010	0.012 ± 0.010
$\Delta R_{1,0,C}$ (s^{-1})	−0.019 ± 0.032	−0.010 ± 0.010	−0.010 ± 0.010
$\Delta R_{1,0,E}$ (s^{-1})	−0.0033 ± 0.032	0.0014 ± 0.010	0.0015 ± 0.010
$\Delta R_{1,0,F}$ (s^{-1})	−0.032 ± 0.032	−0.010 ± 0.010	−0.010 ± 0.010
$\Delta r_{1,A}$ $(mM^{-1}s^{-1})$	−0.031 ± 0.008***	−0.012 ± 0.007	−0.013 ± 0.007
$\Delta r_{1,C}$ $(mM^{-1}s^{-1})$	0.009 ± 0.008	0.006 ± 0.007	0.006 ± 0.007
$\Delta r_{1,E}$ $(mM^{-1}s^{-1})$	−0.011 ± 0.008	−0.013 ± 0.007	−0.012 ± 0.007
$\Delta r_{1,F}$ $(mM^{-1}s^{-1})$	0.027 ± 0.008**	0.017 ± 0.007*	0.017 ± 0.008*
ANOVA P-value	4.2×10^{-4}***	0.15	0.15
AIC_{Full}	−107.33	−170.09	−251.41[a]
AIC_{Null}	−90.22	−172.23	−253.42[a]

OLS suggests effect of facilities is statistically significant in the model, whereas GGLM/LNLS suggests it is not. AIC values for GGLM indicate better fit relative to OLS. Results of the 10 Bonferroni corrected pairwise comparisons between relaxivities measured at each facility can be found in Supporting Information Table S1

[a]AIC values for LNLS can only be compared to each other because of a different response variable of $\log(R_1)$.

*$P < 0.05$.

**$P < 0.01$.

***$P < 0.001$.

TABLE 5 Median MRMSPPE for Table 1 models of LIC versus R_2

Model (# of parameters)	Method	MRMSPPE[a] (%)	95% CI[a] (%)	MRMSPPE[b] (%)	95% CI[b] (%)
Quadratic (3)	OLS	24.91	(4.75, 61.57)	165.29	(94.34, 240.23)
	GGLM-ID	3.50	(3.13, 4.45)	16.15	(14.74, 18.21)
Double log linear (2)	OLS	10.57	(10.10, 11.35)	18.26	(16.68, 20.24)
	GGLM-LOG	11.39	(10.62, 12.47)	19.40	(17.52, 21.73)

Simulation parameters were generated from a lognormal distribution: $LIC_{obs} \sim \mathcal{LN}\left(\log(LIC_{true}), CV^2_{LIC}\right)$ and $R_2 \sim \mathcal{LN}(\log(R_{2,true}), (0.08)^2)$. CV_{LIC} was set to 0.19 for LIC < 4 mg/g and 0.09 for LIC > 9 mg/g. For 4 ≤ LIC ≤ 9, CV was simulated from CV ~ Unif (0.09, 0.19). The ground truth was set using Equation 10 with $R_{2,true} \in [10,275]\, s^{-1}$ spaced by 0.5 s^{-1}. The high MRMSPPE for OLS explains why a quadratic model that looks visually appealing may not have been chosen for R_2-LIC calibration. However, GGLM-ID makes a quadratic model more reasonable, and further improvements are possible in the future through an errors-in-variables GGLM-ID model

[a]Error in LIC but no errors in R_2.

[b]Including errors in both LIC and R_2.

4 | DISCUSSION

The OLS approach to the analysis of MRI relaxometry data suffers from high variability and biased SEs that can invalidate statistical inference. Although the coefficients are still unbiased, the higher MSE value with OLS implies that, on average, the OLS estimates will be further away from the true value. At a concentration error of 10%, the MSE for GGLM-ID (~13) is 4–5× lower than that of OLS (~50). This is a substantial improvement and is in fact equivalent to the accuracy one would obtain with a 4–5× increase in sample size using OLS. Therefore, GGLM-ID is more economical, because fewer runs are required to obtain representative results.

The simulations in Table 2 with 10% concentration error demonstrate that at a relaxivity of 100 mM^{-1} s^{-1}, this error propagates to a large variability of 7% in relaxivity using OLS but only a 3.5% variability using GGLM-ID. This is very promising, because it indicates greater reproducibility for relaxivity studies. Although 10% concentration error is relatively high, it ultimately shows that the GGLM-ID method is quite robust to concentration errors. Concentration errors are anticipated for biological samples or proprietary contrast agents where the metal content is typically determined via atomic absorption spectroscopy or inductively coupled plasma mass spectrometry.

Interestingly, the simulations suggested that WLS, which is an often-recommended method to address heteroscedasticity, is not ideal. The MSE remained low, but the estimate was slightly biased in the presence of concentration errors. In the case of GLMs, weighting is done through

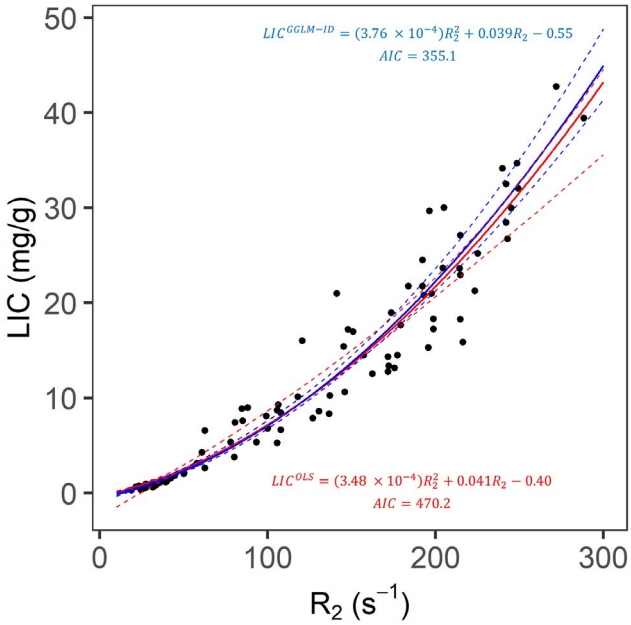

FIGURE 3 OLS and GGLM-ID quadratic fits to an approximation of R_2-LIC data from St. Pierre et al[7] PlotDigitizer were used to obtain approximations to the points. The lower AIC suggests that the GGLM-ID model fits better, which agrees with the MRMSPPE simulation in Table 5. Ninety-five percent prediction intervals depict fluctuations of the calibration curve under different realizations of the approximated data. These simulations used the same lognormal errors of $LIC_{\text{obs}} \sim \mathcal{LN}\left(\log(LIC_{\text{true}}), CV_{LIC}^2\right)$ and $R_2 \sim \mathcal{LN}(\log(R_{2,\text{true}}), (0.08)^2)$. The OLS calibration curve tends to underestimate at high R_2/LIC and has larger variability

an iterative technique that is more flexible and does not require the user to specify weights beforehand. This flexibility is likely what enabled GGLM-ID to be robust against concentration errors up to 10%. Furthermore, the simulations suggest the GGLM-ID works even when the CV is not perfectly constant in T_2 but instead in T_2^P with $P \sim N$ (1.18, 0.12^2). This suggests that GGLM-ID works under a variety of error structures.

As mentioned earlier, the log-transform is an alternative to address heteroscedasticity that has an approximately constant CV. Our simulations confirm this but suggest that GGLM-ID may be slightly superior, likely attributed, again, to the iterative weighting algorithm, which tolerates greater deviations from constant CV.[51] Some work also suggests that GGLM has more power than a log-transform to detect differences.[31] Clinically, it should be noted that substantial differences are possible if outliers are present.[32] Our simulations, which used a heteroscedastic normal distribution, show that the exact distribution of the relaxation time/rate need not follow a perfect gamma distribution. Furthermore, inverting normal T_2 manually creates non-normal R_2, and therefore GGLM-ID also appears to be robust to normality deviations. If the variance structure deviates substantially from constant CV, an option worth exploring is that of semiparametric gamma generalized estimating equations (GEE), which was used very recently to investigate the relationship between C-reactive protein and LIC by MRI.[52] GEE is even more robust to variance structure misspecification.[31,53]

Furthermore, an inherent advantage over log-transform LNLS is that the GGLM-ID does not distort linearity at the cost of stabilizing the variance (Supporting Information Figure S5). Gamma GLMs can readily be extended to existing linear models for statistical parametric mapping whereas a log-transform would change coefficient interpretation, without a nonlinear fit.[15] This is advantageous when there are continuous predictors.[15] When variables are categorical, one additional subtle benefit of using GGLM-ID over a log transform is interpretation. Log-transforms compare % change in R_i, whereas GGLM-ID compares absolute differences. This is important in MRI relaxometry, where the important metric for assessing phenomenon such as CA uptake or cerebral blood volume[16] is ΔR_1 and not $\Delta(\log R_1)$. ΔR_1 is a direct surrogate biomarker for CA concentration, whereas $\Delta(\log R_1)$ can be small at higher concentrations even if ΔC is large and vice versa. Relaxation rates are additive on the original scale,[15] and therefore the GGLM approach fits better with physical theory.

The ANOVA example using data from Waterton et al[20] shows that the choice of OLS versus GGLM-ID has the potential to influence the conclusion of an MRI relaxometry study. It should be noted that the Waterton et al[20] study used a very simple design and the concentration errors for their $NiCl_2$ phantom were very low. Although the statistically significant differences in relaxivity detected by OLS ANOVA with post-hoc Bonferroni test may not necessarily all be practically significant, they serve to highlight the potential for erroneous conclusions. This is evidenced by the bootstrap resampling (Supporting Information Table S2) that was not able to replicate the OLS conclusions. Considering that the conclusions are affected even in a simple scenario, a greater impact can be expected in experiments that contain concentration errors, biological variation, or mixed effects. The conclusions derived from ANOVA using OLS are suspect because of high type I error rates, which could make studies more difficult to replicate.[24] When evaluating imaging biomarkers for suitability, OLS cannot adequately separate true effects from both physiological or instrument noise. Furthermore, the study by Waterton et al[20] used the same RARE pulse sequence in all the facilities. As mentioned earlier, MRI parameters such as pulse sequence, flip angle, etc. also influence the measured relaxation time/rate and its CV.[2,20,23] Intuitively, comparing measurements with different acquisition methods will further increase the overall CV. Although not investigated in this study nor by Waterton et al,[20] we speculate that using the GGLM-ID framework for ANOVA has the potential to lead to greater reproducibility of inferences based on MRI relaxometry images even across different acquisition methods.

Future studies on large data sets obtained in vivo will need to be conducted to assess this.

The example of LIC-R_2 calibration serves to highlight yet another example where considering a GGLM-ID framework can potentially improve prediction. The exact nonlinear fitting method used by St. Pierre et al[7] is unknown, but GGLM-ID is much better than a simple OLS quadratic fit. Simulations in Table 4 indicate that GGLM-ID performs well as assessed by MRMSPPE even in the presence of errors on both axes. This can also be visualized in Figure 3, where the variability of the OLS calibration curve is greater than that of GGLM-ID under different realizations of the same data. For further improvements, the best approach would be to extend the GGLM framework to a heteroscedastic errors-in-variables method.[22,54,55] The validation of LIC measurements by MRI relaxometry is an ongoing field of study.[9]

The GGLM approach can also be integrated with previous guidelines for developing imaging biomarkers by the quantitative imaging biomarker alliance (QIBA).[22,56] QIBA recommends a noninferiority/superiority/equivalence testing framework (over standard hypothesis testing) to assess clinical significance.[56] Unlike OLS, GGLM can be used to construct accurate 95% CIs for such tests, while avoiding the potential problems in interpreting log-transforms that QIBA warns about.[22] GGLM can also be used to more precisely derive the constant and proportional bias of a new biomarker measurement technique relative to the accepted standard in QIBA's linear model.[22,56] For example, the presence of Rician noise is known to bias obtained T_1/T_2 values if the usual Levenberg–Marquardt fitting for signal intensity is used.[57] GGLM could be used to precisely assess this bias with respect to methods that directly incorporate the Rician noise in T_1/T_2 estimation. Although not discussed extensively in the current work, the GGLM framework can also be extended to gamma generalized linear mixed models (GGLMM) that are useful for analysis of longitudinal or spatially correlated data that often arises in imaging. These models can be readily integrated with QIBA's proposed testing framework to accurately quantify changes in R_i over time and ensure that these changes are repeatable and reproducible.

5 | CONCLUSIONS

Although widely used because of simplicity, OLS is a poor method to determine relaxivity. It fails to consider the nonconstant variance of the relaxation time/rate and provides inaccurate estimates of standard errors that invalidate statistical inference. In this work, we have shown that GGLM-ID provides one of the best estimates of relaxivity in terms of both precision and accuracy. Furthermore, among other proposed fits, GGLM factors in the inherent variability within an MRI experiment and therefore is more representative of the physical phenomenon. For in vivo studies, GGLM-ID can be used to make unbiased inferences about relaxometry biomarkers on the original scale of the data, unlike a log transform. The GGLM-ID method can be implemented easily in R with the glm() command, but to make the method more easily adoptable, we have included code for an interactive RStudio interface that can be used for relaxivity, the simplest example of the linear model in MRI relaxometry. Overall, the GGLM-ID framework serves to unify the analysis of MRI relaxometry data from simple in vitro experiments in solution to complex biological and clinical MRI data on developing quantitative imaging biomarkers.[22]

ACKNOWLEDGMENTS

This project was supported by NIH-NHLBI (grant HL137387). Biostatistics assistance was supported by the National Center for Advancing Translational Sciences, National Institutes of Health (grant UL1 TR001860). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We would also like to acknowledge Prof. Danielle Harvey in the UC Davis Department of Public Health, Biostatistics Graduate Group for aiding in the final editing of this manuscript.

DATA AVAILABILITY STATEMENT

Links to Relaxivity Dashboard and Data Analysis Code are openly available. GGLM-INV Relaxivity RShiny Dashboard: https://rokapre.shinyapps.io/GGLMINV_interface/ Github Link with Data Analysis RMarkdown PDF and RStudio Relaxivity Fitting Interface: https://github.com/Relaxivity/Relaxivity-GLM

ORCID

Rohan Kapre https://orcid.org/0000-0001-8210-6930
Junhan Zhou https://orcid.org/0000-0001-8902-7514
Laurel Beckett https://orcid.org/0000-0002-2418-9843
Angelique Y. Louie https://orcid.org/0000-0001-6610-5356

REFERENCES

1. Margaret Cheng H-L, Stikov N, Ghugre NR, Wright GA. Practical medical applications of quantitative MR relaxometry. *J Magn Reson Imaging*. 2012;36:805–824.
2. Deoni SCL. Quantitative relaxometry of the brain. *Top Magn Reson Imaging*. 2010;21:101–113.
3. Worthoff WA, Shymanskaya A, Shah NJ. Relaxometry and quantification in simultaneously acquired single and triple quantum filtered sodium MRI. *Magn Reson Med*. 2019;81:303–315.
4. Scott RC, King MD, Gadian DG, Neville BGR, Connelly A. Hippocampal abnormalities after prolonged febrile convulsion: a longitudinal MRI study. *Brain*. 2003;126:2551–2557.
5. Kawel N, Nacif M, Zavodni A, et al. T1 mapping of the myocardium: intra-individual assessment of post-contrast T1 time evolution and extracellular volume fraction at 3T for Gd-DTPA and Gd-BOPTA. *J Cardiovasc Magn Reson*. 2012;14:26.

6. Wood JC, Enriquez C, Ghugre N, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients | Blood Journal. *Blood*. 2005;106:1460–1465.
7. St. Pierre TG, Clark PR, Chua-anusorn W, et al. Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. *Blood*. 2005;105:855–861.
8. Garbowski MW, Carpenter J-P, Smith G, et al. Biopsy-based calibration of T2* magnetic resonance for estimation of liver iron concentration and comparison with R2 Ferriscan. *J Cardiovasc Magn Reson*. 2014;16:40.
9. Jhaveri KS, Kannengiesser SAR, Ward R, Kuo K, Sussman MS. Prospective evaluation of an R2* method for assessing liver iron concentration (LIC) against ferriscan: derivation of the calibration curve and characterization of the nature and source of uncertainty in the relationship. *J Magn Reson Imaging*. 2019;49:1467–1474.
10. Shen Y, Goerner FL, Snyder C, et al. T1 relaxivities of gadolinium-based magnetic resonance contrast agents in human whole blood at 1.5, 3, and 7 T. *Invest Radiol*. 2015;50:330–338.
11. Shen Y, Goerner FL, Heverhagen JT, et al. In vitro T2 relaxivities of the Gd-based contrast agents (GBCAs) in human blood at 1.5 and 3 T. *Acta Radiol*. 2018.
12. Szomolanyi P, Rohrer M, Frenzel T, et al. Comparison of the relaxivities of macrocyclic gadolinium-based contrast agents in human plasma at 1.5, 3, and 7 T, and blood at 3 T. *Investigative Radiology*. 2019;54:559–564.
13. Knobloch G, Colgan T, Wiens CN, et al. Relaxivity of ferumoxytol at 1.5 T and 3.0 T: investigative radiology. *Invest Radiol*. 2018;53:257–263.
14. Brzozowski P, Penev K, Martínez FJM, Scholl TJ, Mequanint K. Gellan gum-based gels with tunable relaxation properties for MRI phantoms. *Magn Reson Imaging*. 2019;57:40–49.
15. Callaghan MF, Helms G, Lutti A, Mohammadi S, Weiskopf N. A general linear relaxometry model of R1 using imaging data. *Magn Reson Med*. 2015;73:1309–1314.
16. Thrippleton M, Blair G, Valdes-Hernandez M, et al. MRI relaxometry for quantitative analysis of USPIO uptake in cerebral small vessel disease. *Int J Mol Sci*. 2019;20:776.
17. Kolf A-K, Konieczny M, Hesper T, et al. T2* mapping of the adult intervertebral lumbar disc: normative data and analysis of diurnal effects. *J Orthop Res*. 2019;37:1956–1962.
18. Kutner MHN, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*. 5th ed. Columbus, OH: McGraw Hill/Irwin; 2004.
19. Pekar S, Brabec M. *Modern Analysis of Biological Data. Generalized Linear Models in R*. Brno, Czech Republic: Masaryk University Press; 2016.
20. Waterton JC, Hines CDG, Hockings PD, et al. Repeatability and reproducibility of longitudinal relaxation rate in 12 small-animal MRI systems. *Magn Reson Imaging*. 2019;59:121–129.
21. Weiskopf N, Suckling J, Williams G, et al. Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. *Frontiers in Neuroscience*. 2013;7:95. https://doi.org/10.3389/fnins.2013.00095.
22. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res*. 2014;24:27–67.
23. Lankford CL, Does MD. Propagation of error from parameter constraints in quantitative MRI: example application of multiple spin echo T2 mapping. *Magn Reson Med*. 2018;79:673–682.
24. Amrhein V, Trafimow D, Greenland S. Inferential statistics as descriptive statistics: there is no replication crisis if we don't expect replication. *Am Stat*. 2019;73:262–270.
25. Andica C, Hagiwara A, Kamagata K, et al. Gray matter alterations in early and late relapsing-remitting multiple sclerosis evaluated with synthetic quantitative magnetic resonance imaging. *Sci Rep*. 2019;9:8147.
26. Siegel S. Nonparametric statistics. *Am Stat*. 1957;11:13–19.
27. Flandin G, Friston KJ. Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Hum Brain Mapp*. 2019;40:2052–2054.
28. Pirasteh A, Yuan Q, Hernando D, Reeder SB, Pedrosa I, Yokoo T. Inter-method reproducibility of biexponential R2 MR relaxometry for estimation of liver iron concentration. *Magn Reson Med*. 2018;80:2691–2701.
29. Chatterjee S, Commowick O, Afacan O, Warfield SK, Barillot C. Multi-compartment model of brain tissues from T2 relaxometry MRI using gamma distribution. Paper presented at: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 4–7 April 2018, 2018.
30. Fox J. *Applied Regression Analysis and Generalized Linear Models*. 3rd ed. Thousand Oaks, CA: SAGE Publications; 2016.
31. Myers RH, Montgomery DC, Vining GJ, Robinson TJ. *Generalized Linear Models: With Applications in Engineering and the Sciences*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2010.
32. Wiens BL. When log-normal and gamma models give different results: a case study. *Am Stat*. 1999;53:89–93.
33. Toth E, Helm L, Merbach AE. Relaxivity of MRI contrast agents. *Contrast Agents I Topics in Current Chemistry*. 2002;221:61–101.
34. Werner EJ, Datta A, Jocher CJ, Raymond KN. High-relaxivity MRI contrast agents: where coordination chemistry meets medical imaging. *Angew Chem*. 2008;47:8568–8580.
35. Laurent SE, Vander Elst L, Muller RN. Superparamagnetic iron oxide nanoparticles for MRI. In: *The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging*. Hoboken, NJ: John Wiley & Sons; 2013:427–447.
36. Wells J, Kazakova O, Posth O, et al. Standardisation of magnetic nanoparticles in liquid suspension - IOPscience. *J Phys D Appl Phys*. 2017;50.
37. Tellinghuisen J. On the role of statistical weighting in the least-squares analysis of UV-visible spectrophotometric data. *Appl Spectrosc*. 2000;54:1208–1213.
38. Gu H, Liu G, Wang J, Aubry A-F, Arnold ME. Selecting the correct weighting factors for linear and quadratic calibration curves with least-squares regression algorithm in bioanalytical LC-MS/MS assays and impacts of using incorrect weighting factors on curve stability, data quality, and assay performance. *Anal Chem*. 2014;86:8959–8966.
39. Martinez GV, Navath S, Sewda K, et al. Demonstration of a sucrose-derived contrast agent for magnetic resonance imaging of the GI tract. *Bioorg Med Chem Lett*. 2013;23:2061–2064.
40. Jarrett BR, Frendo M, Vogan J, Louie AY. Size-controlled synthesis of dextran sulfate coated iron oxide nanoparticles for magnetic resonance imaging. *Nanotechnology*. 2007;18:035603.
41. Tu C, Ng TSC, Sohi HK, et al. Receptor-targeted iron oxide nanoparticles for molecular MR imaging of inflamed atherosclerotic plaques. *Biomaterials*. 2011;32:7209–7216.
42. *R: A Language and Environment for Statistical Computing* [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2019.

43. Komsta L. mblm: Median-Based Linear Models. *R package version 0.12.1*.. 2019. https://CRAN.R-project.org/package=mblm.
44. Wilcox R. A note on the theil-sen regression estimator when the regressor is random and the error term is heteroscedastic. *Biom J*. 1998;40:261–268.
45. Quinonez C. 2003. CaltechTHESIS. PhD. Thesis, https://doi.org/10.7907/X9G0-2M43.
46. Carroll RJ. *Transformation and Weighting in Regression*. 1st edn. New York: Routledge; 1988: https://content.taylorfrancis.com/books/download?dac=C2006-0-18267-0&isbn=9780203735268&format=googlePreviewPdf.
47. Gareth J, Witten D, Hastie T, Tibishirani R. *An Introduction to Statistical Learning - with Applications in R | Gareth James | Springer*. New York, NY: Springer; 2013.
48. Russell Lenth. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.3.5.1. 2019. https://CRAN.R-project.org/package=emmeans.
49. *Plot Digitizer* [computer program]. Version 2.6.8. Boston, MA: Source Forge; 2015.
50. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Stat Med*. 2000;19:3219–3236.
51. Firth D. Multiplicative errors: log-normal or gamma? *J Roy Stat Soc: Ser B (Methodol)*. 1988;50:266–268.
52. Wurschi GW, Mentzel H-J, Herrmann K-H, et al. MRI as an alternative to serum ferritin for diagnosis of iron overload in children in the context of immune response after stem cell transplantation. *Pediatr Transplant*. e13583.
53. Ballinger GA. Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*. 2004;7:127–150.
54. Cheng C-L, Riu J. On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics*. 2006;48:511–519.
55. Abarin T, Wang L. Instrumental variable approach to covariate measurement error in generalized linear models. *Ann Inst Stat Math*. 2012;64:475–493.
56. Obuchowski NA, Reeves AP, Huang EP, et al. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res*. 2015;24:68–106.
57. Karlsen OT, Verhagen R, Bovée WMMJ. Parameter estimation from Rician-distributed data sets using a maximum likelihood estimator: application to t1 and perfusion measurements. *Magn Reson Med*. 1999;41:614–623.
58. Bürkner P-C. brms: an R package for Bayesian multilevel models using stan. *J Stat Softw*; 2017;1:2017.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

FIGURE S1 Synthesis of SDIO nanoparticles.[40,41] Dextran iron oxide (DIO) and sulfated dextran iron oxide (SDIO) nanoparticles were prepared with slight adjustments to the published protocols by our group. In brief, reduced dextran (1 equiv.) and $FeCl_3 \times 6H_2O$ (27 equiv.) were dissolved in degassed deionized ultra-filtered (DIUF) water and stirred under argon for 30 min at 0–5°C. $FeCl_2 \times 4H_2O$ (18 equiv.) and NH_4OH (432 equiv.) were added dropwise to the mixture. After 3 h at 85°C, DIO was purified by dialysis (50 KDa) against deionized water for 72 h. DIO was obtained as a brown powder after lyophilization. DIO (1 equiv.) was dissolved in dry formamide, to which 2-methyl-2-butene was added dropwise under argon. Sulfur trioxide pyridine complex (1 equiv.) was added in the mixture. Then, the mixture was refluxed for 3.5 h at 30°C with stirring and argon bubbling. Saturated $NaHCO_3$ aqueous solution was used to quench the reaction. The product was purified by dialysis (50 KDa) against deionized water for 72 h. SDIO was obtained as a brown powder after lyophilization. Iron content was determined via inductively coupled plasma mass spectrometry (ICP-MS)

FIGURE S2 T_2 variance function. The variance function was determined by a Bayesian linear regression[58] via the MCMC algorithm of log(SD T_2) versus log(mean T_2). The default (uninformative) prior was used for the intercept and an $N(1, 0.33^2)$ prior was set for the slope. The posterior distribution for slope was found to be $N(1.18, 0.12^2)$. This was used in the Monte Carlo simulations in Table 2 of the main paper that contained mean 0 normally distributed error of $\epsilon \sim N\left(0, \sigma_{T_2}^2 = \left(0.025 T_2^{1.18}\right)^2\right)$

FIGURE S3 (A) Resampling histogram of estimated variance of relaxivity for each fit. Variance follows a χ^2 distribution and is therefore asymmetric for all fits. (B) Resampling histogram for estimated standard error (SE) of relaxivity for each fit

FIGURE S4 7T R_1 data from Waterton et al.[20] R_1 data are shown with all facilities pooled together. The data are clearly heteroscedastic, which impacts statistical inference (because of inefficiency and effects on estimated SEs) even if the fits may visually look similar

FIGURE S5 Log transform (LNLS) SDIO relaxivity determination. Whereas LNLS provides a precise determination of relaxivity, it distorts the linear relationship to stabilize the variance. Although adequate, this may complicate analysis in more complex models, such as those with including mixed effects and both categorical/continuous predictors

TABLE S1 Bonferroni post-hoc contrasts (computed using *emmeans*[48]) on relaxivities using OLS and GGLM-ID. The SE for each difference estimated was roughly 0.0106 $mM^{-1}s^{-1}$ with OLS and roughly 0.0127 $mM^{-1}s^{-1}$ with GGLM-ID. OLS indicates statistically significant differences in relaxivity between A and C/F/G1 as well as E and F. The estimated effect sizes for OLS are also different. Conclusions from OLS are suspect because of heteroscedasticity. The % difference in parenthesis is calculated with respect to the mean relaxivity (0.840 $mM^{-1}s^{-1}$ for OLS and 0.849 $mM^{-1}s^{-1}$ for GGLM-ID).

95% CIs are also shown for % differences with respect to mean relaxivity. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$

TABLE S2 Bootstrap resampling[50] P-values for R_1 repeatability study. Each facility was assigned a random R_1 value (with replacement) among all observed R_1 values at a given concentration. This ensures that in the bootstrap, no statistically significant differences between facilities will exist. This bootstrap OLS and GGLM-ID were fitted to the data and a null distribution of t-statistics (or Wald Z, for GGLM-ID) was generated for both models for all comparisons above. The null distribution of this bootstrap is more robust to heteroscedasticity in R_1 (s^{-1}). The assumption is that the population relaxivity (r_1 mM^{-1} s^{-1}) standard errors among Facilities are roughly the same,[50] which is valid here as the same CA is being tested. The observed t statistic in the original data was compared to this distribution to generate a bootstrap P-value. Bonferroni correction for 10 tests was applied in the end and no statistically significant differences are apparent for either model. Therefore, the OLS conclusion is not replicable. 10K iterations were performed. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$

How to cite this article: Kapre R, Zhou J, Li X, Beckett L, Louie A. A novel gamma GLM approach to MRI relaxometry comparisons. *Magn Reson Med*. 2020;84:1592–1604. https://doi.org/10.1002/mrm.28192.