Assessment of hepatic transporter function in rats using dynamic gadoxetate-enhanced MRI: a reproducibility study

Magnetic Resonance Materials in Physics, Biology and Medicine (2024) 37:697–708
https://doi.org/10.1007/s10334-024-01192-5

RESEARCH ARTICLE

Assessment of hepatic transporter function in rats using dynamic gadoxetate-enhanced MRI: a reproducibility study

Ebony R. Gunwhy[1] · Catherine D. G. Hines[2] · Claudia Green[3] · Iina Laitinen[4,5] · Sirisha Tadimalla[6] · Paul D. Hockings[4,7] · Gunnar Schütz[3] · J. Gerry Kenna[8] · Steven Sourbron[1] · John C. Waterton[8,9]

Received: 3 January 2024 / Revised: 19 July 2024 / Accepted: 23 July 2024 / Published online: 6 August 2024

Abstract

Objective Previous studies have revealed a substantial between-centre variability in DCE-MRI biomarkers of hepatocellular function in rats. This study aims to identify the main sources of variability by comparing data measured at different centres and field strengths, at different days in the same subjects, and over the course of several months in the same centre.

Materials and methods 13 substudies were conducted across three facilities on two 4.7 T and two 7 T scanners using a 3D spoiled gradient echo acquisition. All substudies included 3–6 male Wistar-Han rats each, either scanned once with vehicle ($n=76$) or twice with either vehicle ($n=19$) or 10 mg/kg of rifampicin ($n=13$) at follow-up. Absolute values, between-centre reproducibility, within-subject repeatability, detection limits, and effect sizes were derived for hepatocellular uptake rate (K^{trans}) and biliary excretion rate (k_{bh}). Sources of variability were identified using analysis of variance and stratification by centre, field strength, and time period.

Results Data showed significant differences between substudies of 31% for K^{trans} ($p=0.013$) and 43% for k_{bh} ($p<0.001$). Within-subject differences were substantially smaller for k_{bh} (8%) but less so for K^{trans} (25%). Rifampicin-induced inhibition was safely above the detection limits, with an effect size of 75 ± 3% in K^{trans} and 67 ± 8% in k_{bh}. Most of the variability in individual data was accounted for by between-subject ($K^{trans}=23.5$%; $k_{bh}=42.5$%) and between-centre ($K^{trans}=44.9$%; $k_{bh}=50.9$%) variability, substantially more than the between-day variation ($K^{trans}=0.1$%; $k_{bh}=5.6$%). Significant differences in k_{bh} were found between field strengths at the same centre, between centres at the same field strength, and between repeat experiments over 2 months apart in the same centre.

Discussion Between-centre bias caused by factors such as hardware differences, subject preparations, and operator dependence is the main source of variability in DCE-MRI of liver function in rats, closely followed by biological between-subject differences. Future method development should focus on reducing these sources of error to minimise the sample sizes needed to detect more subtle levels of inhibition.

Keywords Chemical and drug induced liver injury · Rats · Reproducibility of results · Gadolinium ethoxybenzyl DTPA · Magnetic resonance imaging

Introduction

Drug-induced liver injury (DILI) is a serious public health problem and a major concern in drug development. The WITHDRAWN database lists 60 drugs withdrawn or discontinued for DILI, which was fatal for at least 26 of the listed 60 [1]. Impaired liver function also plays a key role in drug-drug interactions (DDI), which remain a cause for concern in drug development and are often caused by hepatic transporter inhibition or induction. A critical challenge in the management of DILI and DDI is the need for better diagnostics to identify the risk earlier in the drug development lifecycle and optimise patient selection for treatment.

Dynamic contrast-enhanced MRI (DCE MRI) with gadoxetate provides potential biomarkers of DILI or DDI occurring through hepatobiliary transporter inhibition [2]. Gadoxetate is a liver-specific contrast agent that is actively taken up into hepatocytes with biliary clearance of 50% in

Ebony R. Gunwhy and Catherine D. G. Hines contributed equally to this work.

Steven Sourbron and John C. Waterton jointly supervised this work.

Extended author information available on the last page of the article

humans [3] and 70% in rats [4]. Gadoxetate is a known substrate for multiple influx and efflux transporters which are responsible for the liver specificity and biliary clearance of this agent [5–10]. Since gadoxetate and "perpetrator" drugs are both substrates of these transporters, an opportunity exists for using gadoxetate DCE-MRI to measure liver transporter inhibition or alteration, and thereby predict the risk of DILI or DDI [2, 11, 12].

The use of gadoxetate DCE-MRI as a biomarker for DILI or DDI risk is supported by an increasing body of evidence both in disease models and in humans [9, 13–20]. In a landmark paper published in 2018, Karageorgis et al. [21] reported on the first multi-centre study using gadoxetate DCE-MRI to measure liver transporter function using the method of Ulloa et al. [2]. The study used a case–control design with different animals given either a vehicle or a potent inhibitor before gadoxetate DCE-MRI. The study showed compelling evidence that inhibition of liver uptake and excretion can be detected with confidence. The data also showed between-centre biases in absolute values of uptake and excretion rates, though the relative differences between treatment and control groups were more consistent between centres.

Future development of the assay can potentially reduce the between-centre biases further, and thereby reduce the sample sizes needed to detect a given level of drug-induced inhibition of liver function. To inform such further development, this study aimed to unravel the sources of variability and identify the main sources of the observed between-centre bias, both in the absence of drugs and in the presence of a known strong inhibitor. Rifampicin was selected due to its potential for causing DILI, as determined from previous in vitro [22] and in vivo analysis [21]. Data were collected across three centres, at two field strengths, and over time periods separated by 2–19 months. In a subset of data, repeated measurements were performed in each subject.

Materials and methods

Assay development

The assay used in this study was developed by the TRISTAN project -an international collaborative private-public partnership aiming, among others, to develop standardised assays for hepatic transporter assessments in rats and humans. The standardised assay in rats was developed in consensus after a comparison of methods in literature and preliminary acquisitions and sensitivity analyses, with the results published and presented at relevant society meetings [23–30]. Subsequently, a systematic program of technical and biological validation studies was initiated to characterise the assay in terms of its parameter uncertainty, sensitivity to effects of drugs with different potency, and ability to detect adaptations after repeat dosing. Differences with the assay used in Karageorgis et al. [21] include the use of a model-based input function [30] and allowing for different relaxivity values between extra- and intracellular spaces. The assay is specified in the supplementary material which includes sufficient detail for independent replication.

Study design

Data were collected prospectively from eight substudies, with side-line data included retrospectively from five additional substudies (see Table 1 for more details). This resulted in 13 substudies available for analysis, covering:

1. Three centres: D, E, and G (see Table 2 for hardware specifications per centre)
2. Two MRI field strengths: 4.7 T, 7 T
3. Eight-time periods between experiments (substudies): 2 months, 10 months, 12 months, 13 months, 14 months, 15 months, 16 months, 19 months
4. Two treatments: saline and rifampicin

Each of the 13 substudies included between three to eight subjects, which were either scanned once (Day 1) or twice for repeatability assessment (Day 1, Day 2). Day 1 and Day 2 repeated MRI measurements were performed two to seven days apart, to ensure adequate wash-out of contrast agent and animal recovery of the anaesthesia between scans. During the baseline scan (Day 1) of each study, all subjects were administered saline/vehicle, while either saline/vehicle or another drug of interest was administered to subjects during the follow-up scan (Day 2). The date that each substudy was conducted is listed in Table 1. For each centre, time periods refer to the interval of time between the date that each separate substudy was conducted.

Substudies 1–4 comprise a multicentre test–retest study, where saline only was administered to half of the subjects at follow-up while rifampicin was administered to the other half. Substudies 5–12 comprise baseline saline/vehicle data from pharmacologic studies using the same protocol [25, 31]. Substudy 13 was a single-centre test–retest cross-over study, where six subjects were observed following the administration of saline in a scanner at 4.7 T on one day and then in a scanner at 7 T on an alternate day. This cross-over study was conducted to determine the effect of field strength on biomarker detection within the same subjects and at the same centre.

Table 1 Summary of subjects included, sorted by centre, field strength, design, and treatment

Substudy	Centre	Field strength (T)	$N_{subjects}$ / group	Baseline (Day 1)	Follow-up (Day 2)	Study date	Study type
1	D	4.7	3	Saline	Saline	July 2018	P
2	E	7	3	Saline	Saline	Sept. 2018	P
3	G	7	3	Saline	Saline	July 2018	P
4	G	4.7	4	Saline	Saline	Sept. 2019	P
1	D	4.7	3	Saline	Rifampicin	July 2018	P
2	E	7	3	Saline	Rifampicin	Sept. 2018	P
3	G	7	3	Saline	Rifampicin	July 2018	P
4	G	4.7	4	Saline	Rifampicin	Sept. 2019	P
5	E	7	6	Vehicle	–	Sept. 2019	R
6	E	7	6	Vehicle	–	Oct. 2019	R
7	D	4.7	6	Vehicle	–	Sept. 2019	R
8	G	4.7	6	Vehicle	–	Nov. 2019	R
9	G	7	4	Vehicle	–	Oct. 2019	R
10	G	7	6	Vehicle	–	May 2019	R
11	G	7	6	Vehicle	–	Nov. 2019	R
12	G	4.7	4	Saline	–	Sept. 2019	R
13	G	7	3[a]	Saline	–	Sept. 2019	P
13	G	7	3[b]	Saline	–	Sept. 2019	P
13	G	4.7	3[a]	–	Saline	Sept. 2019	P
13	G	4.7	3[b]	–	Saline	Sept. 2019	P

Follow-up treatments on Day 2 for substudies 5–12 are beyond the scope of this paper. Substudies 1–4 have been duplicated to differentiate between those treated with saline at follow-up and those with rifampicin

[a]These subjects were imaged by the same centre using 7T at baseline and 4.7T at follow-up

[b]These subjects were imaged by the same centre using 4.7T at baseline and 7T at follow-up

P = prospective; R = retrospective

Table 2 MRI specifications per centre

Centre	Spectrometer	Gradient strength / $mT \bullet m^{-1}$ (model)	Radiofrequency transmitter / receiver volume coil (i.d./mm)	Software
D	Biospec 47/20 USR Avance IIIHD	660 (B-GA12S HP)	Quadrature 200 MHz (72)	ParaVision 6.0.1
E	Biospec 70/30 USR Avance II	440 (B-GA12S)	Single channel 300 MHz (72)	ParaVision 6.0.1
G (4.7 T)	Biospec 47/40 Avance III	200 (B-GA12S)	Quadrature 200 MHz (90)	ParaVision 6.0.1
G (7 T)	Biospec 70/30 USR Avance III	300 (B-GA12)	Quadrature 300 MHz (90)	ParaVision 6.0.1

Data acquisition

Data were acquired as described in more detail in the assay specification (Supplementary material, Sect. 1). Rats were purchased locally from Charles River Laboratories with a weight of 250 g at the time of ordering. Animals were provided enrichment, standard rat chow and water ad libitum, and were pair-housed with 12 h light/dark cycles. Following local veterinary advice, the air mixture was comprised of medical grade air, or combinations of 100% oxygen with either nitrogen or nitrous oxide. Towards the initiation of the exam, animals received saline or rifampicin at -60 min for the administration of gadoxetate, following the protocol reported in Karageorgis et al. [21].

MRI was performed using 4.7 T or 7 T Bruker (Ettlingen, Germany) scanners as shown in Table 1. DCE-MRI was acquired with a 3D Fast Low Angle Shot RF-spoiled gradient echo sequence (FLASH) with an oversampling of 6 and retrospective triggering (IntraGate, Bruker, Ettlingen, Germany) from a respiratory navigator measured in front of each FLASH pulse (slice thickness = 30 mm; FA = 2°). Imaging parameters included the following: TE/TR = 1.1/5.8 ms, 26 slices, slice thickness = 1.35 mm, FOV = 60 × 60 × 35 mm and a 64 × 64 × 26 matrix for a

resolution of 0.94 × 0.94 × 1.35 mm, coronal orientation, readout in rostral-caudal direction, 6-fold phase oversampling and a flip angle of 30 degrees (centre E) or 20 degrees (centres D and G). For DCE, heparinised saline was placed in the tail vein catheter prior to the bolus of gadoxetate (see Supplementary material, Sect. 1.2). Gadoxetate Primovist or Eovist (Bayer AG, Berlin, Germany) was diluted 1:5 in saline and administered at 0.5 mL/kg (25 µmol/kg bodyweight). A separate phantom study [27] was performed prior to this study to characterise the effect of gadoxetate on T1. Relaxivity values were obtained from the literature [26] as described in the Supplementary material of this manuscript (Sect. 1.4). Any additional MRI parameters are also detailed in Sect. 1.4 of the Supplementary material. An example of DCE-MRI data acquired in this study is shown in Fig. 1.

Data analysis

Region of interest (ROI) analysis was performed at the centre of acquisition for each rat using a version of the software PMI [32] tailored to the TRISTAN rat assay including detailed standard operating procedures (SOPs) and training of users by a central centre [26, 32]. Signal-time curves were exported from PMI to csv files and further processed by a Python script [33]. All signal-time curves have also been made available [34]. Rates for hepatic plasma clearance, K^{trans} [mL/min/mL] and biliary efflux, k_{bh} [mL/min/mL] of gadoxetate (i.e., the TRISTAN rat assay biomarker rate constants of interest) were then derived from the signal-time curves via tracer kinetic two-compartment modelling (see Supplementary material, Sect. 1.4).

Each substudy returned a single mean value for each of the two biomarkers K^{trans} and k_{bh}, along with a 95% Confidence Interval (CI, defined as 1.96 × standard error).

Reference values for K^{trans} and k_{bh} were derived as the mean baseline value overall substudies, along with their CI. Reproducibility (of a single measurement) was quantified as the 95% tolerance interval (TI, defined as 1.96 × standard deviation) over substudies and expressed as a percentage of the mean. Repeatability (of a single measurement) was quantified for each of the substudies 1–4 (Table 1, rows 1–4) as the TI of baseline and follow-up saline data, expressed as a percentage of the mean. Rifampicin effect size (of a single measurement) was calculated for each of the substudies 1–4 (Table 1, rows 5–8) as the difference between saline value and inhibited value, expressed as a percentage of the saline value. Final values for repeatability and rifampicin effect size were determined as the mean (with CI) over the 4 substudies.

The absolute detection limit (%) for inhibition of any given biomarker was defined as the smallest change producing a TI that does not overlap with the CI of the saline benchmark value. The TI was estimated by applying the reproducibility error in saline. The relative detection limit (%) for inhibition of any given biomarker was defined as the smallest change producing a CI on the between-day difference that does not contain the value 0. The CI on the difference was estimated by applying the repeatability in saline, multiplied with $\sqrt{2}$ to account for the error in both terms. This quantity is also known as the repeatability coefficient (RC), a recommended metric for repeatability [35].

An analysis of sources of variability was performed on the data from substudies 1–4 using pairwise t-tests and analysis of variance. A value of $p < 0.05$ was considered significant

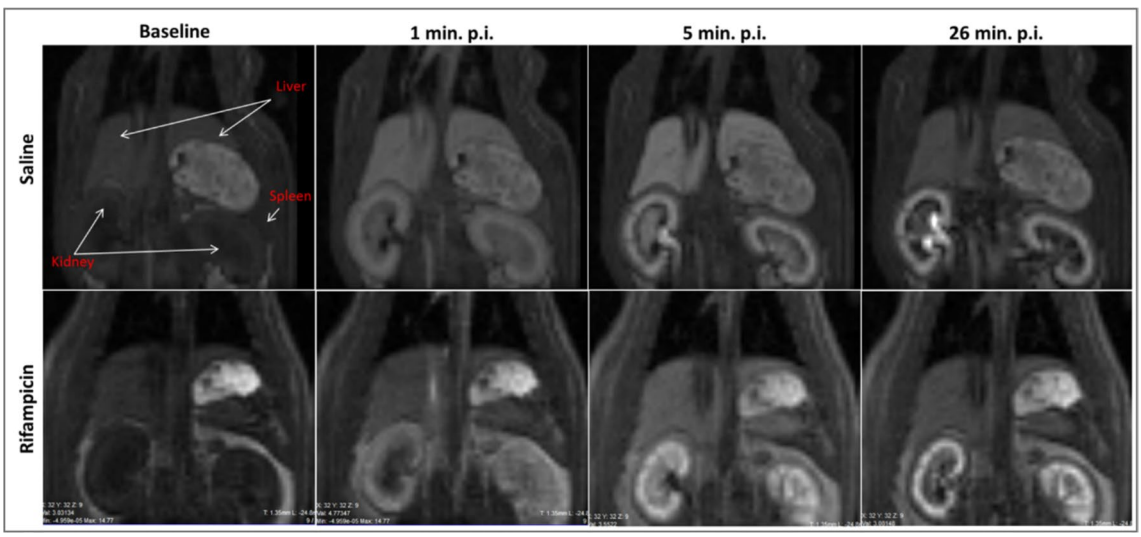

Fig. 1 Sample MRI data comparing scans at multiple time points for a single rat after administration of saline (TOP) and rifampicin (BOTTOM)

for all statistical tests. As these analyses are intended to be exploratory, no adjustment was made for multiple comparisons. The *mixed_anova* function from the Python package *Pingouin* [36] was used to partition the sources of variability into between-substudy (reproducibility), between-day (repeatability), and between-subject components for a two-way mixed-effects model with a 2×3 study design. A one-way ANOVA was conducted to identify systematic differences between substudies. To help understand the impact of different variables in isolation, the Day 1 data were stratified by centre, field strengths, and time periods and tested for significant differences using a one-way ANOVA.

Results

Exclusions

All subjects survived until the end of the procedures without adverse effects. Eight subjects were excluded from analysis due to artefacts in Day 1 data, reducing the $N_{subjects}$ to three subjects in substudy 10, three in substudy 13, and one subject in the repeatability data of substudy 2 (line 2 of Table 1). Since substudy 2 was left with only one set of valid repeatability data, it was excluded from single-substudy repeatability and rifampicin assessment. One subject was also excluded due to artefacts in Day 2 data, reducing the $N_{subjects}$ with valid repeatability data to two in substudy 3 (line 3 of Table 1).

Main results

Figure 2 summarises the baseline values for all 13 substudies, showing significant differences between substudies for K^{trans} ($p=0.013$) and k_{bh} ($p<0.001$). Average values over all substudies were $K^{trans}=0.90\pm0.08$ mL/min/mL and $k_{bh}=0.19\pm0.02$ mL/min/mL. Two substudies produced a K^{trans} measurement outside of this reference range and three other substudies produced a k_{bh} measurement outside this reference range. Figure 3 shows Day 1 vs. Day 2 results for each substudy, showing clear systematic inhibition of K^{trans} and k_{bh} after rifampicin. Day 1 and Day 2 results under the same conditions are similar.

Reproducibility was 31% for K^{trans} and 43% for k_{bh}; repeatability was 25% for K^{trans} and 8% for k_{bh}. Inhibition must be above the absolute detection limit of 30% for K^{trans} and 37% for k_{bh} to be detectable from a single data point. The relative detection limit is substantially lower for k_{bh} (11%), but not for K^{trans} (35%). Rifampicin-induced inhibition was safely above this detection limit, with an effect size of 75 ± 3% in K^{trans} and 67 ± 8% in k_{bh}. In absolute terms, rifampicin reduced K^{trans} significantly from 0.83 ± 0.15 to 0.20 ± 0.04 mL/min/mL and k_{bh} from 0.15 ± 0.08 to 0.05 ± 0.03 mL/min/mL.

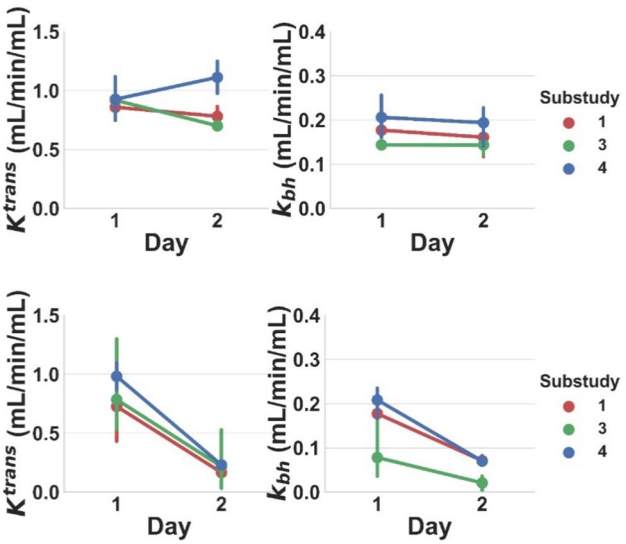

Fig. 3 Baseline and follow-up K^{trans} and k_{bh} [mL/min/mL] values for the control group treated with saline on both days (top row), and the treatment group treated with saline at Day 1 and rifampicin at Day 2 (bottom row). Results are shown for all 3 included substudies. Lines connect the mean value at Day 1 and Day 2 for each substudy. Error bars indicate the 95% confidence interval on the mean

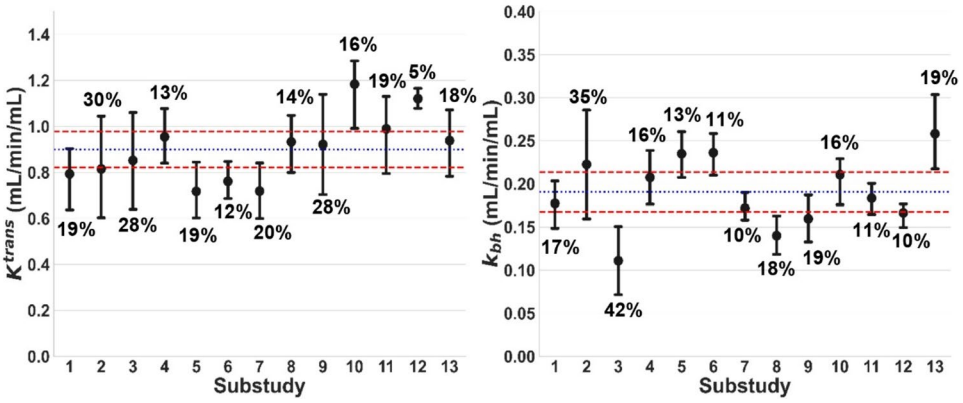

Fig. 2 Distribution of baseline K^{trans} and k_{bh} [mL/min/mL] for all substudies. Error bars show mean values for each substudy and the 95% CI on the mean. The relative error on the mean is shown as a percentage above/below each error bar. Blue lines show benchmark values for each biomarker derived as the mean across all substudies. Red lines indicate the 95% CI on the mean

Sources of variability

Most of the variability in individual data was accounted for by biological between-subject (K^{trans} = 23.5%; k_{bh} = 42.5%) and between-centre (K^{trans} = 44.9%; k_{bh} = 50.9%) variability, substantially more than the between-day variation (K^{trans} = 0.1%; k_{bh} = 5.6%). All other sources of variability combined contributed 31.5% of the total variability in K^{trans} and 1.0% in k_{bh}.

Figures 4–6 show the effect of individual variables (centre, field strength, and time point) after correcting for the others.

No significant differences were found for K^{trans}. Centres E and G, both operating at 7 T in the same time period, produced significantly different values for k_{bh} (Fig. 4).

In centre G, which operated at both 7 T and 4.7 T, significant differences in k_{bh} were found between field strengths (Fig. 5), even if measurements were performed in the same subjects. When the experiments were repeated several months apart at the same centre and field strength (Fig. 6), significant differences were observed in k_{bh} after 2 months (centre G, 4.7 T) and after 10 months (centre G, 7 T).

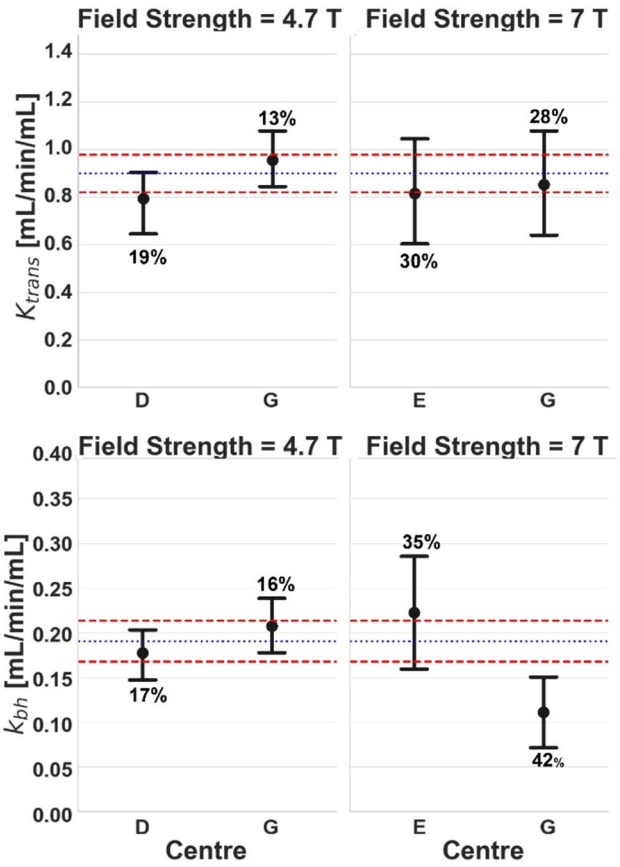

Fig. 4 Comparison of baseline K^{trans} (top row) and k_{bh} (bottom row) [mL/min/mL] for different centres but at the same field strength: 4.7 T (left column) and 7 T (right column). Error bars and colour coding are the same as Fig. 2

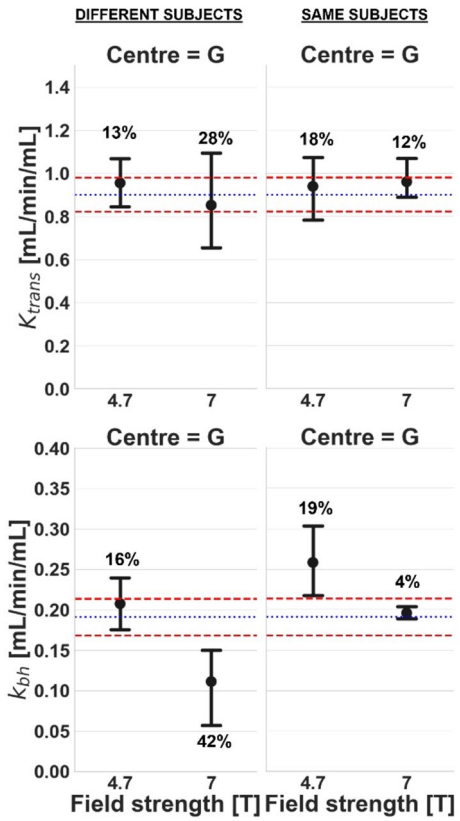

Fig. 5 Comparison of baseline K^{trans} (top row) and k_{bh} (bottom row) [mL/min/mL] for different field strengths at the same centre. Results are shown for a study with different subjects (left column) and a study with the same subjects (right column). Error bars and colour coding are the same as Fig. 2

Discussion

Reproducibility is a critical challenge for DCE-MRI, which currently hampers its use in biomarker-driven drug discovery or identification of safety risks ("de-risking"). The purpose of the current paper was to inform future method development by unravelling the sources of between-centre variability. For this purpose, 89 rat datasets were analysed from 13 separate studies at three institutions with four MRI scanners. Detailed application notes are provided in supplementary material to allow independent implementation and replication of the results.

The data demonstrated that the smallest detectable inhibition of absolute rate constants is about 2 times lower than the effect size of a potent inhibitor. The smallest detectable change in the relative rate constants values is similar for uptake but 6 times less for excretion. This suggests that

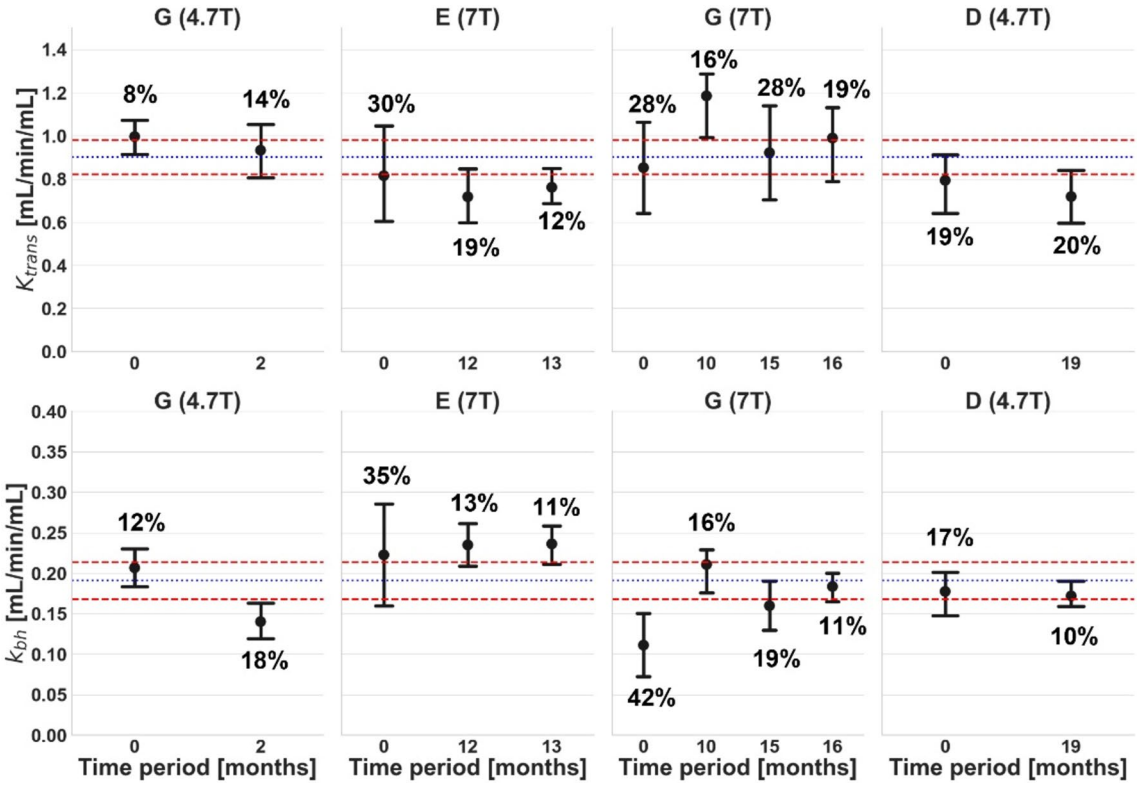

Fig. 6 Comparison of baseline K^{trans} (top row) and k_{bh} (bottom row) [mL/min/mL] at the same scanner but at different time periods. Results are shown for centre G at 4.7 T (left), centre E at 7 T (middle left), centre G at 7 T (middle right), and centre D at 4.7 T (right). Error bars and colour coding are the same as Fig. 2

the assay is sufficiently robust to detect the effect of potent inhibitors and should also be sufficient for substantially weaker inhibitors. Studies with different test compounds of various levels of potency are underway to determine exactly which level of inhibition is detectable. To avoid overinterpretation of results, the reproducibility reported in this study needs to be accounted for when comparing absolute values to the reference range derived in this study.

The detection limit on relative values for k_{bh} is about 6 times lower than that on absolute values. This shows that, in its current form, the assay is best applied by comparing a baseline assessment (saline) against an assessment with the test drug in the same population, performed on consecutive days, and reporting the relative changed caused by the drug rather than the absolute values in the presence of the test compound. The trade-off is the need for an extra control experiment in the same subjects, which at the preclinical level is a relatively small additional burden considering the gain in power of the assay.

Reference values produced in this study differ by several orders of magnitude from values reported by Karageorgis et al. [21]. However, theoretical analysis [37] has demonstrated that the latter are overestimated by a factor of 1000 due to a mistake in units, and are defined relative to a different concentration than uptake rates presented in this paper. After correcting for those differences, the values reported in Karageorgis et al. [21] are $K^{trans} = 1.0$ mL/min/mL and $k_{bh} = 0.07$ mL/min/mL. The K^{trans} value reported in [21] is very close to our reference range of 0.90 ± 0.08 mL/min/mL, however, k_{bh} is considerably outside of the reference range of 0.19 ± 0.02 mL/min/mL. Ignoring differences in definition, relative changes under rifampicin reported in Karageorgis et al. are broadly similar for K^{trans} (70% vs. 75%) but are a factor of 1.8 lower for k_{bh} (38% vs. 67%). A reanalysis of the data from Karageorgis et al. [21] using the modelling approach used in our own work would be required to determine the exact reason for these differences. However, this was beyond the current scope of the work presented in this study. Potentially this higher sensitivity to change can be traced back to a series of refinements introduced in the assay, such as the use of accurate measurements of gadoxetate relaxivities in blood and plasma [26], modelling that accounts for relaxivity differences between liver and blood, or the use of a model-based input function. A more detailed analysis of sources of variability and sensitivity of the results to these changes would be needed to unravel the different contributions of these changes and to understand whether they can explain the differences observed.

Variability analysis indicates that a reduction of biological variability due to environmental conditions should be a priority for future improvement of the assay. The analysis of variances shows that the between-subject variability in any single centre and between-centre variability are large contributing factors to the variability of the assay. This indicates that experience with biological factors from cancer models does not translate to DILI and DDI models. Indeed, Ng et al. measured K^{trans} three times in rat tumours [38] and found inter-rat variance was 28% of the total variability, while intra-rat variance comprised 72%, indicating that the repeatability error was much larger than the heterogeneity of the population.

The aim of our assay was to compare average biomarker values per centre rather than individual values per subject. However, between-subject analyses were included to help understand possible sources of error. In the individual comparisons, the between-subject error was the second-largest contributor to variability ($K^{trans} = 23.5\%$; $k_{bh} = 42.5\%$). It is therefore plausible that biological variability between populations also contributes to the between-centre variability. In terms of the technical sources of variability, the analysis indicates that two of the contributing factors investigated (centre, field strength) can have a substantial effect. This suggests that future efforts on reducing the overall technical uncertainties of the assay need to focus on all three variables. Another issue is the relatively high number of exclusions due to artefacts or fit failures. While this can be compensated in animal studies by increasing the number of subjects, this indicates an inherent instability of the assay that should be reduced in future improvements.

The between-centre variance can potentially be reduced by standardised central training of operators, more detailed standard operating procedures, and removal of variables such as choice of anaesthesia, number of subjects and subjective criteria for exclusions. In terms of field strength, while known differences such as baseline R_1 and relaxivity are corrected for, and the impact on K^{trans} appears small, there remains a significant impact on k_{bh}. Comparison to the benchmark values suggests that the 7 T results for k_{bh} rather than the 4.7 T results are an outlier, potentially indicating typical high-field effects such as increased susceptibility artefacts are responsible. Closer investigation of artefacts at different field strengths is needed to identify the source of these errors and identify practical solutions. In terms of changes over time, the data did not clearly indicate a trend or drift in values over any of the time scales. However, there was significant variability in k_{bh} values measured over time periods of 0–2 months and 0–16 months. The causes are not clear at this moment but–apart from biological factors - could potentially be explained by staff turnover or hardware/software upgrades. This indicates that these factors need to be carefully recorded and controlled in future experiments.

A barrier to the further development and deployment of gadoxetate DCE assays for liver transporter function is the wide range of methods and approaches for acquisition and analysis. This leads to sometimes orders of magnitude difference between published values, and a difficulty in identifying clear relationships between data from different studies [37]. Independent replication of the assays is effectively prohibited due to the lack of detailed standard operating procedures for acquisition and analysis, hardware specifications, sequence cards, software for analysis or implementations of modelling sections. This highlights a clear need for a well-defined, standardised, and authoritative assay that can be replicated independently, is broadly acceptable by the community and produces well-defined biomarkers [39, 40] characterised by benchmark values and their uncertainties. The assay introduced in this study will be versioned methodically to establish clear benchmarks but also account for any future improvements in acquisition or analysis methodology. The assay also served as a critical step in a translational pathway and led to the development of a compatible human assay. First validation studies of the human assay are underway and on 25th February 2021, a letter of intent on the human assay was accepted into the formal biomarker qualification program of the Food and Drug Administration (FDA) [41].

Limitations

Repeatability measurements should ideally be performed under identical conditions, which requires a 'coffee-break' type experiment with a short interval between both scans. This is not feasible in this context because sufficient time must be allowed for gadoxetate to fully clear from the body. Therefore, assessments are performed on successive days, which may cause some added variability due to recovery from anaesthesia or other physiological impacts of the previous day's measurement. Furthermore, the time between successive scans varied for each centre (ranging from two to seven days) for practical reasons due to MRI scanner availability. This may have introduced more variability in the between-subject variance observed in this study.

Reproducibility assessments are ideally performed in the same subjects so that the effect of experimental variables can be separated from population heterogeneity. In rat studies this is not feasible as the transport, changes in housing and diet, and time required for transfer between centres is likely to cause significant changes. Hence assessments at different centres were performed on different subjects, which means that the effects of population heterogeneity could not be fully distinguished.

The confidence intervals on the absolute values for K^{trans} and k_{bh} show that the current sample size is sufficient to estimate these with a high degree of confidence.

Since the analysis reveals systematic differences between centres, field strengths, and time periods, increasing the sample size further will not fundamentally affect the conclusions on sources of variability. This confirms that the study was adequately powered for the primary objective.

However, the sample size of this study is insufficient to provide a high degree of confidence on the parameter errors, such as the reproducibility or repeatability errors provided in this paper, or the detection limits on absolute and relative values. Reliable estimates of the "error on the error" would typically require substantially larger studies. For this reason, confidence intervals on the errors themselves are not provided in this paper, and their values should be seen as order-of-magnitude estimates rather than reliable benchmarks for data interpretation in future studies.

Not all data presented in this paper are prospective, as some studies were planned after the first results were analysed to fill in gaps in the data. Also, while acquisitions at different centres were performed independently by different teams, the nature of the project involved frequent discussion on video calls or in-person meetings, including exchange of knowledge on practical issues encountered during the acquisitions. Moreover, while the images were fully processed locally, the analysts were trained by a central team that had developed the modelling and provided the software. In the initial stages of the project some locally derived results were compared against a central analysis and corrected if needed. As a result, the substudies cannot be considered fully independent, which may have created some bias in the data.

The protocols were set up originally with an aim to match the resolution across all scanners and therefore minimise the risk of differences caused by different levels of partial volume, for instance. However, this resulted in the same low resolution used at 4.7 T being used at 7 T. Another possible limitation is that the sequence parameters in this study were standardised across all field strengths rather than adjusted to account for the field dependency in relaxation times and relaxivity. The flip angle is an exception because one of the centres had inadvertently used an earlier version of the protocol. It is not expected that this would cause significant bias in the results, as all FAs are set to the corresponding value per centre in the signal model used for the analysis. However, the difference may have some effect on the level of B1 effects and possibly inflow suppression. In addition, alternative approaches such as using a slightly bigger field of view rather than phase oversampling were not considered in the initial set up of the sequence. Revisiting the protocol set up in future developments could aid in overcoming some of these limitations.

Conclusion

Between-centre bias caused by factors such as hardware differences, subject preparations, and operator dependence is the main source of variability in DCE-MRI of liver function in rats, closely followed by biological between-subject differences. Future method development should focus on reducing these sources of error in order to minimise the sample sizes needed to detect more subtle levels of inhibition.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10334-024-01192-5.

Acknowledgements Thank you to Carsten Schwenke, an independent statistician from SCO:OSSIS (Minden, Germany) for their statistical guidance. Thank you to Richa Gandhi for performing a ROI analysis that was included in a previous version of this work. We would also like to thank Nico Zimmer and Ute Thoma for performing and analysing the imaging experiments at Sanofi-Aventis GmbH.

Author contributions ERG and CDGH contributed equally to this work. JCW and SS jointly supervised this work. Conceptualization, CDGH, ST, CG, IL, PDH, JCW, SS, GS and JGK; methodology, ERG, SS, CDGH, CG, IL, PDH and JCW; software, ERG, SS and ST; formal analysis, ERG and SS; investigation, ERG, CDGH, IL, CG, SS, PDH and JCW; resources, CDGH, IL, PDH, GS, SS and ERG; data curation, ERG, CDGH, IL, CG, SS and JCW; writing—original draft preparation, ERG, CDGH, SS, IL, CG and ST; writing—review and editing, SS, JCW and IL; visualization, ERG and SS; supervision, CDGH, SS and JCW; project administration, JGK, CDGH and JCW; funding acquisition, JCW and GS. All authors have read and agreed to the published version of the manuscript.

Funding The research leading to these results received funding from the Innovative Medicines Initiatives 2 Joint Undertaking under grant agreement No 116106. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This communication reflects the author's view and neither IMI nor the European Union, EFPIA, or any Associated Partners are responsible for any use that may be made of the information contained therein.

Data and code availability Data and code presented in this study are openly available in Zenodo, as follows: 1. DCE-MRI datasets and results in CSV format: https://doi.org/https://doi.org/10.5281/zenodo.7838397 (published on 17 April 2023). 2. Tracer-kinetic model source code archived from GitHub: https://zenodo.org/record/8372595 (published on 23 September 2023). 3. PMI (Platform for Research in Medical Imaging) v3.1 software: https://doi.org/https://doi.org/10.5281/zenodo.4382479 (published on 21 December 2020).

Declarations

Conflict of interest ERG and ST declare no conflict of interest. CG (at the time of work), and GS are employees of Bayer AG, a for-profit company marketing the MR contrast agent gadoxetate, trade name Primovist (Eovist in the USA). IL was an employee of and received compensation from Sanofi-Aventis Deutschland GmbH at the time the work was performed and is currently an employee of Antaros Medical, a for-profit company engaged in the provision of imaging biomarker services. SS receives research support from Bayer AG for work unrelated to this study. PDH is an employee of Antaros Medical. JGK has received compensation from Bioxydyn Ltd. (a for-profit company

engaged in the discovery and development of MR biomarkers and the provision of imaging biomarker services), provides fee for consultancy advice to various pharmaceutical companies, and undertakes paid consultancy of Cosmetics Europe, which is the European trade association for the cosmetics and personal care industry. JCW holds stock in Quantitative Imaging Ltd. and is a Director of, and has received compensation from, Bioxydyn Ltd. CDGH was an employee of Merck Sharp & Dohme LLC, a subsidiary of Merck & Co., Inc., Rahway, NJ, USA, and owned stock in Merck & Co., Inc., Rahway, NJ, USA at the time the work was performed and is currently an employee of GSK.

Ethics approval All protocols in centre E were approved by the Institutional Animal Care and Use Committee, Merck, Sharp & Dohme LLC, and were in compliance with the 8th edition Guide for the Care and Use of Laboratory Animals published by the National Research Council (National Academies Press, revised 2011). The animal experiments at centre G were approved by Regierungspräsidium Darmstadt Veterinärdezernat (State of Hessen, Germany) and conducted at the AAALACi accredited animal facility. All animal studies at centre D were approved by local authorities in Berlin, Germany, and were conducted in compliance to German national and EU animal welfare regulations.

References

1. Siramshetty VB, Nickel J, Omieczynski C, Gohlke B-O, Drwal MN, Preissner R (2016) WITHDRAWN—a resource for withdrawn and discontinued drugs. Nucleic Acids Res 44(D1):D1080–D1086
2. Ulloa JL, Stahl S, Yates J, Woodhouse N, Kenna JG, Jones HB, Waterton JC, Hockings PD (2013) Assessment of gadoxetate DCE-MRI as a biomarker of hepatobiliary transporter inhibition. NMR Biomed 26(10):1258–1270
3. Weinmann HJ, Schuhmann-Giampieri G, Schmitt-Willich H, Vogler H, Frenzel T, Gries H (1991) A new lipophilic gadolinium chelate as a tissue-specific contrast medium for MRI. Magn Reson Med 22(2):233–237
4. Hamm B, Staks T, Muhler A, Bollow M, Taupitz M, Frenzel T, Wolf KJ, Weinmann HJ, Lange L (1995) Phase I clinical evaluation of Gd-EOB-DTPA as a hepatobiliary MR contrast agent: safety, pharmacokinetics, and MR imaging. Radiology 195(3):785–792
5. Stieger B, Fattinger K, Madon J, Kullak-Ublick GA, Meier PJ (2000) Drug- and estrogen-induced cholestasis through inhibition of the hepatocellular bile salt export pump (Bsep) of rat liver. Gastroenterology 118(2):422–430
6. van Montfoort JE, Stieger B, Meijer DK, Weinmann HJ, Meier PJ, Fattinger KE (1999) Hepatic uptake of the magnetic resonance imaging contrast agent gadoxetate by the organic anion transporting polypeptide Oatp1. J Pharmacol Exp Ther 290(1):153–157
7. Weinmann HJ, Bauer H, Frenzel T, Muhler A, Ebert W (1996) Mechanism of hepatic uptake of gadoxetate disodium. Acad Radiol 3(Suppl 2):S232-234
8. Muhler A, Clement O, Saeed M, Lake JR, Stites DP, Berthezene Y, Brasch RC (1993) Gadolinium-ethoxybenzyl-DTPA, a new liver-directed magnetic resonance contrast agent absence of acute hepatotoxic, cardiovascular, or immunogenic effects. Investig Radiol 28(1):26–32
9. Saito K, Ledsam J, Sourbron S, Otaka J, Araki Y, Akata S, Tokuuye K (2013) Assessing liver function using dynamic Gd-EOB-DTPA-enhanced MRI with a standard 5-phase imaging protocol. J Magn Reson Imaging 37(5):1109–1114
10. Pastor CM, Mullhaupt B, Stieger B (2014) The role of organic anion transporters in diagnosing liver diseases by magnetic resonance imaging. Drug Metab Dispos 42(4):675–684
11. Nassif A, Jia J, Keiser M, Oswald S, Modess C, Nagel S, Weitschies W, Hosten N, Siegmund W, Kuhn JP (2012) Visualization of hepatic uptake transporter function in healthy subjects by using gadoxetic acid-enhanced MR imaging. Radiology 264(3):741–750
12. Kato N, Yokawa T, Tamura A, Heshiki A, Ebert W, Weinmann HJ (2002) Gadolinium-ethoxybenzyl-diethylenetriamine-pentaacetic acid interaction with clinical drugs in rats. Invest Radiol 37(12):680–684
13. Ryeom HK, Kim SH, Kim JY, Kim HJ, Lee JM, Chang YM, Kim YS, Kang DS (2004) Quantitative evaluation of liver function with MRI Using Gd-EOB-DTPA. Korean J Radiol 5(4):231–239
14. Nilsson H, Blomqvist L, Douglas L, Nordell A, Janczewska I, Naslund E, Jonas E (2013) Gd-EOB-DTPA-enhanced MRI for the assessment of liver function and volume in liver cirrhosis. Br J Radiol 86(1026):20120653
15. Yamada A, Hara T, Li F, Fujinaga Y, Ueda K, Kadoya M, Doi K (2011) Quantitative evaluation of liver function with use of gadoxetate disodium-enhanced MR imaging. Radiology 260(3):727–733
16. Sourbron S, Sommer WH, Reiser MF, Zech CJ (2012) Combined quantification of liver perfusion and function with dynamic gadoxetic acid-enhanced MR imaging. Radiology 263(3):874–883
17. Dahlqvist Leinhard O, Dahlstrom N, Kihlberg J, Sandstrom P, Brismar TB, Smedby O, Lundberg P (2012) Quantifying differences in hepatic uptake of the liver specific contrast agents Gd-EOB-DTPA and Gd-BOPTA: a pilot study. Eur Radiol 22(3):642–653
18. Noren B, Forsgren MF, Dahlqvist Leinhard O, Dahlstrom N, Kihlberg J, Romu T, Kechagias S, Almer S, Smedby O, Lundberg P (2013) Separation of advanced from mild hepatic fibrosis by quantification of the hepatobiliary uptake of Gd-EOB-DTPA. Eur Radiol 23(1):174–181
19. Lagadec M, Doblas S, Giraudeau C, Ronot M, Lambert SA, Fasseu M, Paradis V, Moreau R, Pastor CM, Vilgrain V, Daire JL, Van Beers BE (2015) Advanced fibrosis: correlation between pharmacokinetic parameters at dynamic gadoxetate-enhanced MR imaging and hepatocyte organic anion transporter expression in rat liver. Radiology 274(2):379–386
20. Kukuk GM, Schaefer SG, Fimmers R, Hadizadeh DR, Ezziddin S, Spengler U, Schild HH, Willinek WA (2014) Hepatobiliary magnetic resonance imaging in patients with liver disease: correlation of liver enhancement with biochemical liver function tests. Eur Radiol 24(10):2482–2490
21. Karageorgis A, Lenhard SC, Yerby B, Forsgren MF, Liachenko S, Johansson E, Pilling MA, Peterson RA, Yang X, Williams DP, Ungersma SE, Morgan RE, Brouwer KLR, Jucker BM, Hockings PD (2017) A multi-center preclinical study of gadoxetate DCE-MRI in rats as a biomarker of drug induced inhibition of liver transporter function. PLoS ONE 13(5):e0197213

22. Melillo N, Scotcher D, Kenna JG, Green C, Hines CD, Laitinen I, Hockings PD, Ogungbenro K, Gunwhy ER, Sourbron S (2023) Use of in vivo imaging and physiologically-based kinetic modelling to predict hepatic transporter mediated drug-drug interactions in rats. Pharmaceutics 15(3):896
23. Scotcher D, Melillo N, Tadimalla S, Darwich AS, Ziemian S, Ogungbenro K, Schütz G, Sourbron S, Galetin A (2021) Physiologically based pharmacokinetic modeling of transporter-mediated hepatic disposition of imaging biomarker gadoxetate in rats. Mol Pharm 18(8):2997–3009
24. Waterton JC (2021) Survey of water proton longitudinal relaxation in liver in vivo. Magn Reson Mater Phys Biol Med 34(6):779–789
25. Montelius M, Sourbron S, Melillo N, Scotcher D, Galetin A, Schuetz G, Green C, Johansson E, Waterton JC, Hockings P (2021) Acute and chronic rifampicin effect on gadoxetate uptake in rats using gadoxetate DCE-MRI. In: Proceedings of the Annual Meeting & Exhibition, International Society of Magnetic Resonance in Medicine, held virtually. Abstract nr 2764. https://cds.ismrm.org/protected/21MProceedings/PDFfiles/2764.html.
26. Ziemian S, Green C, Sourbron S, Jost G, Schutz G, Hines CDG (2021) Ex vivo gadoxetate relaxivities in rat liver tissue and blood at five magnetic field strengths from 1.41 to 7 T. NMR Biomed 34(1):e4401
27. Waterton JC, Hines CDG, Hockings PD, Laitinen I, Ziemian S, Campbell S, Gottschalk M, Green C, Haase M, Hassemer K, Juretschke HP, Koehler S, Lloyd W, Luo Y, Mahmutovic Persson I, O'Connor JPB, Olsson LE, Pindoria K, Schneider JE, Sourbron S, Steinmann D, Strobel K, Tadimalla S, Teh I, Veltien A, Zhang X, Schutz G (2019) Repeatability and reproducibility of longitudinal relaxation rate in 12 small-animal MRI systems. Magn Reson Imaging 59:121–129
28. Green C, Tadimalla S, Steinmann D, Sourbron S, Koehler S, Juretschke H-P, Laitinen I, Waterton JC, Hockings PD, Hines CD (2019) Inter-site repeatability and quantitative assessment of hepatic transporter function with DCE-MRI in rats. In: Proceedings of the 27th Annual Meeting & Exhibition, International Society of Magnetic Resonance in Medicine, Montréal, QC, Canada. Abstract nr 1773. https://cds.ismrm.org/protected/19MProceedings/PDFfiles/1773.html.
29. Kenna JG, Waterton JC, Baudy A, Galetin A, Hines CD, Hockings P, Patel M, Scotcher D, Sourbron S, Ziemian S (2018) Noninvasive preclinical and clinical imaging of liver transporter function relevant to drug-induced liver injury. Drug-Induced Liver Toxicity. Springer, pp 627–651
30. Gunwhy ER, Tadimalla S, Waterton JC, Hockings PD, Schütz G, Green C, Kenna JG, Sourbron S (2023) A novel approach to derive robust arterial input functions for DCE-MRI in small animals. In: Proceedings of the Annual Meeting & Exhibition, International Society of Magnetic Resonance in Medicine, Toronto, ON, Canada. Abstract nr 2761.
31. Kenna JG, Green C, Hines CD, Laitinen I, Galetin A, Hockings PD, Melillo N, Montelius M, Scotcher D, Sourbron S, Waterton JC, Schütz G (2021) In vivo imaging and evaluation of drug-drug interaction risk arising via hepatobiliary transporters. In: Proceedings of the 60th Annual Meeting, Society of Toxicology, held virtually. p 157. Abstract nr 2358. https://www.toxicology.org/pubs/docs/Tox/2021Tox.pdf.
32. Sourbron S (2020) plaresmedima/PMI-0.4-TRISTAN-RATS: TRISTAN RATS v3.1 (v3.1). Zenodo. 10.5281/zenodo.4382479
33. Gunwhy ER, Sourbron S (2023) TRISTAN-rat (v2.0.0). Zenodo. 10.5281/zenodo.7839203
34. Gunwhy ER, Green C, Laitinen I, Tadimalla S, Sourbron S, Hockings PD, Schütz G, Kenna JG, Waterton JC, Hines CDG (2023) Rat gadoxetate MRI signal dataset for the IMI-WP2-TRISTAN reproducibility study. Zenodo. https://doi.org/10.5281/zenodo.7838397
35. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, Wahl RL, Kurland BF, Schwarz AJ, Gönen M (2015) Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 24(1):27–67
36. Vallat R (2018) Pingouin: statistics in python. J Open Source Softw 3(31):1026
37. Tadimalla S, Green C, Steinmann D, Koehler S, Juretschke H-P, Laitinen I, Waterton JC, Hockings PD, Hines CD, Schütz G (2019) Repeatability of hepatocellular uptake and efflux in the rat liver: A comparison of Gadoxetate DCE-MRI models. In: Proceedings of the 27th Annual Meeting & Exhibition, International Society of Magnetic Resonance in Medicine, Montréal, QC, Canada. Abstract nr 5034. https://cds.ismrm.org/protected/19MProceedings/PDFfiles/5034.html.
38. Ng CS, Waterton JC, Kundra V, Brammer D, Ravoori M, Han L, Wei W, Klumpp S, Johnson VE, Jackson EF (2012) Reproducibility and comparison of DCE-MRI and DCE-CT perfusion parameters in a rat tumor model. Technol Cancer Res Treat 11(3):279–288
39. Center for Devices and Radiological Health at the Department of Health and Human Services (2022) Technical Performance Assessment of Quantitative Imaging in Radiological Device Premarket Submissions: Guidance for Industry and Food and Drug Administration Staff (FDA-2019-D-1470). Food and Drug Administration of the United States of America. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/technical-performance-assessment-quantitative-imaging-radiological-device-premarket-submissions.
40. FDA-NIH Biomarker Working Group (2016) Safety Biomarker. In: BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Co-published by Food and Drug Administration and National Institutes of Health of the United States of America, pp 31–33.
41. Center for Drug Evaluation and Research (CDER) Biomarker Qualification Program (BQP) (2021) Determination Letter in response to Letter of Intent on the TRISTAN human assay (DDT BMQ000096). Food and Drug Administration of the United States of America. https://www.fda.gov/media/149413/download.

Authors and Affiliations

Ebony R. Gunwhy[1] · Catherine D. G. Hines[2] · Claudia Green[3] · Iina Laitinen[4,5] · Sirisha Tadimalla[6] · Paul D. Hockings[4,7] · Gunnar Schütz[3] · J. Gerry Kenna[8] · Steven Sourbron[1] · John C. Waterton[8,9]

✉ Ebony R. Gunwhy
e.gunwhy@sheffield.ac.uk

[1] Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Polaris, 18 Claremont Crescent, Sheffield S10 2TA, UK

[2] GSK, Collegeville, Collegeville, P.A., USA

[3] MR & CT Contrast Media Research, Bayer AG, Berlin, Germany

[4] Antaros Medical, GoCo House, Mölndal, Sweden

[5] Sanofi-Aventis GmbH, Frankfurt, Germany

[6] Institute of Medical Physics, University of Sydney, Sydney, Australia

[7] Chalmers University of Technology, Gothenburg, Sweden

[8] Bioxydyn Ltd, St. James Tower, Manchester, UK

[9] Centre for Imaging Sciences, Division of Informatics Imaging & Data Sciences, School of Health Sciences, Faculty of Biology Medicine & Health, University of Manchester, Manchester, UK