UNIVERSITY OF HELSINKI REPORT SERIES IN PHYSICS

HU-P-D234

Quality control methods for magnetic resonance imaging in a multi-unit medical imaging organization

Toni Ihalainen

Department of Physics
Faculty of Science
University of Helsinki
Helsinki, Finland

HUS Medical Imaging Center
Helsinki University Hospital
Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty of Science of the University of Helsinki,
for public examination in Auditorium D101 of Physicum,
Gustaf Hällströmin katu 2a, Helsinki,
on February 12th, 2016, at 12 o'clock noon.

Helsinki 2016

Supervisors:

Professor Sauli Savolainen
Department of Physics
University of Helsinki
Finland
and
HUS Medical Imaging Center
Helsinki University Hospital
Finland

Adjunct Professor Outi Sipilä
HUS Medical Imaging Center
Helsinki University Hospital
Finland

Official reviewers:

Professor Miika Nieminen
Research Unit of Medical Imaging,
Physics and Technology
University of Oulu
Finland

Adjunct Professor
Eveliina Lammentausta
Department of Diagnostic Radiology
Oulu University Hospital
Finland

Opponent:

Professor Lauri Parkkonen
Department of Neuroscience
and Biomedical Engineering
Aalto University School of Science
Finland

ISSN 0356-0961
ISBN 978-951-51-1570-6 (printed version)
ISBN 978-951-51-1571-3 (pdf version)

Unigrafia Oy
Helsinki 2016

T. Ihalainen: Quality control methods for magnetic resonance imaging in a multi-unit medical imaging organization, University of Helsinki, 2016, 52 pages. University of Helsinki, Report Series in Physics, HU-P-D234

Keywords: MRI, fMRI, diffusion, quality control, quality assurance, test objects

Classification (INSPEC): A8700, A8760I, B7510N

Abstract

Quality control methods and test objects were developed and used for structural magnetic resonance imaging (MRI), functional MRI (fMRI) and diffusion-weighted imaging (DWI). Emphasis was put on methods that allowed objective quality control for organizations that use several MRI systems from different vendors, which had different field strengths. Notable increases in the numbers of MRI studies and novel MRI systems, fast development of MRI technology, and international discussion about the quality and safety of medical imaging have motivated the development of objective, quantitative and time-efficient methods for quality control. The quality control methods need to be up to date with the most modern MRI methods, including parallel imaging, parallel transmit technology, and new diffusion-weighted sequences. The methods need to be appropriate to those organizations that use MRI for quantitative measurements, or for the participation in multicenter studies.

Two different test object methods for structural MRI were evaluated in a multi-unit medical imaging organization, these were: the Eurospin method and the American College of Radiology (ACR) method. The Eurospin method was originally developed as a part of European Concerted Action, and five standardized test objects were used to create a quality control protocol for six MRI systems. Automatic software was written for image analysis. In contrast, a single multi-purpose test object was used for the ACR method, and image quality for both standard and clinical imaging protocols were measured for 11 MRI systems. A previously published method for fMRI quality control was applied to the evaluation of 5 MRI systems and was extended for simultaneous electroencephalography (EEG) and fMRI (EEG–fMRI). The test object results were compared with human data that were obtained from two healthy volunteers. A body-diameter test object was constructed for DWI testing, and apparent diffusion coefficient (ADC) values and level of artifacts were measured using conventional and evolving DWI methods.

The majority of the measured MRI systems operated at an acceptable level, when compared with published recommended values for structural and functional MRI. In general, the measurements were repeatable. The study that used the test object revealed information about the extent of superficial artifacts (15 mm) and the magnitude of signal-to-noise ratio (SNR) reduction (15%) of the simultaneous EEG–fMRI images. The

observations were in accordance with the data of healthy human volunteers. The agreement between the ADC values for different methods used in DWI was generally good, although differences of up to 0.1 $x10^{-3}$ mm^2/s were observed between different acquisition options and different field strengths, and along the slice direction. Readout-segmented echo-planar imaging (EPI) and zoomed EPI in addition to efficient use of the parallel transmit technology resulted in lower levels of artifacts than the conventional methods. Other findings included geometric distortions at the edges of MRI system field-of-view, minor instability of image center-of-mass in fMRI, and an amplifier difference that affected the EEG signal of EEG–fMRI.

The findings showed that although the majority of the results were within acceptable limits, MRI quality control was capable of detecting inferior image quality and revealing information that supported clinical imaging. A comparison between the different systems and also with international reference values was feasible with the reported limitations. Automated analysis methods were successfully developed and applied in this study. The possible future direction of MRI quality control would be the further development of its relevance for clinical imaging.

Preface and acknowledgements

This research was carried out in the HUS Medical Imaging Center and was supported by the State Subsidy for University Hospitals. It has been a forward-looking environment in which to work and advance one's scientific aims. I especially want to thank CEO Jyrki Putkonen, chief physician Pekka Tervahartiala and former MRI process owner Juha Halavaara for developing this environment and creating its positive atmosphere.

I wish to thank both the current and former Heads of the Department of Physics at the University of Helsinki, Professor Hannu Koskinen and Professor Juhani Keinonen, for providing favorable conditions for me to build upon my knowledge of science. I also wish to thank the management of MATRENA doctoral school for the development of a well-organized and a fruitful scientific environment.

The supervisors of this thesis have shown considerable patience with this research for which I am very grateful. Professor Sauli Savolainen has continuously encouraged me to proceed with the thesis, and given invaluable help with his extensive insight into the fascinating world of medical physics. Adjunct Professor Outi Sipilä has provided in-depth guidance in planning the studies and effectively taught me how to write scientific papers, always reading the drafts carefully and squeezing the last drop of literary contamination and vagueness out of my texts.

I wish to thank the official reviewers of this thesis, Professor Miika Nieminen and Adjunct Professor Eveliina Lammentausta, who have given their extremely valuable and constructive comments.

My co-authors also deserve thanks and I am especially grateful to Nadja Lönnroth. Cooperation with her showed me how inspiring writing a scientific article can actually be, and this effectively revived the pulse of my scientific life. Linda Kuusela was a great co-author in three papers we co-authored as she was willing to help and give a sensible opinion whenever needed. Sampsa Turunen kindly provided his expertise in electroencephalography into the research. Marjut Timonen, Jouni Uusi-Simola, Touko Kaasalainen and Juha Peltonen have been enthusiastic about the subject of this thesis and great colleagues throughout the years. Adjunct Professor Sami Heikkinen introduced me to the world of genuine MR science, and later gave his invaluable help in planning and preparing the test objects, together with the kind help and valuable contribution of Maiju Soikkeli. Eila Lantto and Ali Ovissi provided the radiologists' perspective into the last paper of this thesis, for which I am very grateful.

I feel much obliged to all the colleagues and friends who have encouraged me in my research work. Miia Pitkonen, an inspiring and cheerful colleague, has encouraged me at crucial times, which gave me the important spurts of energy to complete this project. Laura Tuomikoski, who is currently also preparing her PhD thesis, has been my warm and encouraging peer supporter during the finalizing phase of my project. The hilarious

company of my friends of the YL Male Voice Choir, where we originally met, kept me lively and open-minded throughout the years. These are Antti, Hannes, Heikki, Jukka, Mikko, Risto, Sampo, Tommi and Tuomas.

I am deeply grateful to my parents Tauno and Tuulikki who provided me with a safe home where education was appreciated, and ensured that I had the possibility to study the things I wanted. Finally, I owe my deepest gratitude to my wife Suvi for persistently but lovingly driving me to complete the project, and while I was writing this thesis, for taking care of our wonderful little daughters Linnea and Sylvia.

Espoo, January 2016

Toni Ihalainen

Contents

List of original publications

This thesis is based on the following publications:

I Ihalainen T, Sipilä O, Savolainen S 2004 MRI quality control: six imagers studied using eleven unified image quality parameters. Eur Radiol 14:1859-1865

II Ihalainen T, Lönnroth N, Peltonen J, Uusi-Simola J, Timonen M, Kuusela L, Savolainen S, Sipilä O 2011 MRI quality assurance using the ACR phantom in a multi-unit imaging center. Acta Oncol 50:966-972

III Ihalainen T, Kuusela L, Turunen S, Heikkinen S, Savolainen S, Sipilä O 2015 Data quality in fMRI and simultaneous EEG–fMRI. Magn Reson Mater Phy 28:23-31

IV Ihalainen T, Kuusela L, Soikkeli M, Lantto E, Ovissi A, Sipilä O 2015 A body-sized phantom for evaluation of diffusion-weighted MRI data using conventional, readout-segmented and zoomed echo-planar sequences (Acta Radiologica, in press)

For Paper I, the author participated in planning and performed the measurements, wrote the automatic analysis software for the data, analysed the results and wrote the majority of the manuscript. For Paper II, the author participated in some of the measurements, participated in summarizing the individual results and participated in analysing them as an entire entity, and wrote the majority of the manuscript. For Paper III, the author planned the study, carried out the fMRI measurements, participated in the simultaneous EEG–fMRI measurements, analysed the fMRI image quality results and wrote the majority of the manuscript. For Paper IV, the author participated in planning of the phantom, planned and carried out the measurements, analysed the results and wrote the majority of the manuscript. The results of these studies have not been used in other PhD studies.

Abbreviations

AAPM	American Association of Physicists in Medicine
ACR	American College of Radiology
ADC	apparent diffusion coefficient
ADNI	Alzheimer's Disease Neuroimaging Initiative
B_1	magnetic field produced by the radiofrequency coil
BIRN	Biomedical Informatics Research Network
BOLD	blood oxygen level dependent
D	self-diffusion coefficient
DICOM	Digital Imaging and Communications in Medicine
DWI	diffusion-weighted imaging
ECG	electrocardiography
EEG	electroencephalography
EEG–fMRI	simultaneous electroencephalography and functional MRI
EOG	electrooculography
EPI	echo-planar imaging
EU	European Union
FLAIR	fluid attenuated inversion recovery
fMRI	functional magnetic resonance imaging
FOV	field of view
FSE	fast spin echo
MRI	magnetic resonance imaging
MRS	magnetic resonance spectroscopy
N/A	not available
NEMA	National Electrical Manufacturers Association
NMR	nuclear magnetic resonance
OECD	Organization for Economic Cooperation and Development
PACS	picture archiving and communications system
QA	quality assurance
QC	quality control
RF	radiofrequency
ROI	region of interest
SD	standard deviation
SNR	signal-to-noise ratio
STIR	short-tau inversion recovery
T_1	longitudinal relaxation time (spin-lattice relaxation)
T_2	transverse relaxation time (spin-spin relaxation)
T_2*	a time constant incorporating the effects of spin-spin relaxation and magnetic field inhomogeneities
TE	echo time
TO	test object
TR	repetition time

1 Background

Magnetic resonance imaging (MRI) is a medical imaging method that exploits a phenomenon of nuclear magnetic resonance (NMR). The principle of MRI was invented in the 1970s by Nobel laureates Paul Lauterbur and Peter Mansfield (Lauterbur 1973, Mansfield and Grannell 1973). The then novel method was soon introduced into clinics, and the number of MRI examinations has soared since then. The mean number of MRI units per million of population in member countries of the Organization for Economic Cooperation and Development (OECD) was 13.3 in 2011 and the mean number of MRI exams per 1000 population was 55.4 (OECD 2013). MRI involves strong magnetic fields, radiofrequency (RF) fields, and magnetic field gradients, which are used to generate images that originate from hydrogen nuclei in water, fat and other molecules in tissue. The contrast of the image is based on the tissue type, its properties and also the set imaging parameters. Therefore, the permutations of different structural and functional contrasts extracted from tissue are manifold. MRI is a non-invasive method and it does not expose the patient to ionizing radiation.

MRI systems are technologically complicated and involve the use of superconducting magnets, fast switching of gradient fields, and demanding computational operations. These systems can be put to heavy use in clinics, which increases their susceptibility to developing faults. MRI systems are expensive, and examination times are rather long (usually 20-60 minutes per patient), it is therefore important to keep the MRI systems operational as much as possible. The complexity of the system and the cutting-edge technology used entails that the operation of MRI systems requires highly specialized personnel. The specialized personnel include medical physicists and other MRI scientists that are capable of introducing and carrying out performance tests on the systems and evaluating the image quality in an objective way.

A multi-unit public or private medical imaging organization may operate more than one MRI system, and may have one common picture archiving and communications system (PACS). The DICOM standard allows the transfer and archiving of the images. The radiologist that interpretes the images may not be in the same building or even in the same city where the imaging itself takes place. Review of images outside the actual imaging unit or department has been continuously increasing, but it is unlikely that the full potential of this practice has been realized yet (Ranschaert and Barneveld Binkhuysen 2013). The capabilities of the MRI systems within an organization should be known to facilitate correct system selection for each clinical question. Whenever the same types of patients are being imaged by different systems, they should fulfil the same minimum image quality standards to ensure reasonably uniform and clinically adequate results of the imaging. Appropriate methods are therefore needed to control the image quality in this kind of environment. The image quality in this regard can be understood as being of technical or diagnostic quality. The technical image quality can be measured by using test objects and represents the ability of the imaging system to function correctly and produce images within the intended technical parameters such as spatial resolution or uniformity. The diagnostic quality

represents the ability of the system to produce adequate images to enable the correct diagnosis to be made. It can be assessed from patient images and involves the adjustment of imaging protocols that can vary from system to system. Linking these two aspects is not straightforward (Martin et al. 1999).

One definition of quality assurance (QA) is: "a program for the systematic monitoring and evaluation of the various aspects of a project, service, or facility to ensure that standards of quality are being met" (Merriam-Webster 2015). A definition of quality control (QC) is "an aggregate of activities (as design analysis and inspection for defects) designed to ensure adequate quality especially in manufactured products" (Merriam-Webster 2015). These concepts can also be applied to medical activities, such as radiological imaging. Using these definitions in the scope of this work, QA is understood to involve the quality evaluation of the whole imaging process, and QC, as a part of it, is considered more appropriate term for technical image quality testing. An integral part of QA is also the acceptance testing of new devices and methods for clinical use.

Since the 1990s there has been a lot of discussion and actions relating to the quality of radiological imaging. This has been especially strong in the imaging modalities that use ionizing radiation. The European Union (EU) directive 97/43/EURATOM obliged the EU member states to implement a clinical audit, a process that ensures the application of generally accepted good practice in all medical uses of radiation (European Commission 2009). Although the quality of practice involves several issues, technical image quality is a key link in the chain. The discussion regarding MRI has been less active and regulations or recommendations have been scarce. MRI is, according to present knowledge, a safe imaging method. The currently identified risks mainly relate to metal objects inside the patient's body and the injection of contrast agents (Kanal et al. 2002). There is, however, an indirect but notable risk for the patient from inadequate QC of the equipment (de Certaines and Cathelineau 2001). A British survey that was published in 2006 reported that 21% of hospitals did not undertake any form of in-house MRI QC (Koller et al. 2006). There have been, however, international efforts to fulfil the MRI QA objectives set by the European Communities (Lerski and de Certaines 1993), American Association of Physicists in Medicine (AAPM) (Price et al. 1990, Och et al. 1992), and American College of Radiology (ACR) (Weinreb et al. 2005).

New innovations, such as imaging sequences, are introduced into the clinical use of MRI every year. These innovations may in turn require new approaches to evaluate and control image quality. The requirements of the diverse range of submethods need to be taken into account in the QC of MRI. Magnetic resonance spectroscopy (MRS) and functional MRI (fMRI) are designed to detect very weak signals that could easily be masked by extraneous background signals, this feature emphasizes the need of noise-related measurements. Another example is the use of MRI in radiotherapy treatment planning, which to date can be based solely on MRI (Kapanen et al. 2013); this requires pronounced accuracy in geometry. In general, an MRI system is an instrument that allows quantitative

measurements. The efforts towards quantitative MRI set requirements for controlling accuracy and precision of the measured values (Tofts 2003).

The motivation for this research project lies in keeping abreast with the growth and organizational concentration of MRI by ensuring and integrating the quality of the images from a wide variety of systems. This research was carried out within an expanding imaging organization of a public hospital district that consisted of a central university hospital, a growing number of regional units joining in, and mobile units. Some of the MRI systems were used for basic clinical imaging, and some for imaging of the most difficult clinical cases in addition to carrying out scientific research. When the research began in 2002, there were 6 MRI systems in 3 different hospital buildings, and the maximum physical distance between the individual systems was 4 km. In 2014, the operational region of the imaging center had grown to 9000 km^2 in size, it served 1.58 million inhabitants, and performed 61000 MRI exams with 14 MRI systems that year. A common PACS system has been in use since the beginning of the work. Economic pressures and growing demand for imaging examinations has led to the emphasis of cost-effectiveness and, consequently, time-effectiveness of MRI practices. The MRI systems of three different vendors have been in use during the entire time period of this study. The general aim of this research was to answer the needs of controlling the image quality of MRI in a growing organization and in balance with other imaging modalities. It was stated at the beginning of the research that regular and objective QC should be initiated in the imaging center to ensure as accurate and precise results for MRI exams as possible. The QC methods should incorporate both structural and functional imaging, and they should be robust enough to enable comparisons of different MRI systems of different field strengths. The test objects and the image analysis should be independent of the MRI system vendors.

2 Aims of the study

The aims of this study were

1. to develop and use vendor-independent and objective methods for QC of structural and functional magnetic resonance imaging in an organization that operates several different MRI systems;

2. to use the methods to compare the performance of the MRI systems with each other and also against international reference values;

3. to extend the QC methods to encompass emerging and evolving MRI techniques, including simultaneous electroencephalography and functional MRI (EEG–fMRI) and diffusion-weighted MRI of the body;

4. to develop and apply automated procedures for the analysis and measurement of quality control data.

3 Principles of image formation and quality

3.1 Structural MRI

The purpose of structural MRI is to get as true information about the patient anatomy as possible. Anatomical structures should be clearly delineated and identifiable. Artifacts i.e. any signals that do not represent the anatomy and that degrade the diagnostic potential of the exam should not be present (McRobbie et al. 2005, Bushberg et al. 2002).

The physical principles of MRI and its underlying phenomenon of NMR have been covered in several comprehensive textbooks (Haacke et al. 1999, Young 2000, McRobbie et al. 2005). The signal generation in MRI begins with RF excitation of hydrogen nuclei with a transmitting coil. The RF excitation is possible when the object is in a magnetic field, which creates a net magnetization within the object, and the frequency of the excitation matches the resonance frequency of the hydrogen nuclei. The orientation and phase of the net magnetization can be controlled by the excitation and by a further combination of RF pulses and gradient fields. Finally, the signal is recorded by a receiving coil. The spatial origin of the signal can be resolved by using a carefully designed sequence of excitations and signal recordings along with the appropriate gradient fields. The samples of the received signals are measured, and a change of variables is made to store the information in a temporary image space called the k-space, with frequency (readout) and phase encoding directions. This information can be transformed into an image of the object with a 2- or 3-dimensional Fourier transform.

The signal in MRI is dependent on chemical properties of tissue. These properties include proton density and two relaxation times, T_1 (spin-lattice relaxation) and T_2 (spin-spin relaxation). The image acquisition parameters define the sequence of RF pulses and gradients, and control the weighting of the effect of the different chemical properties of the tissues. They also control orientation, resolution, acquisition time, and signal-to-noise ratio (SNR) of the image.

The basic pulse sequence of NMR and MRI is based on spin echo (Hahn 1950) whereby the transversal component of excited and subsequently dispersing net magnetization is rephased by a refocusing RF pulse and an echo of the signal is recorded after time TE (echo time) from the original excitation. The acquisition is sequentially repeated after time TR (repetition time) to control the amount of longitudinal relaxation left in the recorded signal. The signal equation in a simple form is

$$S = k\rho(1 - e^{-TR/T_1})e^{-TE/T_2} \quad (1)$$

where selection of timing parameters *TR* and *TE* can be used to create different weightings of the T_1 and T_2 relaxation, and proton density ρ. Factor k represents other

effects, such as flow or sensitivity of the coil (Bushberg et al. 2002). Adjusting the timing parameters is a common control mechanism to alter image contrast in MRI. The spin echo sequence can be modified by the addition of an inversion pulse, which is known as inversion recovery technique. It can be used to nullify the signal from fluids, known as fluid attenuated inversion recovery (FLAIR) or to nullify signal from fat known as short-tau inversion recovery (STIR). The spin echo sequence in image acquisition is slow, and is therefore impractical, however. In present-day clinical structural imaging, the most important sequences are fast spin echo (FSE), in which signals from multiple echoes are collected within one excitation, and various gradient echo sequences that use gradient fields instead of RF pulses for echo formation. The gradient echo sequences allow fast imaging of the abdomen within one breath-hold and thus minimizing the effect of respiratory motion.

In addition to the contrast, the key image quality properties in MRI include SNR, signal uniformity, spatial resolution, and artifacts. The SNR defines the relation of mean pixel intensity and random variations in pixel intensities in the image. A low SNR may prevent detection of low contrast objects in the image. The SNR is proportional to the field strength and it is also a function of voxel size, pixel bandwidth, and the number of repeated signal acquisitions. Signal uniformity on an image plane can be degraded by RF or magnetic field inhomogeneity, eddy currents, gradient miscalibration, or coil structure. Spatial resolution is important in observing fine structures in the image, but is in practice limited by selected voxel size, apart from certain acquisition techniques such as echo-planar imaging (EPI). (Price et al. 1990, Och et al. 1992)

Image artifacts are characteristic to MRI. The artifacts may be inherent to the MRI method, or arise from malfunctioning of the MRI system or from interference from external sources, or from the patient. The chemical shift causes water and fat signals to appear in slightly different positions in the image due to the differences in their respective resonance frequency. The magnitude of chemical shift can be calculated when the acquisition parameters are known. Ghosting is manifested as faint images of the original object in the phase encoding direction and is caused by phase errors in the acquisition process. Interference due to incomplete RF shielding, RF noise spikes, or malfunctioning system components can spoil image quality. Susceptibility artifact appears as a signal loss and distortion and is caused by differences in local magnetic field in the boundary of materials of different magnetic susceptibility. Artifacts that arise from motion or pulsation are often problematic and unavoidable as each signal collection includes information from all spatial locations in the image. There are several other types of artifacts, some of which can be utilized for contrast formation (McRobbie et al. 2005).

In practice, a compromise between spatial resolution, SNR, and acquisition time needs to be made in every acquisition. The sufficiency of image quality can only be determined by reflecting it to the imaging problem at hand, i.e. the specific clinical or scientific question under consideration. The following general requirements of the image quality can however be formulated for clinical structural MRI, based on available clinical guidelines and technical literature (American College of Radiology 2013, Price et al. 1990):

(1) The signal intensity of a homogeneous object should be spatially uniform (nonuniformity may be accepted under certain circumstances)
(2) The SNR should be acceptable for clinical imaging
(3) The geometry of the image should represent the true geometry of the patient
(4) The contrast should appear as intended, and be usable for the clinical problem at hand
(5) The spatial resolution should be good enough for clinical imaging, and as specified by imaging parameters
(6) The thickness and location of the acquired slice should be as specified
(7) Artifacts, such as ghosting, should be absent or minimal

There have been substantial improvements in the receiver coil technology, in the acquisition methods and in reconstruction methods during the period that this research work was carried out. The receiver coils were changed from single-channel coils to multi-element coils, and the MRI systems were able to handle up to 32 receiver channels simultaneously. This development also allowed also the use of parallel imaging (Sodickson and Manning 1997, Pruessmann et al. 1999, Griswold et al. 2002). Sensitivity profiles of individual coil elements are measured and that information is used as part of the image reconstruction, which enabled accelerated imaging. Multi-element coils also take advantage of coil sensitivity normalization. This method corrects for the signal attenuation from the structures that are located deeper from the surface, with the help of coil sensitivity maps (Griswold et al. 2002b).

3.2 Functional MRI and EEG–fMRI

The purpose of functional medical imaging is to get as true information of the physiological or pathological function of the patient as possible, such as information on any changes in metabolic function that might be occurring. The term functional MRI usually refers to the imaging of the blood oxygen level dependent (BOLD) contrast in the brain. In a wider context it can also contain diffusion and perfusion weighted imaging.

The fMRI and diffusion-weighted MRI methods almost exclusively use EPI sequences (Mansfield 1977). The EPI sequence is an extreme example of fast MRI, in which the whole slice can be collected by only one (single-shot EPI), or a few (multishot EPI) RF excitations. The nature of the data collection makes it vulnerable to magnetic susceptibility effects, which cause signal distortion, possible signal void areas, and blurring. There are several approaches to implement the signal acquisition process, and these variants can be utilized to optimize the image quality and level of artifacts. The EPI sequence poses specific requirements to the speed of the RF system, gradient system, and data processing. Images can be produced in less than 100 ms. High temporal resolution, however, is achieved at the expense of spatial resolution and SNR (Price et al. 2002).

The haemodynamic response such as BOLD in fMRI is measured in the brain during the execution of a predetermined stimulus or task (Ogawa 1992) or during the resting state (Biswal 1995). The consumption of glucose is increased in brain cells when they are activated. This triggers processes that increase the flow of oxygenated blood in the activated area. When diamagnetic oxyhaemoglobin molecules give up their oxygen, they become paramagnetic. There is a local relative increase of oxygenated haemoglobin in the activated area and there are fewer susceptibility effects than in non-activated areas. This in turn leads to increases in the local signals in the data acquired with a T_2*-weighted EPI sequence that incorporates the effects of local magnetic field inhomogeneities into the signal. The signal difference between activation and selected baseline is small compared to noise, which makes it necessary to perform a large number of repeated measurements, i.e. long sequential acquisition of echo-planar images, both during activation and baseline state. The MRI system must be stable to enable this, otherwise the small contrast difference is lost or an artificial contrast may even be generated in the data (Purdon 1998).

The fMRI method has been extensively used in cognitive neuroscience (Logothetis 2009). The clinical use of fMRI is mainly related to planning of neurosurgical operations. The common objective in neurosurgical applications is to identify the location of key functional areas in the brain to avoid damage during the operation (McRobbie et al. 2005).

The fMRI can be combined with simultaneous extracranial (Goldman et al. 2000) or intracranial EEG (Carmichael et al. 2012). The extracranial EEG–fMRI method requires a cap containing e.g. 64 electrodes to be placed on the patient's head, in such a way that the electrodes are put into intimate contact with patient's scalp, the contact of which is enhanced by an appropriate gel. The EEG signal during fMRI is directed to a computer in a control room through MRI compatible amplifiers and leads. The simultaneous EEG–fMRI can help in localizing epileptic foci (Lemieux et al. 2001, Gotman et al. 2006). The simultaneous acquisition, however, poses technical challenges. The EEG equipment produces susceptibility artifacts and a reduction in the image SNR (Carmichael 2010, Scarff et al. 2004). The MRI system conversely has effects on the EEG data (Ritter et al. 2010). Much of the artifacts in EEG signals that originate from switching gradient fields can be corrected with an average artifact subtraction method (Allen et al. 2000). The patient safety regarding EEG–fMRI may be a complicating factor and these issues are mainly related to interactions between the patient, the EEG leads and the electromagnetic fields of the MRI system (Lemieux et al. 1999).

3.3 Diffusion-weighted MRI of the body

Water diffusion in tissue is restricted by cell walls, nerve fibers and other anatomical structures. Changes in water diffusion are measured in diffusion-weighted imaging (DWI) and this method is usually used to detect pathological locations in tissues. DWI in the brain is an important tool in the diagnosis of acute ischaemic stroke (Schaefer et al. 2000).

Technological developments have enabled the increased use of DWI in the body area (Thoeny and De Keyzer 2007) where it is mainly used in oncology, and thus its significance has increased during the last few years. Important organs for which the DWI is used include the liver, kidneys and prostate (Sigmund and Jensen 2011).

The basic imaging sequence of DWI in a clinical setting is single-shot spin-echo EPI (Sigmund and Jensen 2011). Imaging gradients are used to measure diffusion in each voxel (Stejskal and Tanner 1965, Le Bihan 1986). The technique is based on two gradient lobes and an RF pulse. The first gradient lobe generates a phase shift for the spins and, after an inverting 180° RF pulse, another gradient lobe generates another phase shift. The phase shifts for static spins cancel each other due to the inversion pulse. However, a certain phase shift remains in the spins that undergo diffusion and the signal is reduced in the measuring pulse sequence part that follows (Wheeler-Kingshott et al. 2003). The strength of the signal is given by

(2) $$S(b) = S(0)e^{-bD}$$

where *S(0)* is the signal value in non-diffusion-weighted image, *b*-value specifies the diffusion weighting, and *D* is a self-diffusion coefficient (McRobbie et al.2005). The value of *b* can be expressed as

(3) $$b = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3})$$

where *G* and δ are amplitude and duration of the gradient lobes, respectively, Δ is a separation time between the two gradient lobes, and γ is the gyromagnetic ratio (McRobbie et al. 2005). The application of the gradients for diffusion weighting is presented in Fig. 1.

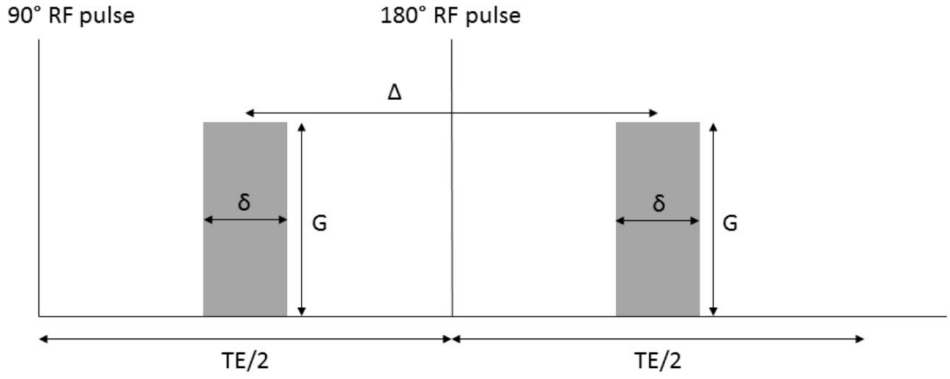


Figure 1 *A pulse diagram for diffusion weighting. Typically, this scheme precedes a spin-echo echo planar imaging acquisition. The Δ is the time interval between gradients; G and δ are amplitude and duration of diffusion gradient, respectively. TE = echo time, RF = radiofrequency (Tofts 2003, McRobbie et al. 2005).*

The diffusion measurement depends on gradient direction, but in clinical imaging the mean value of the diffusion coefficients of three orthogonal directions is often used. It is customary in DWI to use the term apparent diffusion coefficient (ADC) instead of *D*, because the measured signal usually contains components other than those of pure diffusion, such as microcirculation. The ADC value can be calculated from a minimum of two images. Usually pixel-by-pixel ADC values are represented as an ADC map.

The ADC value in the body area could serve as a cancer biomarker, but the reported ADC values of malignant and benign lesions partially overlap. A major problem in DWI has been the lack of standardization of the imaging protocols (Padhani et al. 2009). Sasaki et al. have also shown that variability in ADC values across different vendors can be substantial (Sasaki et al. 2008). If the ADC values measured in different MRI systems are cross-calibrated, then the ADC values could be used for the evaluation of the therapy response of cancer patients. The use of DWI in the body is at present mainly used to detect suspicious lesions and distinguish them from cysts. An MRI examination method that has increasing clinical significance is the whole-body DWI, which can be used for staging and directing the care of cancer patients (Padhani et al. 2011).

In general, abdominal MRI is more problematic when the field strength is 3.0T compared to 1.5T. Image artifacts may often compromise the result despite the increase in SNR. The inhomogeneous distribution of RF energy (or, alternatively, inhomogeneous B_1 field) can cause large signal voids in the image, which can severely hinder the diagnosis. Using parallel transmit technology, the problem can be overcome (Katscher and Börnert 2006). Parallel transmit also allows shimming of the B_1 field and excitation of the selected volume only, which can be exploited in a method called zoomed EPI (Pfeuffer et al. 2002, Rieseberg et al. 2002). This emerging method is compatible with DWI. Another advancement in diffusion-weighted imaging is the application of readout-segmented EPI sequence (Porter et al. 2009). The acquisition obtained by this method is divided into several segments in readout direction, which minimizes the echo spacing parameter and thus reduces artifacts. Other methods such as single-shot FSE and segmented radial FSE sequences for DWI are also available but have mainly been used for imaging of the head (Sigmund and Jensen 2011).

4 Quality control

4.1 Structural MRI

Most of the QC tests in MRI are based on the acquisition of an image of a test object (Figs. 2 and 3). Different image quality parameters are then measured. These measurements can be carried out by visual grading, manual measurement from the image with the appropriate software tools, or by automated analysis.

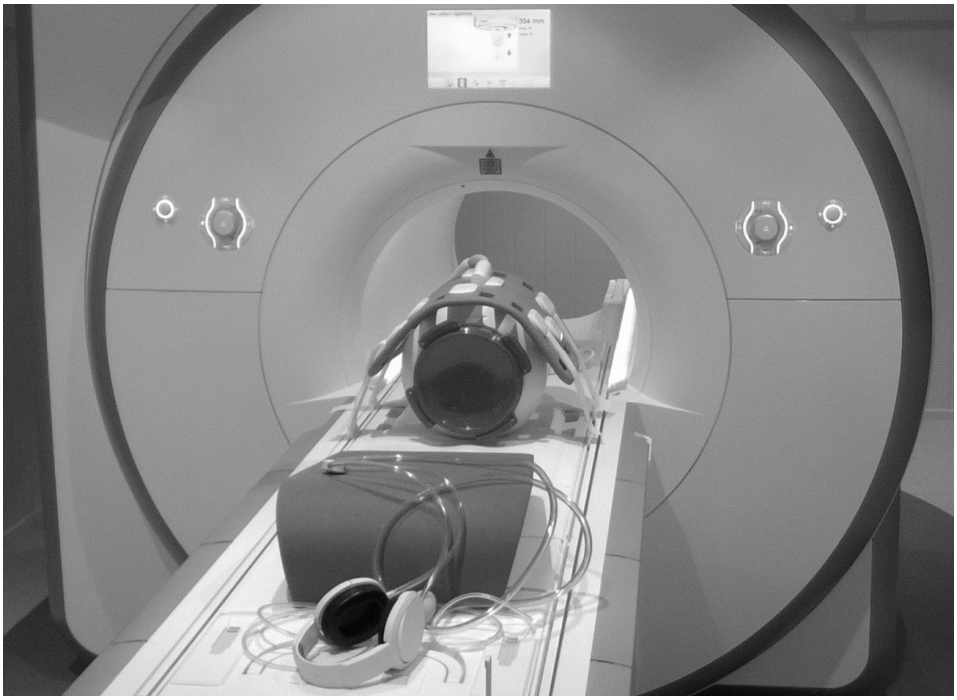


Figure 2 *A test object positioned on an imaging table of an MRI system.*

The important QC parameters include image uniformity, SNR, image geometry and spatial resolution. The image uniformity can be measured by searching a maximum and minimum pixel value (S_{max} and S_{min}) from a homogeneous test object and calculating a value for integral uniformity U (Price et al. 1990)

(4) $$U = \left[1 - \frac{S_{max} - S_{min}}{S_{max} + S_{min}}\right] \cdot 100\%.$$

The SNR can be measured by dividing mean pixel value of a test object by the pixel value standard deviation (SD) measured from background area (air), or from a subtraction of two sequential images (Dietrich et al. 2007). The latter method takes into account the non-uniform noise distribution that originates e.g. from a coil structure. Geometrical accuracy can be evaluated by measuring known distances from the image, and slice thickness can be measured from ramps tilted at a known angle (Price et al. 1990). Spatial resolution can be measured from line pair blocks of different densities, or calculated from a measured edge spread function (Delakis et al. 2009). Other possible QC parameters include ghosting, SNR uniformity, slice thickness, slice position, slice warp, T_1 or T_2 accuracy, and low contrast detectability, which is complementary to the measurement of SNR.

Each MRI system is equipped with the vendor's own test objects that can be used in user QC, servicing, and calibrating typically along with automated analysis procedures. These tools, however, are not designed for comparison of image quality between different MRI systems in a multi-system center or in multicenter studies. There are multiple test objects, or phantoms, available for MRI quality assurance that are independent of the MRI system vendors. The Eurospin phantom set was designed as a part of the European Concerted Action "Tissue Characterisation by MRS and MRI" (Lerski and de Certaines 1993, Lerski 1993). The UK Government's Centre for Evidence-based Purchasing has used the MagNET phantoms in its evaluation program (De Wilde et al. 2002). The American College of Radiology has an accreditation program, which includes a phantom test, and the phantom can also be used in regular quality control (Weinreb et al. 2005). The Alzheimer's Disease Neuroimaging Initiative (ADNI) study has used a multi-purpose phantom (Gunter et al. 2009). Other commercial and non-commercial test objects are also available.

4.2 Functional MRI and EEG–fMRI

The BOLD effect is difficult to mimic with phantoms, although some solutions have been suggested. A minuscule electrical current can be applied through a phantom to affect the T_2* relaxation time of the phantom (Renvall et al. 2006). A two-compartment gel phantom has also been suggested for this purpose (Olsrud et al. 2008). The quality control of temporal stability parameters of the fMRI series has been the subject of several studies (Weisskoff 1996, Simmons et al. 1999, Stöcker et al. 2005). Moreover, artifacts caused by imperfections in data acquisition, including signal loss due to local magnetic field inhomogeneities and geometric distortions characteristic to the EPI sequence, are critical problems in fMRI applications. Geometric distortion has been evaluated using a phantom (Mattila et al. 2007). Phantom experiments for image quality evalution in simultaneous EEG–fMRI have also been published (Krakow et al. 2000, Mullinger et al. 2008).

4.3 Diffusion-weighted MRI of the body

One of the challenges stated in 2009 in the consensus paper by Padhani et al. was the development of quality assurance methods and phantoms (Padhani et al. 2009). There are several solutions to the problem that have been proposed to date. Delakis et al. have developed and applied a QC protocol for diffusion MRI studies (Delakis et al. 2004). Chenevert et al. published a study of an ice-water phantom (Chenevert et al. 2011), followed by a repeatability and reproducibility study (Malyarenko et al. 2012). Lavdas et al. introduced a DWI phantom using different concentrations of agarose, sucrose, and nickel, and used it for comparison of 1.5T and 3.0T MRI systems (Lavdas et al. 2013, Lavdas et al. 2014). Other studies have investigated suitable materials for DWI phantoms (Tofts et al. 2000, Laubach et al. 1998, Matsuya et al. 2009). Miquel et al. have studied the repeatability of abdominal DWI, with phantom and *in vivo* (Miquel et al. 2012). Most of the literature mentioned above has concentrated on the accuracy and repeatability of ADC values.

The ADC measurement is sensitive to temperature. The temperature dependence is approximately 2.4%/K for water (Le Bihan et al. 1989). This is not a problem in clinical abdominal imaging, but may cause significant variation in phantom studies. Therefore the temperature of the phantom should be controlled or measured directly. One such control method was the ice water as described by Chenevert et al. (Chenevert et al. 2011).

4.4 Automated analysis

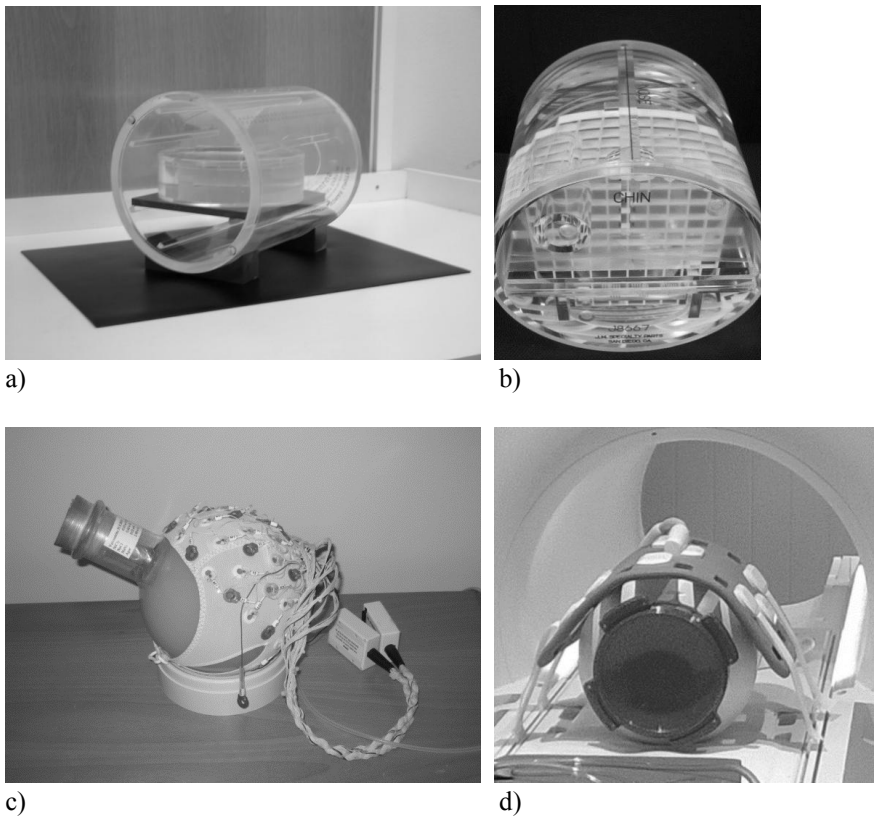
Automated analysis of quality control images has multiple objectives. It is supposed to make tests more objective, faster, and easier to perform on a routine basis.

Automated analysis of MRI or fMRI quality control has been studied by several groups. Covell et al. showed as early as 1986 that automated analysis methods work reliably for MR images (Covell et al. 1986). Gardner et al. have shown that an automated analysis of a test object can reveal image degradation long before trained human observers are able to do so (Gardner et al. 1995). Bourel et al. developed analysis software that could be used for several MRI systems, which was reported to be accurate, reliable and fast (Bourel et al. 1999). Simmons et al. (Simmons et al. 1999) and Stöcker et al. (Stöcker et al. 2005) have both developed an automated analysis procedure for fMRI, and Simmons et al. also applied a Shewhart charting method to detect changes in system performance during long-term quality control. The automated analysis methods have not been particularly widespread, but during the last decade, several groups began to develop and implement the methods, especially for the purposes of multicenter studies (Friedman and Glover 2006, Davids et al. 2014). The MRI system vendors have implemented automated analysis routines into their software for quality control tests. Commercial software solutions for automated analysis are also available.

5 Methods in this study

5.1 Structural MRI

Image quality of MRI systems of a multi-unit medical imaging organization was assessed using two different test object methods. A set of Eurospin phantoms (Diagnostic Sonar Ltd, Livingston, Scotland) was used in Paper I and the ACR MRI accreditation phantom (J.M. Specialty Parts, Inc., San Diego, CA, USA) was used in Paper II. Those phantoms were designed for measurements with the head coil in clinical use on each system.



a) b)

c) d)

Figure 3 *Phantoms used in this study. a) Eurospin phantoms, along with positioning aids (Paper I); b) the ACR MRI accreditation phantom (Paper II); c) the fMRI phantom, which is partly covered by the EEG cap (Paper III); d) the body-sized diffusion phantom (Paper IV)*

The motivation behind Paper I was to provide information to initiate a quality control program that would, for the first time, encompass all MRI systems in our organization. The Eurospin phantom set was selected for this purpose. It consisted of five different test objects (TO1-TO5) filled with copper sulfate solution. TO1 was a uniform cylinder for measurement of SNR and homogeneity. TO2 contained ramps for slice thickness measurements, in addition to a 120x120 mm frame for geometric distortion evaluation. TO3 was used for slice position measurement with the help of angled rods. TO4 contained bar patterns for spatial resolution evaluation. TO5 included test tubes with agarose and gadolinium chloride solutions for T_1 and T_2 measurements. Their T_1 and T_2 values were reported in the handbook of the phantom set (Diagnostic Sonar Ltd 1992–1995). The measurements were to be performed in transversal, sagittal and coronal slice orientations according to the manufacturer's instructions. This required the construction of positioning aids that were made from polyvinyl chloride.

The measurement and image analysis protocol was formulated according to the recommendations laid down in the Eurospin manual (Diagnostic Sonar Ltd 1992-1995) and the AAPM reports (Price et al. 1990, Och et al. 1992), and knowledge of vendor-specific QC measurements performed daily or weekly. The 11 parameters under evaluation were image uniformity, magnitude of ghosting artifact, SNR and its uniformity, geometric distortion, slice thickness, slice position, slice warp, resolution and T_1 and T_2 accuracy. The T_1 relaxation times were calculated by using the inversion recovery sequence with five different inversion times and T_2 relaxation times by using the spin echo sequence with five different TE values. The measurements of the protocol were performed for 6 MRI systems from different vendors and which had different field strengths (Table 1). The measurements were repeated after two weeks on one MRI system and after half a year on all the measured systems to tentatively evaluate the short-term and long-term usability of the measurement protocol in addition to verifying the stability of the systems. The measurement time of the protocol for one system, including phantom positioning, was 3–4 hours.

The quality control protocol initiated in Paper I was in use for several years. When the first system with the field strength of 3.0T was purchased, the phantoms were changed to the MagNET phantoms (MagNET, London, UK), which in turn were in use for several years. It was decided that the feasibility of the ACR phantom for quality control purposes in our organization should be evaluated for practical reasons. The ACR phantom used was cylindrical in shape and had a diameter of 148 mm and length of 190 mm. It had several structures in different layers. These internal structures enabled measurements of geometric accuracy, spatial resolution, slice thickness accuracy, slice position accuracy, image intensity uniformity, percent signal ghosting and low-contrast detectability. The standard T_1- and T_2-weighted spin echo sequences given by the ACR and each site's own clinical T_1- and T_2-weighted sequences were used in the measurement protocol. The images were acquired in transversal plane only, apart from a locator image that was acquired from the sagittal plane.

Table 1 The MRI systems investigated in this study.

System No.	Field strength	Year of purchase	Published papers
1	1.5T	1994	I
2	0.23T	1996	I
3	1.0T	1996	I, II
4	1.5T	1996	I
5	1.5T	2000	I, II, III
6	1.5T	2000	II
7	1.5T	2001	I
8	1.5T	2001	II
9	1.5T	2004	II
10	1.5T	2004	II
11	3.0T	2006	II, III
12	1.5T	2006	II, III
13	1.5T	2009	II
14	1.5T	2010	II, III
15	3.0T	2010	II, III
16	1.5T	2014	IV
17	3.0T	2014	IV

The ACR phantom measurements were performed on 11 MRI systems in the organization (Table 1), and repeated measurements were carried out three to nine months after the first measurement for each system. The measurement time of the protocol for one system was 20 minutes. The images were analysed using the ACR instructions (American College of Radiology 2005). The analysis was partly visual, i.e. it determined the high-contrast spatial resolution from distinguishability of hole-array-pairs, or low-contrast object detectability by calculating number of visible objects with gradually decreasing contrast and object size. Part of the analysis, such as the evaluation of geometric accuracy was carried out by running the measurement tools of an image processing software ImageJ (Schneider et al. 2012). Details of the image analysis methods are presented in the reference document (American College of Radiology 2005). The motivation behind Paper II was to summarize and evaluate the results of the ACR measurements.

5.2 Functional MRI and EEG–fMRI

A question of temporal stability had been raised in connection with the measurements of structural MRI (Paper I). The motivation behind Paper III was to extend the quality control program to include fMRI. For the purposes of our work, we sought a method that

would be suitable for several field strengths and would be vendor-independent. The method should also be internationally acknowledged and enable the measurements with the EEG equipment. Friedman and Glover (Friedman and Glover 2006) published a quality assurance protocol that has been used in the Biomedical Informatics Research Network (BIRN). A uniform gel phantom was used in that protocol. We chose to use and further develop their method. A phantom in a glass ball with an internal diameter of 18 cm (Fig. 3c) was constructed, according to the given instructions, consisting of 3600 mL H_2O, 400 mL 21.8 mM $NiCl_2$, 120 g of agar, 20 g of NaCl (0.5%) and 1 g of sodium azide (Friedman and Glover 2006). This composition was intended to mimic the relaxation properties of grey matter. The basic fMRI measurements were carried out on the five MRI systems utilized for fMRI (Table 1) and repeated after approximately one year. The phantom was positioned in a head coil that was used in clinical or research fMRI studies. An axial gradient-echo based single-shot EPI sequence was used with 200 acquired time frames. The sequence was similar to that described by Friedman and Glover (Friedman and Glover 2006) to enable a direct comparison with their results. The measurements on two 3.0T systems were carried out with and without coil intensity normalization. The images were analysed automatically as described in Section 5.4.

Apart from the fMRI measurements, the protocol was used for carrying out EEG–fMRI measurements on two 3.0T systems. A 64-channel EEG cap (BrainProducts, Inc., Munich, Germany) was placed on top of the phantom and the electrodes were interconnected with ECI Electrogel (Electro-Cap International, Inc., Eaton, OH, USA) or EC2 Genuine Grass Electrode Cream (Grass Technologies, Inc., Warwick, RI, USA) to prevent damage to the EEG equipment. The amplifiers were taken into the imaging room, as in human imaging experiment. The EEG signal was also recorded, and subsequently analyzed by a BrainVision Analyzer 2.0 (BrainProducts Inc., Munich, Germany). A separate EEG recording was performed outside the MRI room. Noise behaviors that were studied included also an analysis of the effects that the MRI system bore ventilation, the lighting, and the helium pump had upon the EEG signal.

The penetration depth of the superficial artifacts was measured slice by slice from the EEG–fMRI images using the measurement tool of ImageJ software, in addition to the same image analysis as used in the fMRI measurements. We also retrospectively examined human volunteer fMRI data obtained from two subjects to evaluate whether our observations related to EEG–fMRI image quality of the phantom would be similar to human data. The ethical issues of this part of the study are described fully in Paper III. The human data consisted of three imaging sessions (one subject participated in two sessions). Each voluntary session consisted of four identical acquisitions with and without the EEG equipment. We first selected two slices from each acquisition, one at the level of the centrum semiovale and the other, 24 mm above it. We calculated the SNR values for these slices in a manner similar to that of the automated analysis for the phantom data. Again, ImageJ software was used. Medial and lateral square ROIs were placed on each of the selected slices, using the same ROI coordinates for each respective slice. We then calculated the

SNR values for each ROI. Finally, we calculated the mean SNRs of four acquisitions and subsequently, calculated the SNR difference caused by the EEG equipment for each ROI.

5.3 Diffusion-weighted MRI of the body

The motivation behind the research published in Paper IV was to develop tools for evaluating image quality of the DWI of the body area, which was a method that had become significant in clinical imaging and was being performed by 1.5T and 3.0T systems in our organization. None of the test objects in the previously published DWI studies had represented dimensions similar to those of a typical patient in abdominal examination. A cylindrical phantom with a diameter of 31 cm and total volume of 26 litres was constructed for this study (Fig. 3d). The phantom contained four samples filled with gels that were intended to produce a clinically relevant range of ADC values that varied from 0.8 to 2.0 $x10^{-3}$ mm^2/s. The gels were prepared in accordance with a previously published study by Lavdas et al. (Lavdas et al. 2013) and consisted of different concentrations of agarose, sucrose and nickel. A plastic frame was constructed and installed inside the phantom to keep the samples in their correct positions.

The phantom measurements of this study were carried out using two new MRI systems (1.5T and 3.0T, Table 1), after their respective installations but prior to when they were taken into actual clinical use. The phantom was taken into the imaging room the day before measurements were taken, to let its temperature stabilize to that of the imaging room temperature. The room temperature was recorded before and after the measurements. The images were acquired by a 32-channel spine coil and an 18-channel body coil, as in the clinical examination. The effects of several imaging options on diffusion-weighted images and ADC values were measured. The following series were acquired: 1) conventional abdominal DWI EPI series, 2) conventional series without parallel imaging, 3) conventional series with coil intensity normalization, 4) readout-segmented EPI, 5) zoomed EPI (3.0T), and 6) conventional series with patient-specific B_1 shim (3.0T). Each series was acquired twice. The acquired b values were 0, 400 and 800 s/mm^2, though b=400 s/mm^2 was not acquired for readout-segmented EPI. Detailed sequence parameters can be found in Paper IV.

The ADC values of all the samples were measured for each ADC map. Two different ROIs were used. A ROI of 1200 mm^2 was placed over the most artifact-free region of the sample. A circular ROI of 3800 mm^2 covered 75% of the sample, and any possible artifacts were ignored. The SNRs of the trace-weighted images (b=800 s/mm^2) were measured for each sample from using a difference image of two consecutive acquisitions for the noise calculation. The level of artifacts was evaluated for trace-weighted images (b=0 s/mm^2 and b=800 s/mm^2) and ADC maps in consensus by two experienced observers. The edge detection images that were calculated from the original images using the Sobel filter in this evaluation were used as a supplement, because artifacts were better visualized in these images. The following image quality score was given to each image: 0 = no artifact or

obtrusive local noise; 1 = small or mild single artifact or some obtrusive noise; 2 = substantial artifact or noise, yet enabling an ADC measurement or reading of the image; 3 = artifact or noise that notably affects the ADC measurement or reading of the image. A relative artifact level index was calculated from these results for each acquisition option.

5.4 Automated analysis

Automated analysis for the images was used in Papers I and III. The software used in Paper I was self-written in the Java programming language (developed by Sun Microsystems, CA, USA; acquired in 2009 by Oracle Corporation, CA, USA) as a plugin for the image processing software ImageJ (Schneider et al. 2012). The first step of the automated analysis was to locate the center point of the phantom in the image with the circular Hough transform (Hough 1962). The same orientation of the phantom between measurements was achieved by using strict positioning, and it was possible to position the ROIs automatically. The ROIs selected by the software were shown to the user as superimposed on the images, for acceptance (Fig. 4). Image uniformity was calculated as in Equation 4. The SNR was measured by two different methods: 1) from a mean pixel value of a ROI that had been drawn on a homogeneous phantom, and an SD value of background ROI, and 2) using a method devised by the National Electrical Manufacturers Association (NEMA) that is based on a difference image (National Electrical Manufacturers Association 1988). The percentage value of signal ghosting was calculated according to directions by the ACR (American College of Radiology 2001). The slice thickness was measured according to directions of the Eurospin manual (Diagnostic Sonar Ltd 1992–1995). Full details of these and the other image analysis methods used are presented in Paper I. All the automatically calculated values were also cross-checked by making manual measurements to ensure the correct functionality of the software. Measurements of slice position and warp were excluded from automatic analysis. The signal values for T_1 and T_2 measurements were extracted automatically, but the relaxation times were calculated by fitting the signal values into signal equations using the Solver utility of Microsoft Excel.

Software that was produced by Biomedical Informatics Research Network (BIRN) was used (Friedman and Glover 2006) in Paper III. The software automatically calculated values of the following parameters from the image series: signal fluctuation and drift, SNR and signal-to-fluctuation noise ratio, radius of decorrelation, the maximum displacement and drift of image center-of-mass in three directions, smoothness, and ghosting. A report and a set of images were produced by the software.

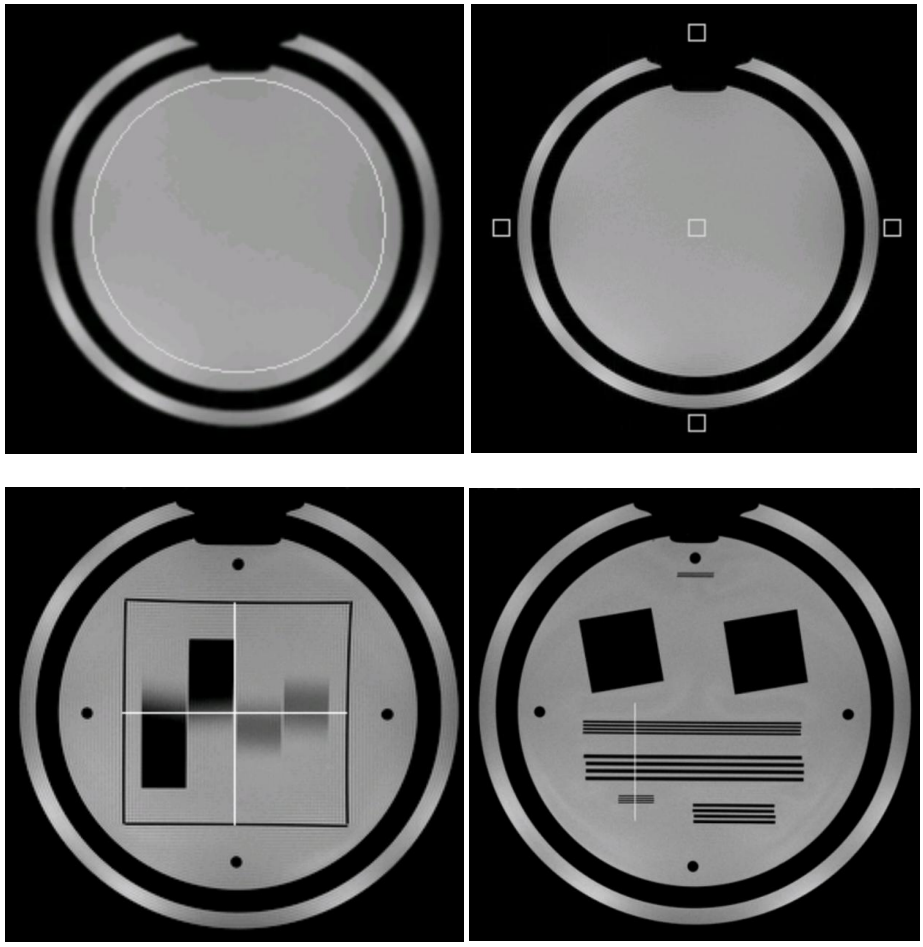


Figure 4 *ROIs and measurement lines selected by the automated analysis software used in Paper I. The selections were used to evaluate image uniformity and SNR (top left), ghosting (top right), geometric distortion (bottom left) and spatial resolution (bottom right).*

6 Results

Two studies with extensive sets of measurements of the image quality of structural MRI were carried out with two different phantom protocols. The majorities of the measurement findings in both studies were in accordance with typical or recommended values.

The results of eleven measured parameters in the first study (Paper I) were compared with recommended values of the Eurospin protocol or other published values. The sources of selected reference values and also detailed results for individual parameters can be found in Paper I. The results are summarized in Table 2 in this dissertation. In general, the results were repeatable within 5%, and the automatic analysis software produced similar results to those of manual measurements. The measurements also revealed interesting incidental observations. For example, a notable geometric distortion was observed in the sagittal and coronal planes in one MRI system in the first study. The distortion occurred only at the edges of field-of-view (FOV) and did not affect the geometric accuracy measurement. The distortion was attributed to the system and was not known to its users, thus a correction filter was subsequently applied into the clinical protocols to remove the distortion.

Table 2 Summary of the results of the Eurospin phantom measurements. SNR = signal-to-noise ratio, N/A = not available

Parameter	Results	Recommended values
Uniformity	Majority of the results 85–95% in transversal plane, 60–85% in the sagittal and coronal planes	>80%
SNR	Approximately proportional to field strength, unexpectedly high values on one system	N/A
SNR uniformity	±5% on the transversal plane, larger variation on other planes	N/A
Ghosting	<1%, with one exception (4%)	<3%
Geometric accuracy	120 mm ± 1 mm	120 mm ± 1 mm
Slice thickness	5 mm ± 0.9 mm	5 mm ± 1.0 mm
Slice position error	±2 mm on transversal plane	± 2 mm
Slice warp	None	None
Spatial resolution	0.5 mm (=pixel size)	pixel size
T_1 and T_2 accuracy	10%	N/A

Table 3 Summary of the results of the ACR phantom measurements. Some of the results are given separately for ACR and site sequences.

Parameter	Results	Acceptance criterion
Uniformity	74–88% (3.0T), 88–97% (1.5T)	82% (3.0T), 87.5% (1.5T)
Slice thickness	4.5–5.8 mm (ACR T_1-weighted, ACR T_2-weighted, Site T_1-weighted), 2.9–6.3 mm (Site T_2-weighted)	5 mm ±0.7 mm
Spatial resolution	1.0 mm for ACR sequences, pixel size dependent for Site sequences	1.0 mm (pixel size) for ACR sequences
Low contrast detectability	38–40 objects (3.0T, T_1-weighted), 25–39 objects (1.5T, T_1-weighted), 10–40 objects (3.0T, T_2-weighted), 11–38 objects (1.5T, T_2-weighted)	37 objects (3.0T), 9 objects (1.5T)
Ghosting	<2.5%; majority of the results <1%	<2.5%
Slice position error	<3.6 mm	<5 mm
Geometric accuracy	±2 mm of 190 mm, apart from two systems that failed the test in one measurement	±2 mm of 190 mm

Table 4 Passing rates (%) of the individual ACR phantom tests for the ACR and site sequences on the 1st and the 2nd round of measurements (modified from Paper II).

	ACR sequences		Site sequences	
Measurement	1st	2nd	1st	2nd
Uniformity	91	91	91	91
Slice thickness	100	91	55	55
Spatial resolution	100	82	36	36
Low contrast detectability	100	100	91	82
Ghosting	100	100	100	100
Slice position error	100	100	100	100
Geometric accuracy	100	91	100	91

The results of the second study (Paper II) were compared with given acceptance limits of the ACR measurement protocol. The results are summarized in Table 3. The passing rates of the individual tests are presented in Table 4. The criteria for acceptance of the parameters were given in the ACR phantom instructions. The test of seven parameters was passed when the performance criteria were met either for the ACR or the site sequences. As much as 91%

of the systems passed the overall test in the first measurements and 73% in the second measurements.

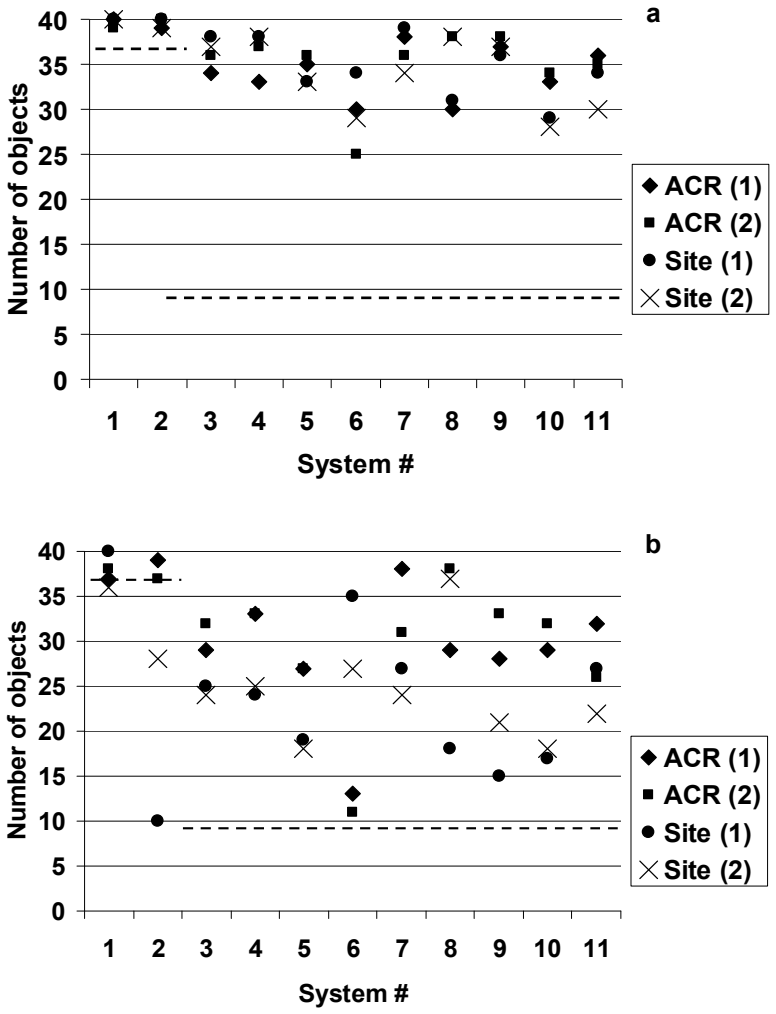


Figure 5 *Low contrast detectability of the ACR phantom measurements 1 and 2 for a) T_1-weighted sequences and b) T_2-weighted sequences, as presented in Paper II. The dashed lines indicate the ACR recommended acceptance values for 1.5T (9 objects) and 3.0T (37 objects). The system numbering in this figure is not consistent with Table I.*

There were notable differences in T_1- and T_2-weighted sequences used in routine head imaging. The differences caused variation in the ACR phantom results, which also affected the passing rates of the tests. The variation was present especially in the results of low contrast detectability, presented in Fig. 5.

The fMRI study (Paper III) included measurements of 5 MRI systems. The temporal stability parameters were in accordance with typical values that had been reported previously (Friedman and Glover 2010). Apart from 4 measurements, the temporal fluctuation was <0.2% for all the measurements. The signal drift was <1.0% in all the measurements. Maximum center of mass displacement varied from 0 to 0.4 mm (mean 0.09 mm). Again an incidental finding was obtained: an abrupt but small movement of the image center of mass a few times during the time series on the MRI systems of one vendor. The use of coil intensity normalization led to a reduction of 8% for SNR on one 3.0T system.

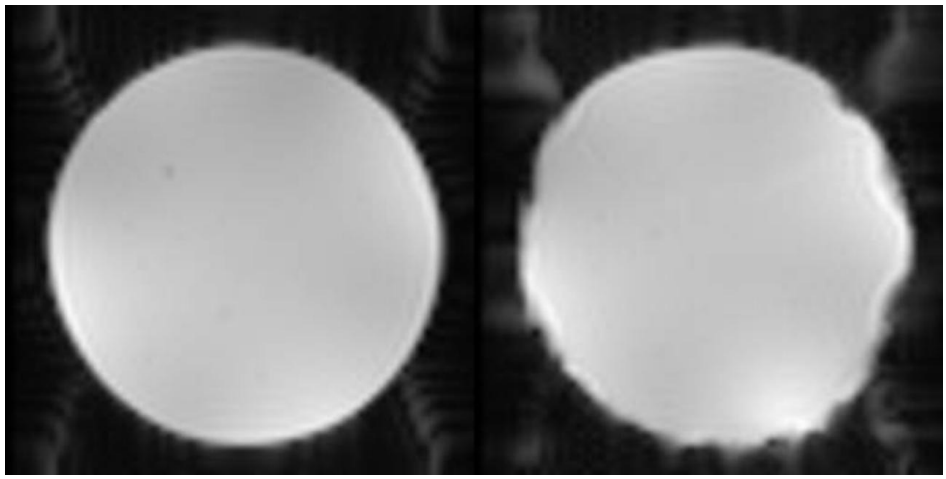


Figure 6 *Images representing the mean pixel values of an fMRI series, without (left) and with (right) the EEG cap that generated superficial artifacts (Paper III).*

Image artifacts (Fig. 6) and SNR reduction in the fMRI data were generated when the EEG equipment were used. The electrocardiography (ECG) and electrooculography (EOG) electrodes were observed to cause strong artifacts, the total signal dropout extended to 25 mm and effects on signal intensity extended to 50 mm from the surface. The typical artifact depth caused by the EEG electrodes was 15 mm or less. The artifacts were not visible in the images of the two human volunteers. When the EEG equipment was used, the average SNR reduction was 15% in the phantom studies, and 18% and 30% in the volunteer studies using medial and lateral ROIs, respectively. No effects on temporal stability nor any other image quality parameters due to EEG equipment were observed.

The EEG signal was also recorded during image acquisition. The noise was approximately five times as high inside the MRI room as outside the room. An interference signal was observed for one MRI system, and it disappeared when the helium pump was switched off. A difference was found in noise properties of two amplifiers that had identical

technical properties (Fig. 7). The difference did not depend on the position of the two amplifiers inside the MRI system bore.

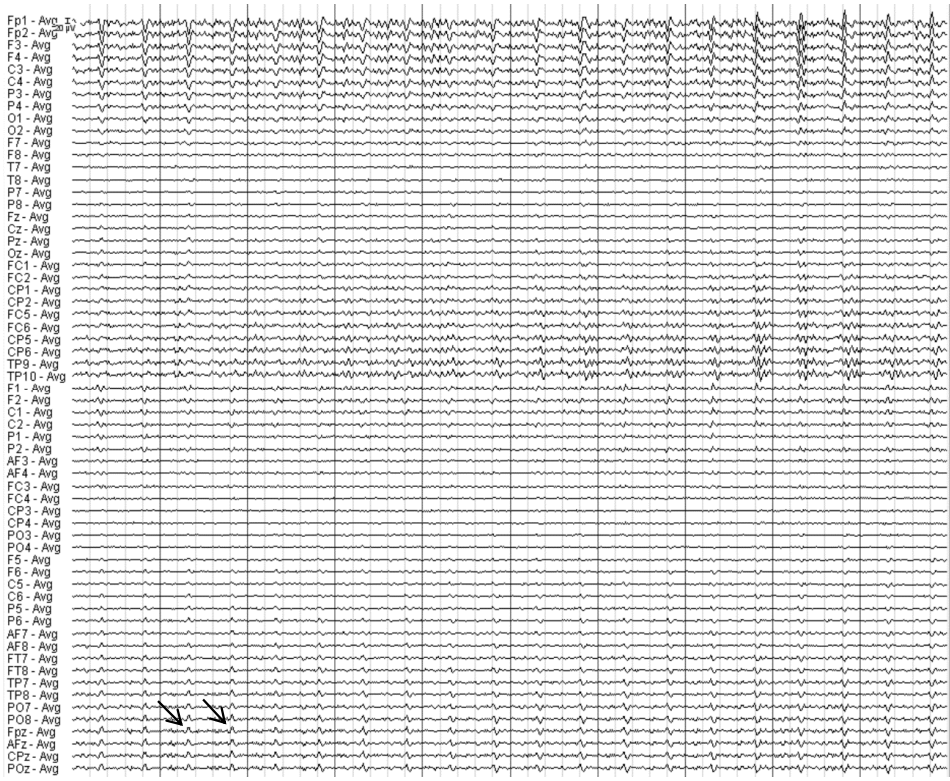


Figure 7 *An example of an EEG measurement. The EEG cap was placed on top of the phantom, which was positioned in the MRI system bore. The cap was connected to two amplifiers that were also positioned in the bore. The MRI sequence was not switched on. Larger noise fluctuations can be seen on measurement channels connected to the first amplifier (upper half) than for the second amplifier (lower half). A 2 Hz interference pattern that originated from the helium pump of the MRI system can also be seen (arrows).*

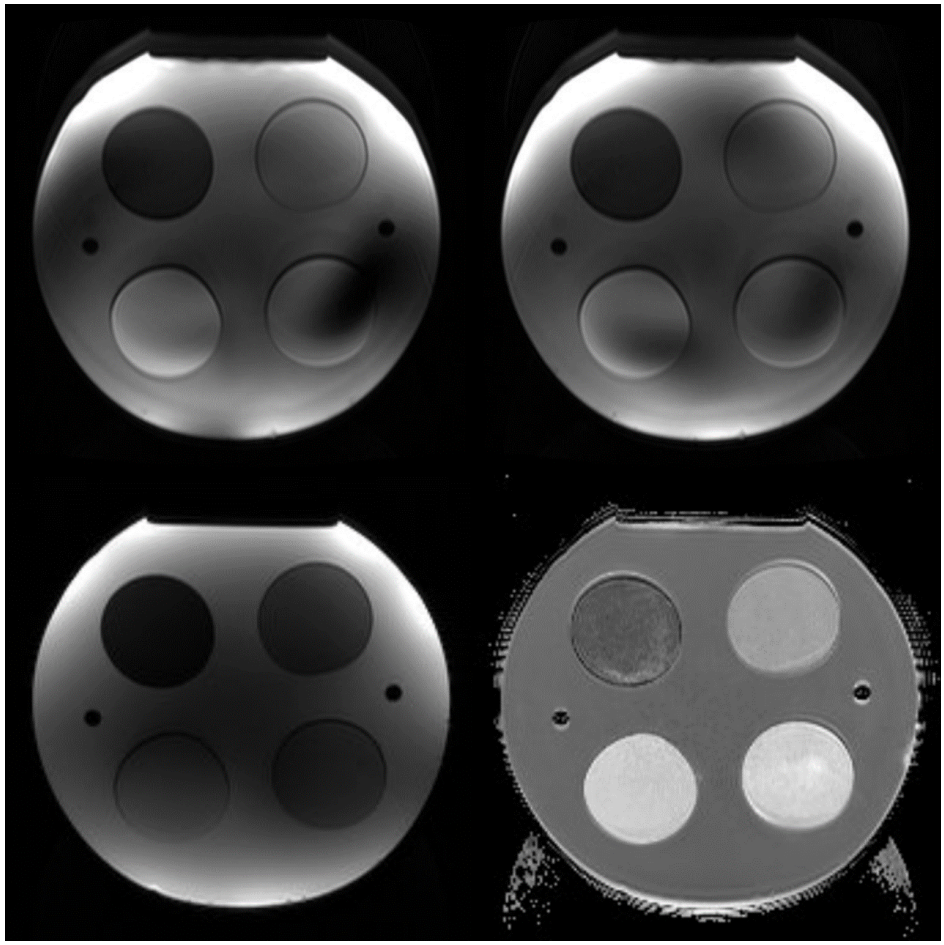


Figure 8 *Images of the diffusion phantom (Paper IV). Top row: trace-weighted images without (left) and with (right) the patient specific B_1 shim at 3.0T. Bottom row: a trace-weighted image (left) and an ADC map (right) acquired with readout-segmented EPI method at 1.5T.*

As in the case of the fMRI measurements, the question of image artifacts was a central issue regarding the DWI measurements with a body-sized phantom (Paper IV). The measurements clearly demonstrated the artifacts characteristic of the EPI sequence, of the parallel imaging, and of the B_1 inhomogeneity. When a conventional abdominal DWI series was given an artifact level index value of 100, the lowest relative artifact level index was calculated for readout-segmented EPI (64), which was followed by zoomed EPI (78) and by patient-specific B_1 shim option (89), which reduced the strong B_1 inhomogeneity artifact at 3.0T (see Fig. 8). The acquisition options also had an effect on contrast. Although the coil

intensity normalization option did not have a relevant effect on the ADC map, it changed the signal of individual objects notably in the trace-weighted images.

Uniformity of ADC values between several acquisition options were evaluated. Higher ADC values (by 0.07 x10^{-3} mm^2/s on average) were observed with a readout-segmented EPI than with the conventional clinical technique. A difference in the ADC values (0.05 x10^{-3} mm^2/s on average) was observed between 1.5T and 3.0T systems. ADC variations of approximately 0.1 x10^{-3} mm^2/s in slice direction were observed (Fig. 9).

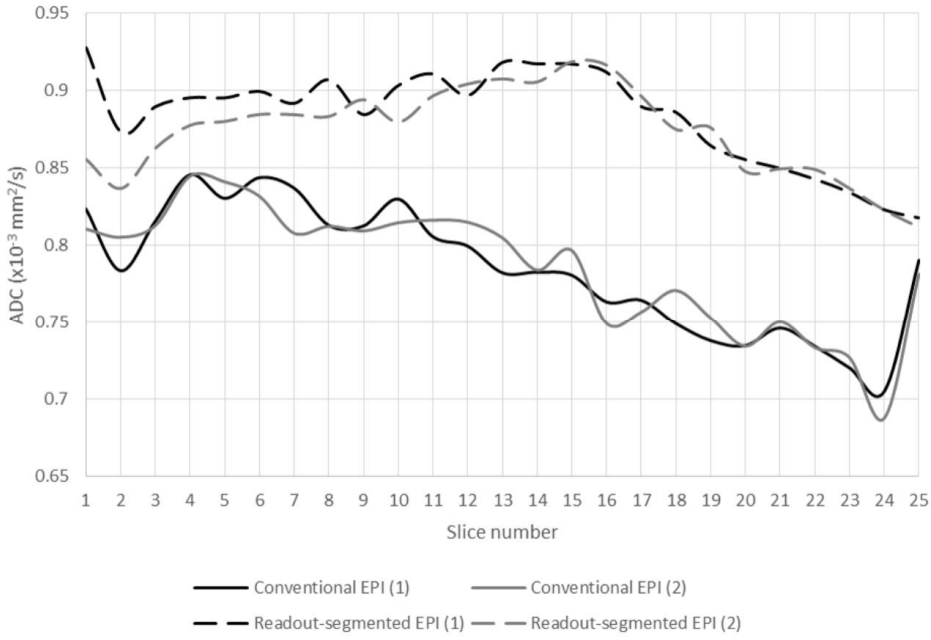


Figure 9 *The ADC values for different slices of two measurements of conventional clinical series and readout-segmented EPI series at 1.5T, as presented in Paper IV. The values in first and last slices have been affected by the edges of the phantom.*

7 Discussion

Despite the technological challenges of dealing with potential instabilities, the MRI systems were usually able to produce signals that were acceptable and repeatable in this study. Indeed a result that exceeded the tolerances was an exception, not a rule. Inferior image quality poses a relevant risk to patients and also causes problems in creating delays in the processing of patients. The QC measurements presented in this study are methods that can mainly be used in annual or acceptance tests that are carried out by medical physicists or MR scientists. It is also customary and recommended to control the short-term repeatability daily or weekly and regularly test the receiver coils (Firbank et al. 2000, Peltonen et al. 2014, American College of Radiology 2015).

7.1 Observations about the results

The Eurospin phantoms and also the ACR phantom enabled the evaluation of different aspects of the image quality in the structural MRI. The Eurospin phantoms did not enable a direct measurement of low contrast detectability, which is one of the essential parameters from the clinical point of view. The ACR phantom protocol only included measurements for the transversal plane, and the protocol did not contain a measurement of SNR. However, measurement of low contrast detectability directly reflects the SNR of the system.

The results of image uniformity were affected by coil sensitivity profiles in the sagittal and coronal planes of Eurospin phantom measurements. The results were also probably affected by coil structure, reconstruction methods, image correction filters, and nonuniform B_1 fields in the 3.0T systems. The results of distance measurements used in geometic accuracy evaluation were acceptable, but the distances were short (120 mm in Eurospin phantoms and 190 mm in ACR phantom) compared with full FOV allowed by the MRI system (up to 500 mm in all directions, dependent on the system). Some of the methods to analyze the phantom images, such as the slice thickness measurement of the ACR phantom, were open to interpretation and thus produced uncertainty in the results. Occasionally the analysis methods themselves also have to be re-evaluated as a whole as technologies evolve. An example of this evolution is the SNR measurement, which had been obtained by using two separate regions of a single image and was used in the Eurospin phantom measurements. This method was valid in the era of single-channel receiver coils. Modern MRI systems, however, use multi-channel coils, parallel imaging and different reconstruction filters that all affect the conventional approach of measuring SNR. The noise should be calculated from pixel-by-pixel SD in multiple repeated acquisitions, or from a difference image (Dietrich et al. 2007).

Both types of evaluated structural MRI quality control methods were internationally acknowledged: the Eurospin phantoms and methods were based on an EC consortium study, and the ACR phantom and measurement protocol were a part of the ACR MRI accreditation

program. By the time the study described in Paper II was implemented, the quality control protocol that had been created for Paper I had already been replaced with MagNET phantoms that closely resembled the Eurospin phantoms in phantom structure and analysis. The introduction of a more time-efficient protocol with the multi-purpose ACR phantom was one of the motivations of this part of the study. The Eurospin and subsequently the MagNET phantoms had long measurement histories in our organization. It was possible to carry out measurements in three orthogonal slice planes (transversal, sagittal, and coronal). Automatic software was used to analyse the images. That software had been originally developed for the first study and then further developed and reprogrammed with Matlab thereafter. However, the measurements of five Eurospin phantoms for the three orthogonal slice planes with instructed spin-echo sequences were time consuming (2-3 hours). A shorter acquisition time (20 min) and the inclusion of clinical sequences in the protocol were considered to be advantages of using the ACR phantom based method. In addition, the ACR phantom had probably become globally more widespread than other phantom sets by the time of the publication of Paper II.

The fMRI systems investigated in this study were concluded to operate at an acceptable level of stability. Changes in image center of mass on the MRI systems for one vendor were considered small compared to voxel size or physiological movement and were probably not able to alter the results. The observed superficial artifacts in EEG–fMRI were comparable with those reported by previous studies (Krakow et al. 2000, Mullinger et al. 2008). The ECG and EOG electrodes produced strong artifacts, thus we recommended positioning them outside the proximity of the FOV for a phantom test, as would be done in patient studies. The SNR reduction due to EEG equipment may restrict the use of the imaging sequences in which the SNR is critical. SNR reduction of only 8% was observed in a previous study (Luo and Glover 2012), but their study and current study were not entirely comparable due to different imaging protocols and voxel sizes used. The type of EEG cap and gel may also have affected the SNR results (Bonmassar et al. 2001, Negishi et al. 2008). The reduction in SNR had probably originated from increased impedance of the coil when the EEG equipment was present (Carmichael 2010). We suggested in Paper III that the test should be run whenever new EEG equipment is introduced, and acceptance limits for artifact depth and SNR reduction could be set to 20 mm and 20%, respectively. The findings also supported the expedient of switching off the helium pump of the MRI system during EEG–fMRI, as also shown in a previous study (Nierhaus et al. 2013).

The results of the DWI phantom study (Paper IV) showed agreement between the ADC values of the different acquisition options. The observed ADC difference between 1.5T and 3.0T systems, and the ADC variation along the slice direction were in line with observations of previous studies (Lavdas et al. 2014, Malyarenko et al. 2013). Our experiences of ADC measurements with the body-sized phantom allow us to make the following observations: 1) the samples should be large enough to enable artifact-free ROIs, 2) temperature of the phantom should be known, 3) possibility of off-center variation should be taken into account, 4) available options should be used to make the B_1 field more uniform, and 5) the

ADC values should be measured by using different DWI methods, such as readout-segmented EPI, if applicable.

The diffusion-weighted images are important in diagnosis in addition to the ADC maps. The use of coil intensity normalization should be standardized in clinical protocols due to its possible effects on contrast in these images. The new techniques that can be used in DWI (readout-segmented EPI, zoomed EPI, patient-specific B_1 shim) showed fewer artifacts than conventional methods, which is encouraging for increasing the use of these approaches in clinical settings.

7.2 Limitations of the study

A central limitation of this work was the inadequacy in the number of repeated measurements for the reliable analyzes of the repeatability of the results. Repeated measurements were carried out in short and long term in Papers I–III. However, more measurement data would be needed for the reliable definition of tolerances for the QC tests.

The stability of agar gel phantoms is unclear (Tofts 2003). The agar gel phantom and samples in Papers III and IV were self-prepared and their possible instability may have caused bias in the results. This would be especially the case in Paper III in which the phantom was used for a long period of time. However, similar phantoms have been used in the BIRN project for comparable time frames and Friedman and Glover (Friedman and Glover 2010) did not report problems in phantom stability. Lavdas et al. reported stable ADC values and a drop in T_2 of about 1 ms a week during a time frame of nine weeks for similar gel compositions to those used in Paper IV (Lavdas et al. 2013). Probably the air-tight sealing of the phantoms minimizes the evaporation of water from the gels and maximizes their stability.

The ADC measurement in DWI is sensitive to temperature. The temperature measurement reported in Paper IV did not take into account the possibility of phantom warming in the MRI system bore. In addition, the measured ADC values from the phantom samples differed slightly from the intended values. The inclusion of a bottle of pure water as a reference standard and the direct measurement of the temperature of the phantom would increase the reliability of the measurements described in Paper IV.

The measurements of the accuracy of relaxation times in Paper I had several limitations. The number of measured time points was limited to only five, the shortest echo time for T_2 measurements was quite long (20 ms) and the accuracy of the T_1 and T_2 calculations was not evaluated. Moreover, the measurements in clinical and research imaging are carried out using multi-slice multi-echo sequences instead of repeated single-slice spin-echo measurements, thus the resulting T_2 values may differ. The increasing role of quantitative imaging may increase the significance of this measurement in the future.

7.3 General observations and insights

Although difficult to accomplish in practice, MRI allows quantitative measurements to be made. There should be a method to ensure that the measurement gives a correct result for every quantitative MRI measurement taken. For example, the use of ADC values in diagnostics has been limited, partly because those values have generally been considered to be vendor dependent (Sasaki et al. 2008). The computed tomography systems by comparison are universally calibrated to give the same values in Hounsfield units for each material imaged. In addition, QC methods are more standardized and regulated in imaging methods that use ionizing radiation.

The value of standardized QC protocols was demonstrated in practice by the results of the ACR phantom study (Paper II) in which standard and site protocols were measured and compared. There was notably more variation in the results of site protocols than for the standard protocol. We faced difficulties in standardizing the QC protocols between different MRI systems in this work. The test objects did not always fit inside the clinically relevant head coils, thus either the positioning of the phantom in the center of the coil was compromised, or an alternative head coil had to be used. The number of sequence parameters, filters and other options has been growing throughout the years. The vendors have different names and implementations for different options, and they provide different possibilities to adjust the imaging parameters. In practice these differences in image acquisition could not be fully removed. All relevant parameters such as receiver bandwidth had not been taken into account in published QC protocols. Obviously, these difficulties are also relevant to clinical work. In addition to the afore mentioned issues, the differences in site protocols of routine head imaging (Paper II) arose from the following: the incoherent selection of clinical sequences for our measurements, differences of the MRI systems, nonstandardized imaging practices and different patient materials. MRI could benefit from a process similar to clinical audit (European Commission 2009) or a practice of clinical image quality evaluation (European Commission 1996) in standardization of QC and imaging practices.

The automated analyses of QC images were successfully applied in this work, in addition to many other published studies. The use of automated methods has saved time (Bourel et al. 1999) and automation in fMRI is essential because of the large amount of data generated. Automated QC methods can probably detect degradation of image quality before the system users notice it in the clinical images, as was demonstrated by Gardner et al. (Gardner et al. 1995). The exacting requirement for accurate phantom positioning was also essential in our studies that used automatic analysis methods, especially in Paper I. The uniformity of methods in the single organization where this work was carried out has made it possible to introduce automated and centralized processing of the QC images, follow-up of the results, and introduction of appropriate actions for dealing with inferior results. These actions have facilitated the management of QA in a multi-unit and geographically scattered organization. A follow-up method such as Shewhart charting as suggested by Simmons et al. could be used in this kind of setting (Simmons et al. 1999). Automated analysis has been developed

for the ACR phantom using the measurement data obtained for Paper II, and that software has been used in the QC program of our organization (Mäkelä et al. 2012). Other groups have also implemented automated analyses for the ACR phantom data (Davids et al. 2014).

The purpose of clinical MRI quality control is to serve and support clinical imaging. The research for this dissertation has shown that appropriate QC tools were able to reveal image quality issues that had not been obvious in the patient images, such as geometric distortions in Paper I, and the image center-of-mass transitions in Paper III. Furthermore, results of the EEG–fMRI and DWI measurements supported the importance of having data quality evaluation tools available whenever a new imaging technique is introduced into clinical use. It is possible with the help of the QC tools developed in this study to estimate the effects of SNR reduction and superficial artifacts on image quality in EEG–fMRI in addition to the increasing level of confidence in the clinical use of new applications of DWI. However, there were also limitations in the QC methods in relation to clinical imaging, which raised questions throughout the study. The spin echo sequences in QC protocols of structural imaging may not reflect clinical imaging, in which mostly turbo spin echo and gradient echo sequences are used. A successful fat suppression is often a critical part of a clinical study, and fat suppression is often degraded for technical reasons. However, it was not feasible to evaluate fat suppression with the phantoms and protocols used in this study. We were able to evaluate stability parameters of EPI, but not the BOLD contrast itself in fMRI. There have been some solutions devised for mimicking the BOLD contrast with phantoms (Renvall et al. 2006), but the adoption of these methods is not yet widespread in clinical sites. We were not able to include important effects such as microcirculation in the test object of the DWI study. However, we were able to mimic the contrast of DWI by using different gels and, along with the limitations of the phantom compared to human imaging, we were able to measure clinically relevant differences in image quality between different methods.

Time management is often critical in clinical work. MRI systems are expensive, imaging times are long and demand for the MRI exams are constantly increasing. The extent to which the MRI systems are currently being used is high. Time consuming QC tests with no obvious clinical input cannot be accepted under these circumstances. The purpose of QC is to detect system faults and degrading image quality sufficiently early, and the results are expected to lead to remedial actions that fully restore the state of the system. Measures for increasing time effectiveness and that were applied in this work include the use of a single multi-purpose phantom instead of multiple phantoms, and the use of automated analysis. Another measure is the division of the tests to necessary short-term (daily) and more comprehensive long-term (annual) tests.

The development of the MRI technology during the period the research for this thesis was being conducted has been rapid, and this development has also affected imaging practices (and vice versa). A large number of significant or small improvements have been introduced. Up-to-date QC methods are needed to ensure good accuracy and precision of the results. The increased number of permutations for building an MRI protocol for a certain

clinical or scientific question can also be a complicating factor in an imaging organization that aims for having reasonably uniform imaging results between MRI systems, or in multi-center studies. The lack of standardization of the imaging protocols has been recognized as a problem in several areas, such as for DWI in oncology (Padhani et al. 2009). Development of new quality control methods is needed for the standardization work, especially in quantitative measurements. An effective MRI QC can support clinical imaging, if clinically relevant parameters are measured. Such a QC method will not use too much system time, and will be able to detect inferior performance of the systems.

8 Conclusions

Novel methods for objective and vendor-independent QC of MRI systems were developed, adapted and used in a multi-unit medical imaging organization in this research. The developed methods included automated analysis software for structural MRI QC images (Paper I), a protocol for QC of simultaneous EEG–fMRI (Paper III) and a body-sized phantom for diffusion measurements (Paper IV). Feasibility of a single multi-purpose phantom in the QC of structural MRI was studied (Paper II). At the end of the study the following conclusions were made:

1) The QC protocols for two internationally acknowledged test object methods for structural MRI and test objects for functional and diffusion-weighted body MRI were generated and measured in a single organization that has multiple MRI systems. These QC methods allowed a comparison of the systems with the reported limitations.

2) A majority of MRI QC results in this study were close to the recommended or to typical values and were therefore acceptable. However, the methods developed and used were also capable of detecting inferior image quality when it arose.

3) Appropriate image and data quality assessment tools revealed clinically relevant information about emerging MRI techniques, such as the reduction in SNR for simultaneous EEG–fMRI and the level of artifacts for different acquisition options of diffusion-weighted MRI of the body.

4) Automated analysis software was developed for QC of structural imaging, and its correct functionality was verified after being compared with manual measurements. Another automated analysis method was successfully adapted for fMRI and for simultaneous EEG–fMRI measurements on different systems.

The clinical relevance of phantom-based QC could be further improved in both structural and functional MRI in future research and development.

References

Allen PJ, Josephs O, Turner R (2000) A method for removing imaging artifact from continuous EEG recorded during functional MRI. NeuroImage 12:230-239

American College of Radiology (2001) Magnetic resonance imaging quality control manual. Reston, VA, USA

American College of Radiology (2005) Phantom test guidance for the ACR MRI accreditation program. Reston, VA, USA.

American College of Radiology (2013) MRI Accreditation program. Clinical image quality guide.
www.acr.org/~/media/ACR/Documents/Accreditation/MRI/ClinicalGuide.pdf, accessed on October 11th 2015

American College of Radiology (2015) MRI accreditation program requirements. www.acr.org/~/media/ACR/Documents/Accreditation/MRI/Requirements.pdf, accessed on August 6th 2015

Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34: 537-541

Bonmassar G, Hadjikhani N, Ives JR, Hinton D, Belliveau JW (2001) Influence of EEG electrodes on the BOLD fMRI signal. Hum Brain Mapp 14: 108-115

Bourel P, Gibon D, Coste E, Daanen V, Rousseau J (1999) Automatic quality assessment protocol for MRI equipment. Med Phys 26: 2693-2700

Bushberg JT, Seibert JA, Leidholdt EM, Boone JM (2002) The Essential Physics of Medical Imaging, 2nd ed. Lippincott Williams & Wilkins, Philadelphia, USA

Carmichael D (2010) Image quality issues. In: Mulert C, Lemieux L (eds) EEG–fMRI. Springer, Berlin, pp 173-199

Carmichael DW, Vulliemoz S, Rodionov R, Thornton JS, McEvoy AW, Lemieux L (2012) Simultaneous intracranial EEG–fMRI in humans: protocol considerations and data quality. NeuroImage 63: 301-309

de Certaines JD, Cathelineau G (2001) Safety aspects and quality assessment in MRI and MRS: a challenge for health care systems in Europe. J Magn Reson Imaging 13: 632-638

Chenevert TL, Galbán CJ, Ivancevic MK, Rohrer SE, Londy FJ, Kwee TC, Meyer CR, Johnson TD, Rehemtulla A, Ross BD (2011) Diffusion coefficient measurement using a temperature-controlled fluid for quality control in multicenter studies. J Magn Reson Imaging 34: 983-987

Covell MM, Hearshen DO, Carson PL, Chenevert TP, Shreve P, Aisen AM, Bookstein FL, Murphy BW, Martel W (1986) Automated analysis of multiple performance characteristics in magnetic resonance imaging systems. Med Phys 13: 815-823.

Davids M, Zöllner FG, Ruttorf M, Nees F, Flor H, Schumann G, Schad LR (2014). Fully-automated quality assurance in multi-center studies using MRI phantom measurements. Magn Reson Imaging 32: 771-780

Delakis I, Moore EM, Leach MO, De Wilde JP (2004) Developing a quality control protocol for diffusion imaging on a clinical MRI system. Phys Med Biol 49: 1409-1422.

Delakis I, Xanthis C, Kitney RI (2009) Assessment of the limiting spatial resolution of an MRI scanner by direct analysis of the edge spread function. Med Phys 36: 1637-1642

De Wilde J, Price D, Curran J, Williams J, Kitney R (2002) Standardization of performance evaluation in MRI: 13 years' experience of intersystem comparison. Concepts Magn Reson 15: 111-116

Diagnostic Sonar Ltd (1992-1995) Eurospin II magnetic resonance quality assessment test objects. Instructions for use, Livingston, Scotland

Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO (2007) Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. J Magn Reson Imaging 26: 375-385

European Commission (1996) European guidelines on quality criteria for diagnostic radiographic images. Report EUR 16260 EN. Luxembourg; Office for official publications of the European Communities

European Commission (2009) Radiation protection No. 159. European Commission guidelines on clinical audit for medical radiological practices (diagnostic radiology, nuclear medicine and radiotherapy). Luxembourg: Publications Office of the European Union

Firbank MJ, Harrison RM, Williams ED, Coulthard A (2000) Quality assurance for MRI: practical experience. Br J Radiol 73: 376-383

Friedman L, Glover GH (2006) Report on a multicenter fMRI quality assurance protocol. J Magn Reson Imaging 23: 827-839

Gardner EA, Ellis JH, Hyde RJ, Aisen AM, Quint DJ, Carson PL (1995) Detection of degradation of magnetic resonance (MR) images: comparison of an automated MR image-quality analysis system with trained human observers. Acad Radiol 2: 277-281

Goldman RI, Stern JM, Engel J, Cohen MS (2000) Acquiring simultaneous EEG and functional MRI. Clin Neurophysiol 111: 1974-1980

Gotman J, Kobayashi E, Bagshaw AP, Bénar CG, Dubeau F (2006) Combining EEG and fMRI: A multimodal tool for epilepsy research. J Magn Reson Imaging 23: 906-920

Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47: 1202-1210

Griswold M, Walsh D, Heidemann R, Haase A, Jakob P (2002b) The use of an adaptive reconstruction for array coil sensitivity mapping and intensity normalization. Proc Intl Soc Magn Reson Med, Honolulu, Hawaii, 2410

Gunter JL, Bernstein MA, Borowski BJ, Ward CP, Britson PJ, Felmlee JP, Schuff N, Weiner M, Jack CR (2009) Measurement of MRI scanner performance with the ADNI phantom. Med Phys 36: 2193-2205.

Haacke EM, Brown RW, Thompson MR, Venkatesan R (1999) Magnetic Resonance Imaging. Physical Principles and Sequence Design. John Wiley & Sons, New York, US

Hahn EL (1950) Spin echoes. Phys Rev 80: 580-594

Hough PVC (1962) Methods and means for recognizing complex patterns. US Patent 3 069 654

Kanal E, Borgstede JP, Barkovich AJ, Bell C, Bradley WG, Felmlee JP, Froelich JW, Kaminski EM, Keeler EK, Lester JW, Scoumis EA, Zaremba LA, Zinninger MD (2002) American College of Radiology white paper on MR safety. Am J Roentgenol 178: 1335-1347

Kapanen M, Collan J, Beule A, Seppälä T, Saarilahti K, Tenhunen M (2013) Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate. Magn Reson Med 70: 127-135

Katscher U, Börnert P (2006) Parallel RF transmission in MRI. NMR Biomed 19: 393-400

Koller CJ, Eatough JP, Mountford PJ, Frain G (2006) A survey of MRI quality assurance programmes. Br J Radiol 79: 592-596

Krakow K, Allen PJ, Symms MR, Lemieux L, Josephs O, Fish DR (2000) EEG recording during fMRI experiments: image quality. Hum Brain Mapp 10: 10-15

Laubach HJ, Jakob PM, Loevblad KO, Baird AE, Bovo MP, Edelman RR, Warach S (1998) A phantom for diffusion-weighted imaging of acute stroke. J Magn Reson Imaging 8: 1349-1354.

Lauterbur PC (1973) Image formation by induced local interactions: examples employing nuclear magnetic resonance. Nature 242: 190-191

Lavdas I, Behan KC, Papadaki A, McRobbie DW, Aboagye EO (2013) A phantom for diffusion-weighted MRI (DW-MRI). J Magn Reson Imaging 38: 173-179

Lavdas I, Miquel ME, McRobbie DW, Aboagye EO (2014) Comparison between diffusion-weighted MRI (DW-MRI) at 1.5 and 3 tesla: a phantom study. J Magn Reson Imaging 40: 682-690

Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M (1986) MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology 161: 401-407

Le Bihan D, Delannoy J, Levin RL (1989) Temperature mapping with MR imaging of molecular diffusion: application to hyperthermia. Radiology 171: 853-857

Lemieux L, Allen PJ, Krakow K, Symms MR, Fish DR (1999) Methodological issues in EEG-correlated functional MRI experiments. Int J Bioelectromagn 1: 87-95

Lemieux L, Krakow K, Fish DR (2001) Comparison of spike-triggered functional MRI BOLD activation and EEG dipole model localization. NeuroImage 14: 1097-1104

Lerski RA, de Certaines JD (1993) Performance assessment and quality control in MRI by Eurospin test objects and protocols. Magn Reson Imaging 11: 817-833

Lerski RA (1993) Trial of modifications to Eurospin MRI test objects. Magn Reson Imaging 11: 835-839

Logothetis N (2009) What we can do and what we cannot do with fMRI. Nature 453: 869-878.

Luo Q, Glover GH (2012) Influence of dense-array EEG cap on fMRI signal. Magn Reson Med 68: 807-815

Malyarenko D, Galbán CJ, Londy FJ, Meyer CR, Johnson TD, Rehemtulla A, Ross BD, Chenevert TL (2013) Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. J Magn Reson Imaging 37: 1238-1246.

Mansfield P and Grannell PK (1973) NMR 'diffraction' in solids? J Phys C 6: L422-L426

Mansfield P (1977) Multi-planar image formation using NMR spin echoes. J Phys C 10: L55-L58

Martin CJ, Sharp PF, Sutton DG (1999) Measurement of image quality in diagnostic radiology. Appl Radiat Isot 50: 21-38

Matsuya R, Kuroda M, Matsumoto Y, Kato H, Matsuzaki H, Asaumi J, Murakami J, Katashima K, Ashida M, Sasaki T, Sei T, Himei K, Katsui K, Katayama N, Takemoto M, Kanazawa S, Mimura S, Oono S, Kitayama T, Tahara S, Inamura K (2009) A new phantom using polyethylene glycol as an apparent diffusion coefficient standard for MR imaging. Int J Oncol 35: 893-900

Mattila S, Renvall V, Hiltunen J, Kirven D, Sepponen R, Hari R, Tarkiainen A (2007). Phantom-based evaluation of geometric distortions in functional magnetic resonance and diffusion tensor imaging. Magn Reson Med 57: 754-763

McRobbie DW, Moore EA, Graves MJ and Prince MR (2005) MRI from Picture to Proton, 3rd ed. Cambridge University Press, Cambridge, UK

Merriam-Webster (2015), www.merriam-webster.com, accessed on 13th June 2015

Miquel ME, Scott AD, Macdougall ND, Boubertakh R, Bharwani N, Rockall AG (2012) In vitro and in vivo repeatability of abdominal diffusion-weighted MRI. Br J Radiol 85: 1507-1512.

Mullinger K, Debener S, Coxon R, Bowtell R (2008) Effects of simultaneous EEG recording on MRI data quality at 1.5, 3 and 7 tesla. Int J Psychophysiol 67:178-188

Mäkelä T, Ihalainen T, Sipilä O (2012) Automated analysis of ACR phantom images. ESMRMB 2012 Congress Book of Abstracts, Lisbon, Portugal, 620-621

National Electrical Manufacturers Association (1988) Determination of signal-to-noise ratio (SNR) in diagnostic magnetic resonance images. Washington, DC, USA

Negishi M, Abildgaard M, Laufer I, Nixon T, Constable RT (2008) An EEG (electroencephalogram) recording system with carbon wire electrodes for simultaneous EEG–fMRI (functional magnetic resonance imaging) recording. J Neurosci Methods 173: 99–107

Nierhaus T, Gundlach C, Goltz D, Thiel SD, Pleger B, Villringer A (2013) Internal ventilation system of MR scanners induces specific EEG artifact during simultaneous EEG-fMRI. NeuroImage 74: 70-76

Och JG, Clarke GD, Sobol WT, Rosen CW, Mun SK (1992) Acceptance testing of magnetic resonance imaging systems: report of AAPM Nuclear Magnetic Resonance Task Group No. 6. Med Phys 19: 217-229

OECD (2013) Health at a Glance 2013: OECD Indicators, OECD Publishing

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci USA 89: 5951-5955

Olsrud J, Nilsson A, Mannfolk P, Waites A, Ståhlberg F (2008) A two-compartment gel phantom for optimization and quality assurance in clinical BOLD fMRI. Magn Reson Imaging 26: 279-286

Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, Dzik-Jurasz A, Ross BD, Van Cauteren M, Collins D, Hammoud DA, Rustin GJ, Taouli B, Choyke PL (2009) Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia 11: 102-125.

Padhani AR, Koh DM, Collins DJ (2011) Whole-body diffusion-weighted MR imaging in cancer: current status and research directions. Radiology 261: 700-718

Peltonen J, Mäkelä T, Sofiev A, Salli E (2014) Automatic daily MRI quality control analysis system applied in a large hospital district. Radiological Society of North America 2014 Scientific Assembly and Annual Meeting, Chicago, IL, USA.

Pfeuffer J, van de Moortele PF, Yacoub E, Shmuel A, Adriany G, Andersen P, Merkle H, Garwood M, Ugurbil K, Hu X (2002) Zoomed functional imaging in the human brain at 7 Tesla with simultaneous high spatial and high temporal resolution. NeuroImage 17: 272-286

Porter DA, Heidemann RM (2009) High resolution diffusion-weighted imaging using readout-segmented echo planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. Magn Reson Med 62: 468-475

Price RR, Axel L, Morgan T, Newman R, Perman W, Schneiders N, Selikson M, Wood M, Thomas SR (1990) Quality assurance methods and phantoms for magnetic resonance imaging: report of AAPM nuclear magnetic resonance Task Group No. 1. Med Phys 17: 287-295

Price RR, Allison J, Massoth RJ, Clarke GD, Drost DJ (2002) Practical aspects of functional MRI (NMR Task Group #8). Med Phys 29: 1892-1912

Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P (1999) SENSE: Sensitivity encoding for fast MRI. Magn Reson Med 42: 952-962

Purdon PL, Weisskoff RM (1998) Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. Hum Brain Mapp 6: 239-249

Ranschaert ER, Barneveld Binkhuysen FH (2013) European teleradiology now and in the future: results of an online survey. Insights Imaging 4: 93-102

Renvall V, Joensuu R, Hari R (2006) Functional phantom for fMRI: a feasibility study. Magn Reson Imaging 24: 315-320

Rieseberg S, Frahm J, Finsterbusch J (2002) Two-dimensional spatially-selective RF excitation pulses in echo-planar imaging. Magn Reson Med 47: 1186-1193

Ritter P, Becker R, Freyer F, Villringer A (2010) EEG quality: the image acquisition artefact. In: Mulert C, Lemieux L (eds) EEG– fMRI. Springer, Berlin, pp 153-171

Sasaki M, Yamada K, Watanabe Y, Matsui M, Ida M, Fujiwara S, Shibata E (2008) Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor, multi-institutional comparison study. Radiology 249: 624-630

Scarff CJ, Reynolds A, Goodyear BG, Ponton CW, Dort JC, Eggermont JJ (2004) Simultaneous 3-T fMRI and high-density recording of human auditory evoked potentials. NeuroImage 23: 1129-1142

Schaefer PW, Grant PE, Gonzales GR (2000) Diffusion-weighted MR imaging of the brain. Radiology 217: 331-345

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9: 671-675

Sigmund EE, Jensen J (2011) Basic physical principles of body diffusion-weighted MRI. In: Taouli B (Ed.) Extra-Cranial Applications of Diffusion-Weighted MRI. Cambridge University Press, Cambridge, the UK, pp 1-17

Simmons A, Moore E, and Williams SC (1999) Quality control for functional magnetic resonance imaging using automated data analysis and Shewhart charting. Magn Reson Med 41: 1274-1278

Sodickson DK, Manning WJ (1997) Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. Magn Reson Med 38: 591-603

Stejskal EO, Tanner JE (1965) Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J Chem Phys 42: 288-292

Stöcker T, Schneider F, Klein M, Habel U, Kellermann T, Zilles K, Shah NJ (2005) Automated quality assurance routines for fMRI data applied to a multicenter study. Hum Brain Mapp 25: 237-246

Thoeny HC, De Keyzer F (2007) Extracranial applications of diffusion-weighted magnetic resonance imaging. Eur Radiol 17: 1385-1393

Tofts PS, Lloyd D, Clark CA, Barker GJ, Parker GJ, McConville P, Baldock C, Pope JM (2000) Test liquids for quantitative MRI measurements of self-diffusion coefficient in vivo. Magn Reson Med 43: 368-374

Tofts P (ed.) (2003) Quantitative MRI of the Brain: Measuring Changes Caused by Disease. John Wiley & Sons, Chichester, UK

Weinreb J, Wilcox PA, Hayden J, Lewis R, Froelich J (2005) ACR MRI accreditation: yesterday, today, and tomorrow. J Am Coll Radiol 2: 494-503

Weisskoff RM (1996) Simple measurement of scanner stability for functional NMR imaging of activation in the brain. Magn Reson Med 36: 643-645

Wheeler-Kingshott CA, Barker GJ, Steens SC, van Buchem MA (2003) D: the diffusion of water. In: Tofs PS (ed.) Quantitative MR of the Brain. John Wiley & Sons, Chichester, UK, pp 203-256

Young IR (ed.) (2000) Methods in Biomedical Magnetic Resonance Imaging and Spectroscopy. John Wiley & Sons, Chichester, UK