

“Number needed to read”—How to facilitate clinical trials in MR-angiography

M. Voth · U. I. Attenberger · A. Luckscheiter ·
S. Haneder · T. Henzler · S. O. Schoenberg ·
C. Schwenke · H. J. Michaely

Received: 11 July 2010 / Revised: 30 August 2010 / Accepted: 29 September 2010 / Published online: 23 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract

Purpose To evaluate the effect of the number of readers on the statistical results in peripheral MRA.

Materials and methods 40 patients with peripheral arterial occlusive disease were included as a sample dataset in this study, randomly separated into two matched groups with $n=20$ patients (group 1—gadobutrol, group 2—gadoterate meglumine) who underwent a continuous table movement MRA of the peripheral vessels at 3 T. Image quality (IQ) of 17 vessel segments was evaluated by 5 independent readers. The effect of the number of readers on significance and statistical power was statistically analyzed.

Results Image quality in group 1 (gadobutrol) ranks significantly higher compared to group 2 (gadoterate meglumine) with a diagnostic IQ in 97% vs. 78% ($p < 0.0001$). For the diagnostic/non-diagnostic IQ assessment significance was reached with one reader 1/5 times (20%), with two readers in 4/10 (40%), with three readers in 6/10 (60%), with four readers in 4/5 (80%), with five readers in 1/1 (100%). Power considerations showed considerable gain when increasing the number of readers.

Conclusion Increasing the number of readers in a diagnostic MRA-study can be used to achieve a higher power or to

decrease the number of subjects included with maintained statistical validity.

Keywords MRA · Comparison study · Power analysis · Vessel segments · Statistical significance

Introduction

Diagnostic studies are clinical studies, designed to evaluate different diagnostic methods/approaches. Thus, they are subject to Good Clinical Practice and guidelines from health authorities. Besides, due to additionally specific guidelines [FDA [1], CPMP [2]] the efficacy of a diagnostic approach needs to be proven by a blinded read performed by multiple, blinded and independent clinical experts. In the majority of cases, statistical evaluation these days is based on two or three reader analyses. Moreover, contrary to other clinical studies the experimental unit in diagnostic studies, i.e., the unit to which the diagnostic procedure is applied, and the observational unit, i.e., the unit from which the observation is obtained, are different. For example in studies to detect focal liver lesions by an imaging technique, the patient serves as experimental unit, whereas the single liver lesions are the observational unit in this instance. Statistical methods like generalized estimation equations and the modified adjusted Chi²-test [3] allow for analyses taking into account correlations between units within a patient and multiple measurements by different readers.

Up to now, sample size and power are most often based on the assessment of a single reader whereas the analysis is done on two or more readers' assessments. To optimize study designs with respect to a minimal number of patients required and minimal overall costs, the effect of increasing the number of readers should be considered. In addition

M. Voth · U. I. Attenberger · A. Luckscheiter · S. Haneder ·
T. Henzler · S. O. Schoenberg · H. J. Michaely (✉)
Institute of Clinical Radiology and Nuclear Medicine,
University Medical Center Mannheim,
Medical Faculty Mannheim, University of Heidelberg,
Theodor-Kutzer-Ufer 1-3,
68167 Mannheim, Germany
e-mail: henrik.michaely@umm.de

C. Schwenke
SCOSSIS Statistical Consulting,
Berlin, Germany

underpowered studies expose subjects to unnecessary risks, as excessively overpowered studies [4]. In our study we focused on ordinal and binary endpoints in a parallel-group setting comparing image quality of two gadolinium-based contrast agents (GBCAs). Currently, contrast-enhanced magnetic resonance angiography (CE-MRA) using conventional extracellular GBCA is the most widely used technique in all day clinical routine for visualization and assessment of vascular disease.

The purpose of our study was therefore three-fold: First, to demonstrate the effect of different numbers of readers on the success of an image quality comparison of two macrocyclic GBCAs in peripheral MRA evaluating, whether more readers reduce the number of patients required or a fixed sample size increases the chance of success; Second, to compare the performance of two statistic methods (generalized estimation equations (GEEs) and the modified adjusted Chi²-test) for binary endpoints. Third, to compare image quality of two different macrocyclic GBCAs for peripheral MRA using a low dose regimen of 0.07 mmol/kg BW.

Material and methods

Patients

This monocentric, randomized, retrospective open-label study was approved by the institutional review board and written informed consent was obtained from all patients included in the study. Between October 2008 to November 2009 two sets of 20 age and gender-matched patients (mean age 72 years in gadobutrol group and 69 years in gadoterate meglumine group, 24 men/16 women) were sampled (Table 1) out of 172 routine patients suffering from PAOD at Fontaine stages II–IV and referred for MRA with either gadobutrol or gadoterate meglumine. The randomization of the administered contrast agent was based on the day of the treatment (randomization by treatment day). The data were stored prospectively and retrospectively analyzed.

Table 1 Demography and baseline characteristics for randomization

	Gadobutrol (<i>n</i> =20)	Gadoterate meglumine (<i>n</i> =20)	<i>p</i> -value*
Age (years)	72±11	69±10	0.2030
Male	12	12	1.0000
Female	8	8	
Weight (kg)	76±15	81±23	0.5237
Renal Function (MDRD group)			
1 (>89 ml/min/1,73 m ²)	4	5	0.3947
2 (>60 to ≤89 ml/min/1,73 m ²)	6	9	
3 (>30 to ≤60 ml/min/1,73 m ²)	7	6	
4 (>15 to ≤30 ml/min/1,73 m ²)	0	0	
5 (≤15 ml/min/1,73 m ²)	3	0	

* *p*-value for difference between treatment groups

MR-hardware

All MRA examinations were performed on a 3.0 T 32 channel whole-body MR system (MAGNETOM Tim Trio [102×32], Siemens AG, Healthcare Sector, Erlangen, Germany). For signal reception, a dedicated peripheral angiography matrix coil with 36 independent coil elements, 2 body matrix coils each with 6 independent coil elements and 2 clusters of the inbuilt spine matrix were used to cover the entire field of view (FoV) from the diaphragm to the feet. All these coils can be tightly fitted to the patients to allow for high SNR. The patients were positioned supine and feet-first. In all patients 18 G intravenous access was obtained in the left or right cubital vein. For the administration of contrast agent an automated power injector (Medrad Spectris Solaris EP, Medrad Indianola, PA) was used.

Contrast agents

For this study two macrocyclic GBCAs were used as high complex stability [5] owing to the kinetic stability characteristic. The 1.0 molar formulated gadobutrol (Gadovist[®], Bayer Schering Pharma AG, Berlin, Germany) is a hydrophilic, neutral (nonionic) contrast agent. The 0.5 molar formulated gadoterate meglumine (Dotarem[®], Guerbet, France) represents a hydrophilic, ionic contrast agent. The T1 relaxivity (r1) of gadobutrol vs. gadoterate meglumine is 5.0±0.3 vs. 3.5±0.2 lmmol⁻¹s⁻¹ (in plasma, at 3.0 T and 37°C) [6]. To allow for a sufficient comparison between the two different GBCAs the 1 M gadobutrol was diluted 1:1 with NaCl. This results in similar contrast agent bolus geometry as form equivalent to the 0.5 M gadoterate meglumine.

MR imaging

To allow for correct positioning of the MRA, 2D gradient-echo sequences localizers in coronal and transversal

orientation were acquired first. In addition, a phase-contrast vessel scout and a fast-view localizer were acquired to obtain the adjustment data required for continuous table movement (CTM), for further details see Kramer et al. [7]. A test bolus technique at the level of the renal arteries was used to calculate the patient's individual circulation time. For this purpose 1 ml was injected at 1.5 ml/s followed by a 30 ml NaCl chaser at the same injection rate. The sequence parameters of the CTM-MRA are specified in Table 2. The z-axis field-of-view reached from the abdominal aorta to the distal calves. The CTM-MRA sequence was acquired before and after the administration of the contrast agent to allow for mask subtraction. For the CTM-MRA 0.07 mmol/kg body weight of the respective GBCA was administered at 1.5 ml/s followed by a 30 ml NaCl chaser. The CTM-MRA slab was positioned to include the abdominal aorta, the pelvic vessels as well as the entire vasculature of the leg. Because of the coronal image orientation of the entire FoV in z direction an angulation of the FoV is not possible. The only parameter permitting adjustment of the spatial resolution was slice thickness. In this study, a spatial resolution of $1.2 \times 1.2 \times 1.2 \text{ mm}^3$ was realized. This value reflects the limit of the current implementation of the method that is imposed by memory constraints. During the CTM-MRA acquisition, the coil elements required to cover the FOV around the isocenter of the magnet are selected automatically. To cover a readout FOV of 38 cm, 18 elements are sufficient. Before table movement is initiated in CTM-MRA, a number of lines are acquired without moving the table. Likewise, the data acquisition is prolonged for a few seconds at the end of the imaging range after the table has stopped moving [7]. Table velocity during data acquisition is influenced by several parameters; the most important ones are the acquired spatial resolution and the applied parallel imaging (PI) factor. In our setting the table velocity was 22.3 mm/s.

Table 2 Sequence parameters for CTM-MRA

	CTM-MRA
Parallel Imaging	GRAPPA 2
Acquisition time (s)	62
Spatial resolution (mm^3)	$1.2 \times 1.2 \times 1.2$
FoV (mm)	1280×337
TR (ms)	2.43
TE (ms)	1.02
Flip angle ($^\circ$)	21
Matrix	384×312
Slices/slab	88
Bandwidth (Hz/Px)	1000
Orientation	Coronal

Image evaluation

The respective image quality of CTM-MRA was evaluated by 5 independent imaging experts with 3–10 years of expertise in consensus according to a 4-point Likert-like rating scale assessing overall image quality as previously used in other studies [8, 9]. Scores allocated were: 4 = excellent (strong enhancement of the vessels, small side-branches seen throughout the course of the vessel, no venous overlay), 3 = good (strong enhancement of the vessels, some side-branches seen, non-disturbing venous enhancement), 2 = moderate (moderate enhancement of the vessels or no side branches seen or moderate venous contamination), 1 = non diagnostic (poor opacification of the vessels or disturbing venous signal).

Image quality for CMT MRA was scored for 17 vessel segments per patient. The pre-defined segments evaluated are shown in Table 3. As additional endpoint, the assessments were dichotomized to provide information whether the image was diagnostic (image quality scores 3 and 4) or non-diagnostic (image quality score 1 and 2).

Statistical analysis

Differences for baseline characteristics between groups were evaluated by Wilcoxon rank sum tests for continuous data and Fisher's exact tests for categorical data. The image quality was assessed across all five readers and all 17 segments on a binary scale (diagnostic/non diagnostic) as overall analysis by a modified adjusted Chi²-test.

Then three segments with small differences in image quality between the two contrast agents were chosen to analyse the image quality for each reader separately, for all five readers and for all combinations of 2, 3, and 4 readers to evaluate the effect of more readers' assessments on the significance of the differences between the contrast agents. These were the Right common iliac artery (AIC right), Right deep femoral artery (AFP right), and Right posterior tibial artery (ATP right) marked bold in Table 3.

For the ordinally scaled image quality (4 point scale), multinomial regression analysis was used based on generalized estimation equations (GEEs) taking into account multiple observations per patients (several segments within a patient) and repeated measurements by up to five readers of the same observational unit. GEEs are the standard method to analyse data with multiple measures in a single patient as in our study. It takes into account the correlation between observations within the same patient and therefore provides most appropriate summary statistics. Independence was used as working correlation matrix and a cumulative logit function as link function. The hypotheses were defined as follows: H0: $\text{Distribution}_{\text{groupA}} = \text{Distribution}_{\text{groupB}}$ vs. H1: $\text{Distribution}_{\text{groupA}} \neq \text{Distribution}_{\text{groupB}}$.

Table 3 Vessel segments and results (Median values across patients by reader for all vessel segments given)

No	Vessel segment	MRA Gadobutrol Reader					MRA Gadoterate meglumine Reader				
		1	2	3	4	5	1	2	3	4	5
1	Infrarenal aorta	4	4	4	4	4	3	3	3	4	4
2	Left common iliac artery	4	4	4	4	4	3	3	3	3	3
3	Right common iliac artery	4	4	4	4	4	3	3	3	3	3.5
4	Left superficial femoral artery	4	4	4	4	4	3	3	3	2.5	3
5	Right superficial femoral artery	4	4	4	4	4	3	3	3	3	3
6	Left deep femoral artery	3	3	3	4	4	3	3	3	3	3
7	Right deep femoral artery	4	3.5	3	4	4	3	3	3	3	3
8	Left popliteal artery	4	4	4	4	4	3	3	3	3	3
9	Right popliteal artery	4	4	4	4	4	3	3	3	3	3
10	Left tibiofibular trunc	3	3	3	4	4	3	3	3	3	3
11	Right tibiofibular trunc	4	4	4	4	4	3	3	3	3	3
12	Left anterior tibial artery	3	3	3.5	4	4	3	3	3	3	3
13	Right anterior tibial artery	4	4	4	4	4	3	3	3	3	3
14	Left posterior tibial artery	3	3	3	4	4	3	3	2	3	3
15	Right posterior tibial artery	3	3	3	4	4	3	3	3	3	3
16	Left peroneal artery	3	3	3	4	4	3	3	3	3	3
17	Right peroneal artery	3	3	3	4	4	3	3	3	3	3

The dichotomized image quality (diagnostic yes/no) was analyzed using logistic regression analysis based on GEEs as well as the modified adjusted Chi-square approach [3] also taking into account the correlations between observational units within a patient and multiple assessments per unit. Again in the GEEs independence was used as working correlation matrix. The hypotheses were defined as follows. $H_0: P_{\text{groupA}} = P_{\text{groupB}}$ vs. $H_1: P_{\text{groupA}} \neq P_{\text{groupB}}$.

Basis for the sample size considerations was the dichotomized endpoint in a parallel group design. The power considerations were done for a fixed sample size of 40 patients with 20 per group. In addition, the sample size was calculated to reach a power of about 83%, i.e. more than 80%. Basic assumptions for the differences between the groups were gained from the across reader analysis of three vessel segments by all five readers with proportions of 96% vs. 86% of segments with diagnostic image quality. The ratio between groups was set to 1:1. The power considerations were done by simulation studies of 1000 runs each, simulating studies based on the results of the average across the five readers' assessments for between one and four observational units per patient. The more conservative approach (GEE or modified adjusted Chi²-test) was used for the simulation. Two-sided p -values < 0.05 were regarded as statistically significant. A flow chart of the analyses can be found in Fig. 1.

Statistical calculations were done with software SAS Version 9.2 (SAS Institute Inc., Cary, NC, USA).

Results

Image quality of two different GBCA

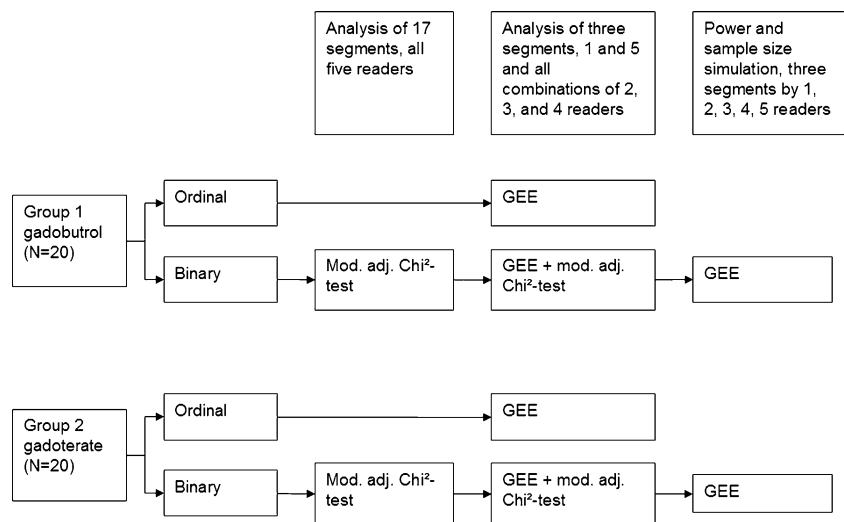
All contrast agent administrations were performed without complications. No adverse events were observed.

For image quality analysis 680 judgments ($[20 \times 17] \times 2$) were made in total by each reader; 310 vessel segments were assessed in group 1 (gadobutrol), 281 vessel segments were assessed in group 2 (gadoterate meglumine). In group 1, 30 vessel segments could not be assessed whereas in group 2 59 vessel segments were not assessable. For all readers the overall median value was 4 for gadobutrol whereas gadoterate meglumine revealed a lower overall median value of 3 (Table 3). Figure 2 illustrates an example for image quality achieved with gadoterate meglumine and gadobutrol, respectively. With all 17 segments and five readers, the proportions of segments with diagnostic image quality across readers was found to be 0.97 (95% CI=(0.94; 0.99)) for gadobutrol and 0.78 (95% CI=(0.70; 0.86)) for gadoterate meglumine leading to a difference of 0.19 (95% CI=(0.10; 0.27)). This difference was already highly significant ($p < 0.0001$). Also each reader's assessment showed a highly significant difference when evaluated separately.

Effect of different numbers of readers

To evaluate the effect of different numbers of readers on the significance and power, we restricted the analysis to

Fig. 1 Flow chart of statistical analyses performed



three segments, where the difference between the contrast agents was less dominant. The statistical significances for the ordinal (4 point scale) and dichotomized image quality (diagnostic yes/no) on the three segments per patient are summarized in Tables 4 and 5. For the ordinal endpoint, again the differences were significant in all scenarios, even with any single reader assessment. For the binary endpoint, with one reader assessment, only in 1 of 5 cases (20%), significance was reached, with two readers in 4 of 10 cases (40%), with three readers in 6 of 10 cases (60%), with four readers in 4 of 5 cases (80%) and with all five readers. The GEE approach for binary data was found to be slightly more conservative compared to the adjusted modified Chi²-test as the *p*-values in the GEE analysis were found to be slightly higher in all scenarios analyzed. Therefore the GEE approach was used for the simulation study on power and sample size.

Power and sample size

In Figs. 3 and 4 the power and sample size considerations are summarized for two, three, and four observational units per patient. For a sample size of 40 and 3 units per patient, the power starts at 29% for a single reader and increased to 79% when five readers would be included into the study, see Fig. 4. The required sample size for a power of about 83% consequently decreased from 120 to 44 for one and five readers, respectively (Fig. 3). Overall, the required number of patients needed for the analysis can be reduced by increasing the number of observational units per patient where possible (e.g. eight liver segments instead of two liver lobes per patient) and/or by increasing the number of readers assessing all images. With an intra-individual comparison design, where the contrast agents would be applied in a paired fashion, the power is already close to 80% with one reader and 40 patients and reaches 96% with

Fig. 2 A Full-thickness coronal MIP of the CTM MRA of a patient illustrates the image quality of gadoterate meglumine (a) vs gadobutrol (b). Especially the distal calf vessels could be depicted more clearly using gadobutrol, which was also reflected by the statistically significant higher median values for gadobutrol for all readers

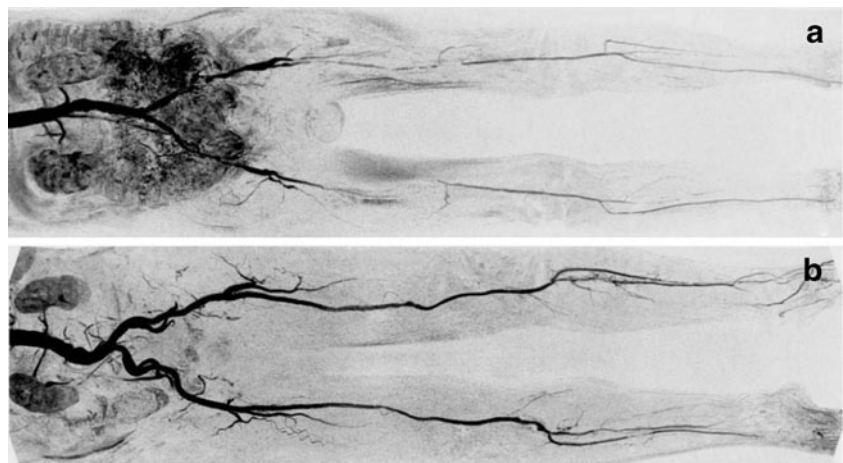


Table 4 Proportions of patients with diagnostic image quality along with differences between contrast agents and 95% confidence intervals (modified adjusted Chi2-test, $N=2*20$, 3 segments per patient)

Reader included	Gadobutrol (95% CI)	Gadoterate meglumine (95% CI)	Difference in proportions (95% CI)
1	0.94 (0.88; 1.00)	0.83 (0.70; 0.97)	0.11 (-.04; 0.26)
2	0.96 (0.91; 1.00)	0.85 (0.73; 0.97)	0.11 (-.02; 0.24)
3	0.96 (0.91; 1.00)	0.76 (0.62; 0.90)	0.20 (0.05; 0.35)
4	0.96 (0.91; 1.00)	0.92 (0.84; 0.99)	0.04 (-.05; 0.14)
5	0.96 (0.91; 1.00)	0.92 (0.84; 0.99)	0.04 (-.05; 0.14)
1 2	0.95 (0.90; 1.00)	0.84 (0.72; 0.97)	0.11 (-.03; 0.25)
1 3	0.95 (0.90; 1.00)	0.80 (0.67; 0.93)	0.16 (0.01; 0.30)
1 4	0.95 (0.91; 0.99)	0.88 (0.80; 0.95)	0.08 (-.01; 0.16)
1 5	0.95 (0.91; 0.99)	0.88 (0.80; 0.95)	0.08 (-.01; 0.16)
2 3	0.96 (0.92; 1.00)	0.81 (0.68; 0.93)	0.16 (0.03; 0.29)
2 4	0.96 (0.93; 1.00)	0.89 (0.81; 0.96)	0.08 (-.00; 0.16)
2 5	0.96 (0.93; 1.00)	0.89 (0.81; 0.96)	0.08 (-.00; 0.16)
3 4	0.96 (0.93; 1.00)	0.84 (0.76; 0.92)	0.12 (0.04; 0.21)
3 5	0.96 (0.93; 1.00)	0.84 (0.76; 0.92)	0.12 (0.04; 0.21)
4 5	0.96 (0.91; 1.00)	0.92 (0.84; 0.99)	0.04 (-.05; 0.14)
1 2 3	0.95 (0.91; 1.00)	0.81 (0.69; 0.94)	0.14 (0.01; 0.27)
1 2 4	0.96 (0.92; 0.99)	0.87 (0.78; 0.96)	0.09 (-.01; 0.18)
1 2 5	0.96 (0.92; 0.99)	0.87 (0.78; 0.96)	0.09 (-.01; 0.18)
1 3 4	0.95 (0.92; 0.99)	0.84 (0.74; 0.93)	0.12 (0.02; 0.22)
1 3 5	0.96 (0.92; 0.99)	0.84 (0.74; 0.93)	0.12 (0.02; 0.22)
1 4 5	0.96 (0.92; 0.99)	0.89 (0.82; 0.96)	0.07 (-.01; 0.14)
2 3 4	0.96 (0.93; 0.99)	0.84 (0.76; 0.93)	0.12 (0.03; 0.21)
2 3 5	0.96 (0.93; 0.99)	0.84 (0.76; 0.93)	0.12 (0.03; 0.21)
2 4 5	0.96 (0.92; 1.00)	0.90 (0.83; 0.96)	0.07 (-.01; 0.14)
3 4 5	0.96 (0.92; 1.00)	0.87 (0.80; 0.93)	0.10 (0.02; 0.17)
1 2 3 4	0.96 (0.92; 0.99)	0.84 (0.74; 0.94)	0.12 (0.01; 0.22)
1 2 3 5	0.96 (0.92; 1.00)	0.84 (0.75; 0.94)	0.12 (0.01; 0.22)
1 2 4 5	0.96 (0.92; 0.99)	0.88 (0.81; 0.95)	0.08 (-.01; 0.16)
1 3 4 5	0.96 (0.92; 0.99)	0.86 (0.78; 0.93)	0.10 (0.02; 0.18)
2 3 4 5	0.96 (0.93; 0.99)	0.86 (0.79; 0.93)	0.10 (0.02; 0.18)
1 2 3 4 5	0.96 (0.92; 0.99)	0.86 (0.78; 0.94)	0.10 (0.01; 0.19)

2 readers. In this example, the required sample size is also much lower than in an inter-individual comparison study leading to sample sizes below 20 when including more than 2 readers.

Discussion

Several publications exist on sample size considerations for multi-reader, multi-case receiver operating characteristic (ROC) studies and how to choose/reduce the number of readers [10, 11]. To our knowledge, our study is the first to demonstrate the effect of different numbers of independent and blinded readers on the statistical significance and power of image quality evaluation based on the comparison of two different GBCAs for peripheral MRA. Since image

quality of the two contrast agents was so different, we restricted the analysis to three vessel segments with minor differences to evaluate significance and power of the study design. Another reason was that GEEs should only be used with cluster sizes of four or less.

Image quality evaluation is often based on the analyses of two blinded readers. However, a higher number of readers could reduce the sample size or, at a fixed sample size, increase the power in a parallel-group setting, as used in our study. When designing a specific study, costs, access to patients and qualified readers need to be taken into account to optimize the study design in terms of sample size and number of readers. In this context it also needs to be considered which of both, the number of patients recruited, or the number of experienced readers available, are easier to accomplish for the specific study design. The

Table 5 *P*-values of the blinded reading for the binary and ordinal image quality comparisons of contrast agents ($N=2*20$, 3 segments per patient)

Reader included	GEE Ordinal*	GEE Binary*	Mod. adj. Chi ² -test Binary*
1	0.0017	0.1472	0.1344
2	0.0011	0.1065	0.1019
3	0.0005	0.0131	0.0129
4	<.0001	0.3523	0.3334
5	<.0001	0.3417	0.3245
1 2	0.0013	0.1207	0.1189
1 3	0.0004	0.0371	0.0301
1 4	<.0001	0.0896	0.0840
1 5	<.0001	0.0879	0.0821
2 3	0.0004	0.0259	0.0198
2 4	<.0001	0.0719	0.0666
2 5	<.0001	0.0704	0.0649
3 4	<.0001	0.0087	0.0047
3 5	<.0001	0.0085	0.0045
4 5	<.0001	0.3469	0.3548
1 2 3	0.0005	0.0469	0.0405
1 2 4	<.0001	0.0831	0.0772
1 2 5	<.0001	0.0821	0.0762
1 3 4	<.0001	0.0243	0.0176
1 3 5	<.0001	0.0239	0.0173
1 4 5	<.0001	0.1033	0.0949
2 3 4	<.0001	0.0173	0.0117
2 3 5	<.0001	0.0171	0.0115
2 4 5	<.0001	0.0952	0.0870
3 4 5	<.0001	0.0181	0.0120
1 2 3 4	<.0001	0.0340	0.0270
1 2 3 5	<.0001	0.0335	0.0267
1 2 4 5	<.0001	0.0754	0.0683
1 3 4 5	<.0001	0.0242	0.0170
2 3 4 5	<.0001	0.0191	0.0127
1 2 3 4 5	<.0001	0.0302	0.0228

* significant *p*-values (< 0.05) are shown in bold font, ordinal = 4-point scale, binary = dichotomous scale

sample size and power considerations of our parallel-group study demonstrated, that first, at a constant sample size more readers increase the chance of a positive study outcome and second, an increased number of readers allows to reduce the sample size without diminishing the power of a study. Thus, the highest gain in statistical power could be derived by increasing the number of readers from one to two as well as from two to three. For more than three readers, the gain becomes less significant. The power considerations, performed on 3 vessel segments per patient, demonstrate intra-individual study designs, if feasible, being advantageous over parallel-group designs.

Diagram of readers vs. sample size to reach a power of 83%

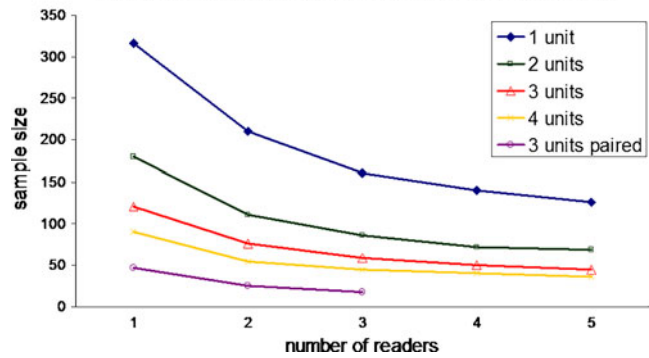


Fig. 3 Sample size for a power of ~83% for different numbers of readers and different numbers of units per patient in a parallel group design, “3 units paired” = intra-individual comparison study

However, there are some limitations to the statistical approaches used. If the sample size is below $n=20$, statistical methods like GEEs or the adjusted modified Chi-square test may not be appropriate as the approximation to the Chi-square distribution is not given any more. For cluster sizes of more than four, i.e. when more than four observational units are observed per patient, one should avoid GEEs and use the adjusted modified Chi-square test or mixed effects models for data evaluation instead. The current study was evaluated by five readers only, so that extrapolations of these results to more than five readers are speculative. However, the results show that already increasing the number readers from 4 to 5 is less effective than increasing the number of readers from 3 to 4. Nevertheless, when planning a new study, the possible effect of additional readers on the power as well as the costs of the study needs to be considered. The questions to be answered are, whether the gain in power, i.e. the reduced costs due to a reduced sample size, justifies increased costs due to the higher number of readers, and, whether it is easier to recruit more patients or more readers. It should also be kept in

Diagram of readers vs. diagnostic power

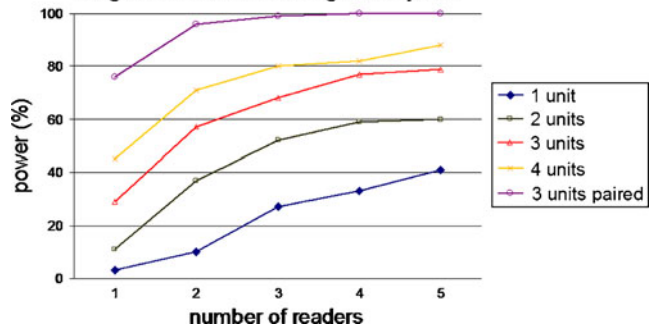


Fig. 4 Power for a sample size of 40 patients overall (2*20) for different numbers of readers and different numbers of units per patient in a parallel group design, “3 units paired” = intra-individual comparison study

mind that the effect of a greatly differing image quality analysis by one single reader would better be balanced by a large than a low number of readers.

Peripheral MRA is an excellent, non-invasive imaging method that routinely guides clinical decisions and has been widely used as the first-line diagnostic tool in arterial vessel imaging [12–14]. A number of studies have already demonstrated the value of contrast-enhanced, high spatial resolution MRA for the peripheral vasculature [15, 16]. The success of peripheral MRA is thereby based on recent technical developments like higher field strengths, multi-RF receiver channels and dedicated receiver coils as the 36-element coil that allows for a more effective implementation of parallel imaging [17, 18] without a significant loss in SNR [19, 20]. The beneficial SNR gains of 3.0 T scanners can thereby be translated in a higher temporal and/or spatial resolution. Besides, the field strength reduces the T1-shortening of gadolinium chelates while increasing the T1-relaxation of the protons of the stationary background tissue [21, 22] and thus improving the detectability of even small vessels of the vessel periphery. In this study, we used a dose level 0.07 mmol/kgBW for the entire run-off vasculature using a low dose regimen, which was evaluated for gadobutrol before in a separate study [9], the overall diagnostic image quality was significantly higher for gadobutrol than for gadoterate meglumine, which may be mainly related to the 40% higher relaxivity at 3.0 T compared to gadoterate meglumine [6]. The 1:1 dilution of the 1.0 M gadolinium chelate, gadobutrol, with saline, resulting in a contrast agent bolus geometry similar to the 0.5 M gadolinium chelate, gadoterate meglumine, might have particularly disadvantaged gadobutrol compared to gadoterate meglumine by reducing the higher T1-shortening, a combination effect of relaxivity and concentration.

The superior image quality of gadobutrol has already been proven in many instances [20, 23, 24]. Goyen et al. demonstrated superior image quality of gadobutrol vs 0.5 M Gd-DTPA, both injected at the same dose level, for whole body-MRA in healthy volunteers [23]. Mean signal-to-noise ratio and contrast-to-noise ratio values were significantly higher using gadobutrol. In an intraindividual comparative study at abdominal contrast-enhanced 3D MR angiography, depiction of small abdominal vessels was significantly better and vessel-to-tissue contrast significantly higher with 1.0 M gadobutrol than with an equimolar dose of 0.5 M Gd-DTPA [20].

Our comparison of two different macrocyclic GBCA, with the lowest propensity to release gadolinium [5], in clinically evaluated low-dose protocols at 3.0 T, favors the high T1-relaxivity [6] GBCA, gadobutrol over gadoterate meglumine. A gadobutrol enhanced low-dose protocol at 3.0 T can be considered to be an appropriate strategy especially in risk patients (e.g. severe renal impairment) [9].

The median image quality was thoroughly assessed as good or excellent by all readers, except for one segment in the gadoterate meglumine group, assessed as moderate by two readers. In the calf station image quality of many vessels segments were assessed as non-excellent, mainly due to venous enhancement and delayed flow due to filling over collateral vessels. Altered hemodynamics and venous overlay are a common problem in diagnostic MRA, which can be solved by time-resolved MRA, which was not part of the current study, but could deliver purely arterial images without venous overlay and therefore increase diagnostic accuracy for the calf station as already proven in a previous study [9].

One major limitation of this study is that no independent standard of truth was available. Therefore, the sensitivity and specificity of the gadobutrol-enhanced or the gadoterate meglumine-enhanced MRA could not be established. The results of the analysis of significance and power might be influenced by our specific clinical study used for the data generation, but mainly on the size of the effect and not the direction.

Conclusion

This study demonstrates three major facts: First, peripheral MRA with gadobutrol allows for a higher diagnostic confidence than peripheral MRA with gadoterate meglumine. Second, based on these data it could be shown that increasing the number of readers can be equally effective as including a higher number of patients for the evaluation of efficacy in diagnostic contrast agent studies. Third, the statistical approaches GEE and modified adjusted Chi²-test lead to similar *p*-values when analyzing binary endpoints.

Acknowledgements M. Voth and U. Attenberger contributed equally to this manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Food and Drug Administration (FDA) (2004) Developing medical imaging drugs and biological products. Part 3: Design, Analysis, and interpretation of clinical studies. Available via Food and Drug Administration, <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071604.pdf>, Accessed 30 Aug 2010
2. CPMP (2001) Committee for Proprietary Medicinal Products (CPMP) of the European Agency for the Evaluation of Medicinal Products (EMA). Points to consider on the evaluation of diagnostic agents

3. Schwenke C, Busse R (2007) Analysis of differences in proportions from clustered data with multiple measurements in diagnostic studies. *Methods Inf Med* 46:548–552
4. Eng J (2003) Sample size estimation: how many individuals should be studied? *Radiology* 227:309–313
5. Frenzel T, Lengsfeld P, Schirmer H, Hutter J, Weinmann HJ (2008) Stability of gadolinium-based magnetic resonance imaging contrast agents in human serum at 37 degrees C. *Invest Radiol* 43:817–828
6. Rohrer M, Bauer H, Mintonovitch J, Requardt M, Weinmann HJ (2005) Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest Radiol* 40:715–724
7. Kramer H, Zenge M, Schmitt P, Glaser C, Reiser MF, Herrmann KA (2008) Peripheral magnetic resonance angiography (MRA) with continuous table movement at 3.0 T: initial experience compared with step-by-step MRA. *Invest Radiol* 43:627–634
8. Herborn C, Watkins D, Runge V, Gendron J, Montgomery M, Naul L (2006) Renal arteries: comparison of steady-state free precession MR angiography and contrast-enhanced MR angiography. *Radiology* 239:263–268
9. Voth M, Haneder S, Huck K, Gutfleisch A, Schonberg SO, Michaely HJ (2009) Peripheral magnetic resonance angiography with continuous table movement in combination with high spatial and temporal resolution time-resolved MRA With a total single dose (0.1 mmol/kg) of gadobutrol at 3.0 T. *Invest Radiol* 44(9):627–633
10. Chakraborty DP (2010) Prediction accuracy of a sample-size estimation method for ROC studies. *Acad Radiol* 17:628–638
11. Obuchowski NA (2009) Reducing the number of reader interpretations in MRMC studies. *Acad Radiol* 16:209–217
12. Hany TF, Debatin JF, Leung DA, Pfammatter T (1997) Evaluation of the aortoiliac and renal arteries: comparison of breath-hold, contrast-enhanced, three-dimensional MR angiography with conventional catheter angiography. *Radiology* 204:357–362
13. Ruehm S, Goyen M, Barkhausen J et al (2001) Rapid magnetic resonance angiography for detection of atherosclerosis. *Lancet* 357:1086–1091
14. Ersoy H, Rybicki FJ (2008) MR angiography of the lower extremities. *AJR Am J Roentgenol* 190:1675–1684
15. Kramer H, Michaely HJ, Matschl V, Schmitt P, Reiser MF, Schoenberg SO (2007) High-resolution magnetic resonance angiography of the lower extremities with a dedicated 36-element matrix coil at 3 Tesla. *Invest Radiol* 42:477–483
16. Nael K, Krishnam M, Nael A, Ton A, Ruehm SG, Finn JP (2008) Peripheral contrast-enhanced MR angiography at 3.0 T, improved spatial resolution and low dose contrast: initial clinical experience. *Eur Radiol* 18:2893–2900
17. Griswold MA, Jakob PM, Heidemann RM et al (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med* 47:1202–1210
18. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P (1999) SENSE: sensitivity encoding for fast MRI. *Magn Reson Med* 42:952–962
19. Attenberger UI, Runge VM, Jackson CB et al (2009) Comparative evaluation of lesion enhancement using 1 M gadobutrol vs. 2 conventional gadolinium chelates, all at a dose of 0.1 mmol/kg, in a rat brain tumor model at 3 T. *Invest Radiol* 44:251–256
20. Hadzadeh DR, Von Falkenhausen M, Kukuk GM et al (2010) Contrast material for abdominal dynamic contrast-enhanced 3D MR angiography with parallel imaging: intraindividual equimolar comparison of a macrocyclic 1.0 M gadolinium chelate and a linear ionic 0.5 M gadolinium chelate. *AJR Am J Roentgenol* 194:821–829
21. Leiner T, de Vries M, Hoogeveen R, Vasbinder G, Lemaire E, van Engelsehoven J (2003) Contrast-enhanced peripheral MR angiography at 3.0 Tesla: initial experience with a whole-body scanner in healthy volunteers. *J Magn Reson Imaging* 17:609–614
22. Michaely HJ, Kramer H, Dietrich O et al (2007) Intraindividual comparison of high-spatial-resolution abdominal MR angiography at 1.5 T and 3.0 T: initial experience. *Radiology* 244:907–913
23. Goyen M, Herborn CU, Vogt FM et al (2003) Using a 1 M Gd-chelate (gadobutrol) for total-body three-dimensional MR angiography: preliminary experience. *J Magn Reson Imaging* 17:565–571
24. Kim ES, Chang JH, Choi HS, Kim J, Lee SK (2010) Diagnostic yield of double-dose gadobutrol in the detection of brain metastasis: intraindividual comparison with double-dose gadopentetate dimeglumine. *AJNR Am J Neuroradiol* 31:1055–1058