Commissioning of MR–only simulation
for radiotherapy planning

Gerald Schubert, Teuvo Vaara, Matti Lindström,
Reko Kemppainen, and Marieke van Grootel–Rensen

When it comes to radiotherapy planning, clinical requirements for precision and accuracy of imaging and planning are extremely high. To ascertain that Philips MR–only simulation for prostate with MRCAT (Magnetic Resonance for Calculating ATtenuation) can meet these demands, medical physicists perform commissioning of MR–only simulation, prior to clinical implementation. This white paper will present a variety of methods for MRCAT commissioning reported by Philips users to help you meet your key commissioning goals: comparing the accuracy of dose delivery calculated with MRCAT against CT–based plans, and verifying proper patient positioning on the linear accelerator using MR–based image sets.

Approaches and guidelines for MRCAT commissioning

Introduction

Philips MR-only simulation is an add-on to Philips Ingenia MR-RT. This tool enables you to calculate attenuation information required for radiotherapy planning – from a single MRI scan.[1] The first step in the commissioning of MR-only simulation is execution of the recommended acceptance tests of Ingenia system and the MR-RT extension. Figure 1 provides an overview of recommended acceptance tests for each individual subsystem. Details of commissioning for Ingenia MR-RT fall beyond the scope of this white paper. Please refer to the relevant system manuals, instructions for use and white papers.

The American Association of Physicists in Medicine (AAPM) and the American College of Radiology (ACR) have issued guidelines[2][3] on how to evaluate the basic device performance and image quality of CT and MRI scanners. To date, there are no comparable commissioning guidelines for MR-only simulation available. In particular, there is no established advice on how to perform MRCAT to CT dose comparisons or to verify that the patient is correctly positioned on the linear accelerator.

Upon introducing MR-only simulation for prostate, Philips has performed performance tests in collaboration with clinical partners.[1] After commercial release, several sites have performed commissioning of MR-only simulation, and results have been published in peer-reviewed publications.[4],[5],[6],[7].

This document summarizes different approaches for MRCAT commissioning as reported by our users, with the aim of providing guidance for sites that are at the start of this process.

Evaluating the performance of MRCAT

The most straightforward way to evaluate MR-only radiation planning is to compare the dose distributions calculated from identical RT-plans (using identical gantry angles, monitoring units, etc.). The calculation is performed twice: once using MRCAT and once using CT images as the underlying density information. For the substitution of CT with MRCAT images, similar expectations on dose accuracy to those for recalculated plans with daily variations in CT scans apply: the introduced dose differences must not exceed the average difference that is observed between two fractions.

The purest approach to dosimetric agreement is to compare the obtained dose from identical plans with proper setup of the underlying images. By excluding all other contributing factors, the remaining dose difference can be traced back to the density information captured in the MRCAT and CT images. The majority of comparisons published to date adopt this approach, though they differ in their method and the metrics considered.

In MR-only based workflows, the dose is optimized using density information from MRCAT. For this reason, an additional commissioning step is required: comparison of the resulting dose for plans that have been optimized on their respective underlying density information, implying a separate optimization for MRCAT and CT.
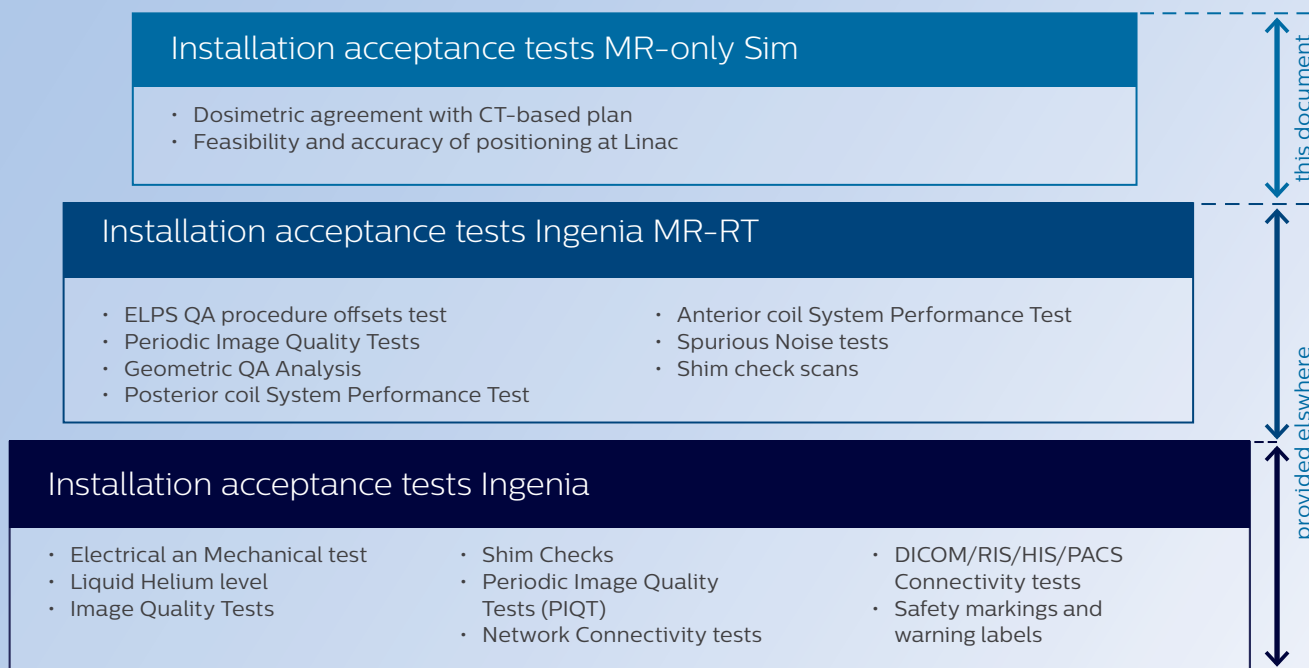

Installation acceptance tests MR-only Sim

- Dosimetric agreement with CT-based plan
- Feasibility and accuracy of positioning at Linac

Installation acceptance tests Ingenia MR-RT

- ELPS QA procedure offsets test
- Periodic Image Quality Tests
- Geometric QA Analysis
- Posterior coil System Performance Test
- Anterior coil System Performance Test
- Spurious Noise tests
- Shim check scans

Installation acceptance tests Ingenia

- Electrical an Mechanical test
- Liquid Helium level
- Image Quality Tests
- Shim Checks
- Periodic Image Quality Tests (PIQT)
- Network Connectivity tests
- DICOM/RIS/HIS/PACS Connectivity tests
- Safety markings and warning labels

this document

provided elswhere

Figure 1. Hierarchy of proposed acceptance tests for commissioning of MR-only sim.

Technical parameters and considerations

Positioning and registration

Ideally, it would be possible to achieve identical patient positioning in both the CT and MRI scanners. Transferring from one imaging modality to another, however, entails inevitable movement and there is an unavoidable time lapse between acquisitions. As a result, there may be differences in bladder filling, positioning of bowel loops and mismatched body outline contours. Therefore, the importance of good and careful patient positioning, with less than two degrees of tilting in any direction, cannot be overstated. Errors made here will not be fully recoverable in a later dose analysis.

Re-gridding

Varying voxel-sizes, relative position of origin, and rotations of the DICOM coordinate system all have an impact on the resulting dose and dose difference. To avoid these issues, interpolating and re-gridding of both image-sets to the same coordinate system after rigidly registering them is a best practice (Figure 2). The MR-image should remain fixed while moving the CT-image and subjecting it to tri-linear interpolation. In this way, the five discrete Hounsfield units of air, fat, water, cortical bone and spongy bone in the MRCAT are preserved. By matching the voxel boundaries, the re-gridding step minimizes partial volume effects. Furthermore, using this method means that all structures can be delineated on slices of the same geometry. This prevents discretization errors arising from a potential inclination of the two underlying image-sets in relation to each other.

Partial volume effects increase with voxel size. Since the dose grids are usually coarser than the underlying density images, re-gridding should be the first step. While dose grids are, by default, aligned with the underlying images, it is important to also ensure that the origins match.

There are many numerical software packages and toolboxes available that offer ready-to-use implementations for performing re-gridding and interpolation.

Priorizations in registration

In practice, the visual quality of the registration mostly depends on which part of the anatomy is given priority. Possibilities in this pelvic CT context include the body outline, the bony structure or the fiducial markers inside the prostate. A perfect whole-body match is rarely achievable. In general, either the bony structures or the body outline display positioning-related differences. We suggest focusing on registration of the bones, since with them it is easier to re-align mismatches in body outline. For the purpose of dose comparison, we discourage registering the positions of the fiducial markers. This approach usually results in the largest positional differences of the overall body.
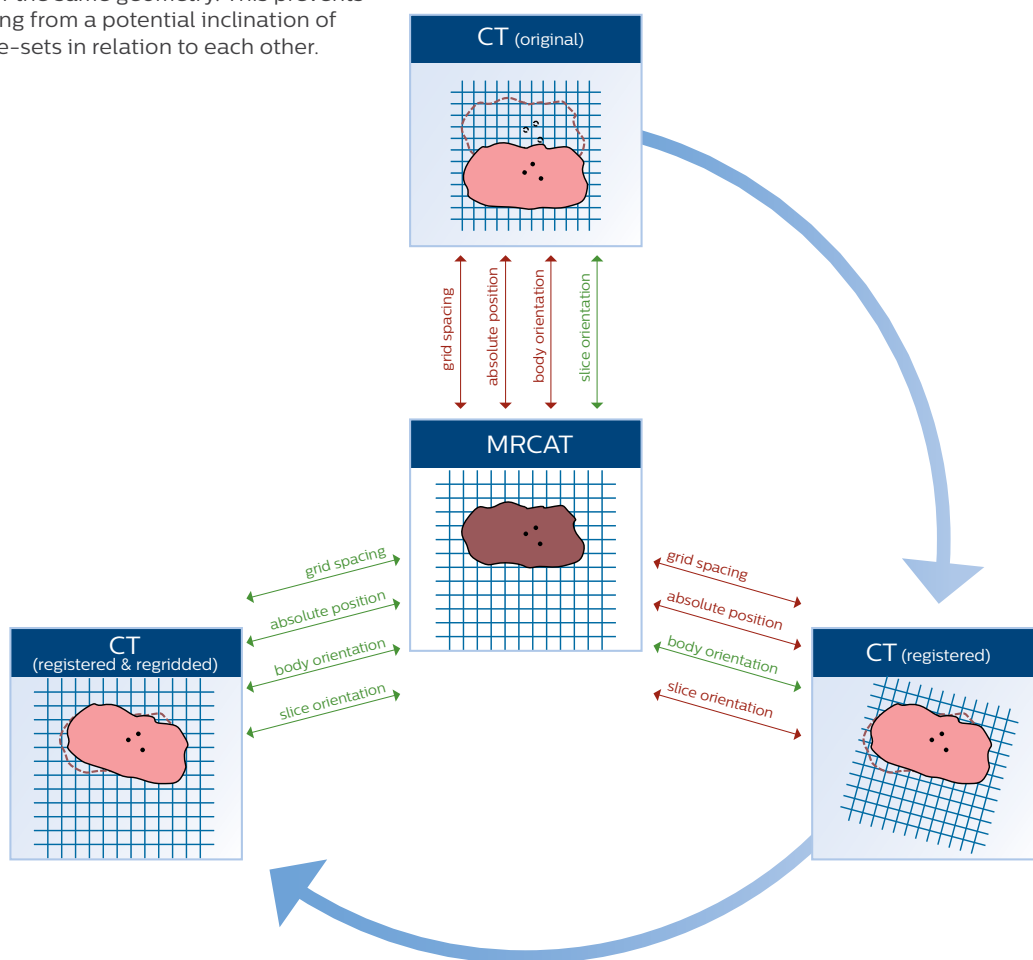

Figure 2. Impact of registration and re-gridding. The green arrows refer to matching characteristics while red arrows indicate mismatch. Residual differences between MRCAT and CT in the registered and re-gridded image are intrinsic to the images.

Matching the body outline
Due to repositioning of the patient between the two imaging sessions for CT and MRI, unavoidable differences in the body outline will occur that are not related to the performance of MRCAT.

One feasible method of reducing differences in body outline is to use the intersection of both body outlines and clip any volume outside to air density. Alternatively, you can retain one of the two body outlines, while the other one is clipped to air density on the outside and any voids inside are filled with water density.[8] An obvious drawback of this clipping approach is that any geometrical distortions of the body outline are concealed. However, a specially designed quality assurance phantom can independently assess the geometric accuracy of the MR-images.

Deformable image registration
In the quest to achieve perfect correlation of the MRCAT and CT images, deformable registration methods can also play a role. However, the appealing appearance and visually perfect match contrast with an intrinsic lack of volume conservation and the challenges in quality assurance of this method. Although deformable image registration lacks widespread acceptance in clinical practice, several groups have reported using it in their clinical workflow.[5] For the purposes of this discussion, we will restrict ourselves to rigid registration.

Aspects of dose (re-) calculation
Once you have registered and re-gridded the two image-sets, the next step is to calculate an RT-dose distribution on each of them, that originates from the same RT-plan and uses identical RT-structures (Figure 3). To this end, the RT-structures can either be delineated on the registered image-sets or propagated from the original CT image using the transformation matrix already identified. Next,

you can optimize the plan on either CT or MRCAT images along with the corresponding dose calculation for this plan. To calculate the dose on the other image set, it is essential that you refrain from repeating the optimization step. Instead, calculate the dose from the unedited plan but use a different underlying image-set. Most treatment planning systems offer this option as a quality assurance or adaptive planning tool, e.g. the Dynamic Planning option in Philips Pinnacle[3]. Whether it is better to perform initial optimization on CT or MRCAT is subject to discussion. Assessing the dose difference in the results from both approaches provides additional insights into the impact of the changed workflow.

For treatment planning systems that base their dose calculations on direct Monte-Carlo simulation, such as Monaco, it is crucial to choose suitable convergence criteria. The dose distributions from such algorithms show larger numerical fluctuations compared to the results from model-based algorithms such as Collapsed Cone Convolution (Pinnacle[3]) or Anisotropic Analytic Algorithm (Eclipse). In practice, a statistical accuracy of 1% per control point sufficiently reduces the stochastic noise at a cost of prolonging calculation time, when compared to current clinical practice.[4]

Metrics for dosimetric comparison

MR-only simulation is designed to provide accurate attenuation information that is sufficient for radiotherapy planning. That is to say that for identical plans obtained using CT or MRCAT, the resulting dose distributions are dosimetrically equivalent within clinically acceptable limits. MRCAT, however, is not intended to be visually similar to CT.
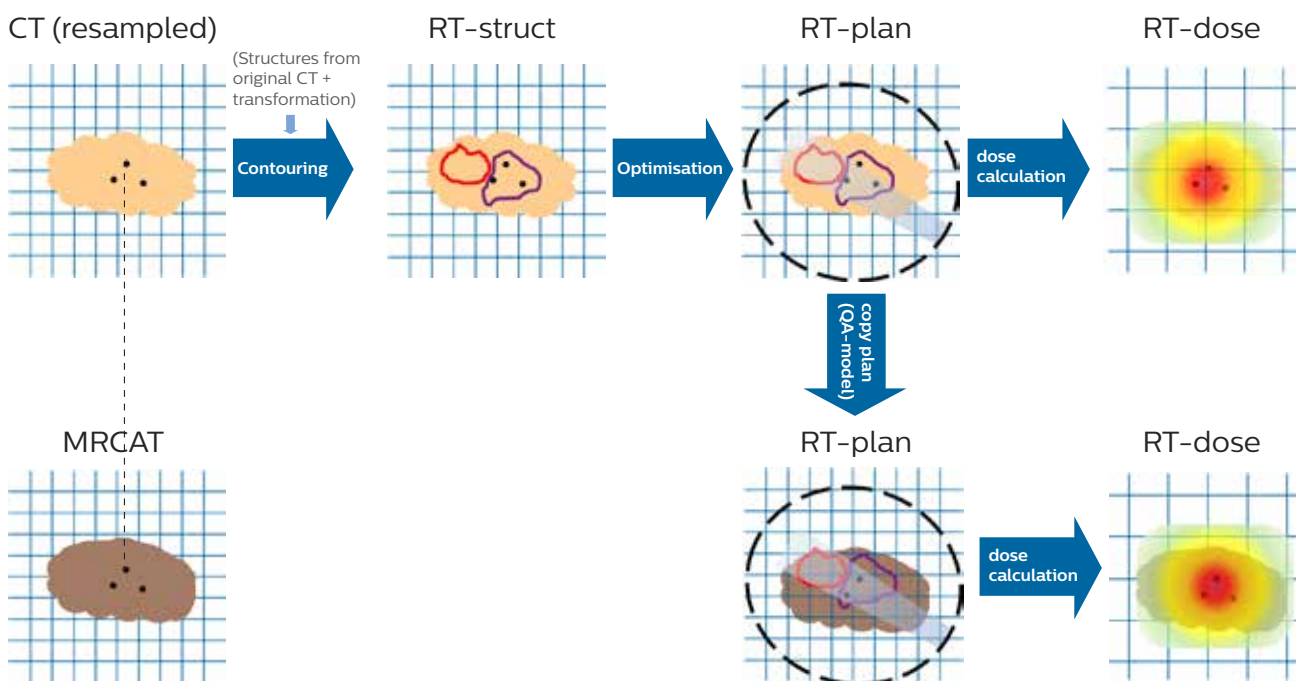

Figure 3. Schematic overview of a best practice approach to setting up a dose comparison between CT- and MRCAT-based dose plans

Voxel-wise comparison

Validations in this document focus on dose comparison or derived metrics, such as gamma analysis (Figure 4), rather than a comparison of the underlying density information. Dose values are the result of accumulated tissue information along each of the beam directions and therefore they 'have a memory'. Due to its local nature, local density information may prove useful in determining where observed dose differences originate (e.g., mismatched bones, internal air cavities, etc.).

The most comprehensive way to investigate dose differences is to perform a full, three-dimensional, voxel-wise comparison. Provided that the data have been properly aligned and re-gridded, this is a straightforward approach.

Creating dose volume histograms

For most practical purposes, the full volumetric information is too dense. Ideally, the aim is to compress the information to allow easier comparison between cases. The natural way to do this is to create dose volume histograms (DVHs) over certain volumes and quantify the difference between the DVHs for these volumes. Suitable volumes can be based on either actual RT-structures, or on more general definitions, such as the region that receives more than 75% of the prescribed dose.

Utilizing scorecards

However, querying the full information contained in the DVH might prove to be cumbersome. You can further reduce the information while retaining the major characteristics of the DVH by recording the obtained values, for quantities that have also been used as optimization objectives, e.g., D_{98}, D_{mean}, D_2. Clinically, such values are expressed in scorecards. These provide a quick and convenient way to verify that plan criteria have been met. The difference between these values when calculated from CT- and MRCAT-based dose distributions provides the highest level comparison metric which can also be extracted and reported for a larger patient cohort.

Performing gamma analysis

Direct dose comparison is susceptible to slight spatial misregistrations which, as explained above, cannot be fully excluded and which, as such, place limitations on this method. The established way to deal with these is to consider the gamma values instead of voxel-wise dose difference. In this respect, gamma analysis acts like a local degree of freedom in the registration, by relaxing the need for a perfect spatial match of corresponding dose values.

Moreover, we can view the gamma values as a three-dimensional quantity that can be evaluated across the patient. It is, however, more common to condense the gamma values into a histogram over a certain volume of interest and/or to characterize the histogram by its key parameters. Likely the most relevant quantity is the gamma pass ratio, which is the percentage of voxels for which gamma is less than one.

Acceptance criteria

The acceptance values for gamma analysis reported in literature vary. It is widely accepted that to assure the quality of the end-to-end workflow, a criterion of 3%/3mm must be met. However, several groups report even stricter requirements (2%/2mm, 2%/1mm, or 1%/1mm) when evaluating the impact of replacing CT with MRCAT. In general, for the test to pass, 99% of the considered voxels must fulfill the defined gamma criterion (some groups specify 95%).
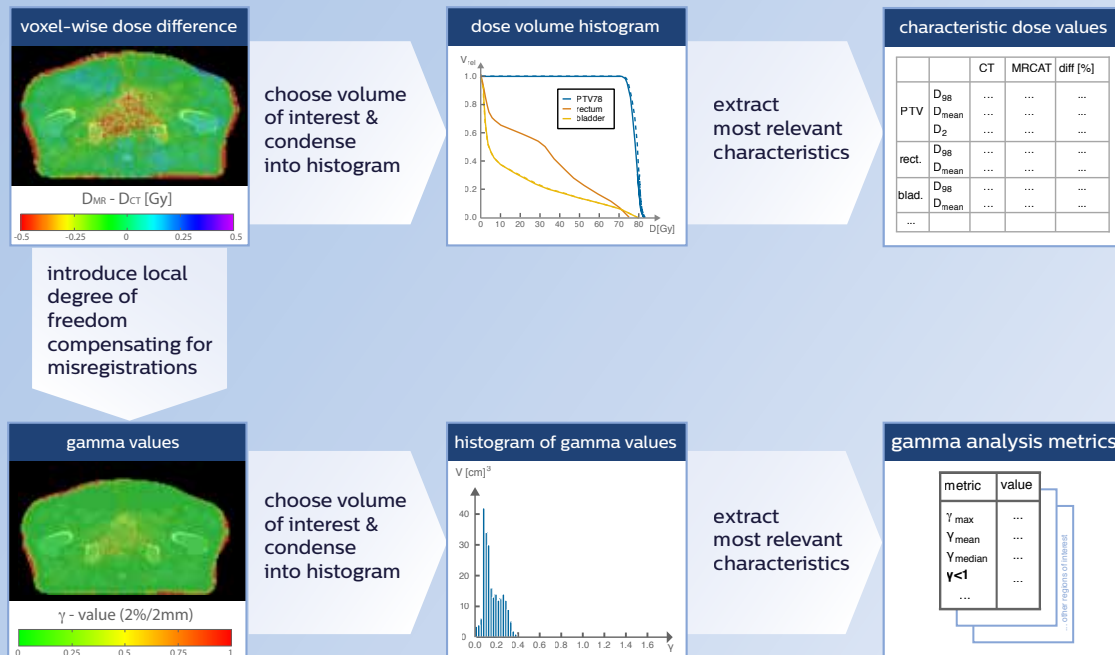

Figure 4: Metrics for dosimetric comparison. For detailed explanation on the methods, see the body text

Evaluating the dose differences helps us arrive at a finer granular judgement. Standard gamma analysis is unable to discern a systematic over- from an under-dosage, as long as their magnitude remains below the threshold. There seems to be consensus that average dose differences to target structures above 1% would not be acceptable for releasing MRCAT for clinical use. For systematic differences below 1%, the situation is more complex. Experts have worked on disentangling the relative contributions to the observed differences before proceeding to a clinical workflow.[4]

A straightforward implementation of scorecard-based acceptance criteria would only enforce consistency: a plan that meets the acceptance criteria on CT should also fulfill these criteria on MRCAT and vice versa. The same should hold for plans that fail to meet the requirements. In practice, this seemingly trivial request is difficult to fulfill. Since in the optimization process, multiple objectives must be balanced against each other, some plans result in some criteria only just reaching the target. A tiny dose difference can make this target fail on the other image-set. In this respect, scorecard-based acceptance criteria are only useful in cases where the plan is re-optimized for the two image-sets.

Practical considerations in dosimetric comparison

Hounsfield unit to density conversion
Of course, the result of a dose comparison depends on the ground-truth that is used in the comparison. Modern treatment planning systems base their dose calculation on electron or mass densities rather than Hounsfield units. A conversion table supplied by the operator and established during commissioning of CT governs conversion between one and the other. The resulting Hounsfield unit to density calibration tables differ according to the chosen CT-calibration phantom at commissioning.[9] The conversion table supplied with Philips MR-only simulation (Figure 5) reflects the best match to an 'average' CT calibrated with the CIRS-062-phantom. Sites that use a different calibration table for CT can expect larger dose differences. In principle, matching the two calibration tables can reduce these differences.[4] When it comes to making an informed decision as to which calibration table to deploy, an awareness of this aspect is crucial. Medical physicists can use the default table from Figure 5, or an adapted one that better reflects the characteristics of the local CT and ensures continuity of the dose prescription.

When using a custom conversion table, dosimetric equivalence of plans should be tested and the customer should measure the correctness of their CT end-to-end performance with the CT (reference) calibration curve.

Overcoming internal air cavities
Intestinal gases are regarded as a transient effect and the corresponding air pockets are assigned a soft tissue value in the MRCAT images. This is in line with current practice at several sites at which internal air cavities are manually overwritten in the CT-based workflow.

Similarly, the MRCAT algorithm will not differentiate the presence of permanent, artificial air-filled objects inside the patient's body. In this case, a manual overwrite
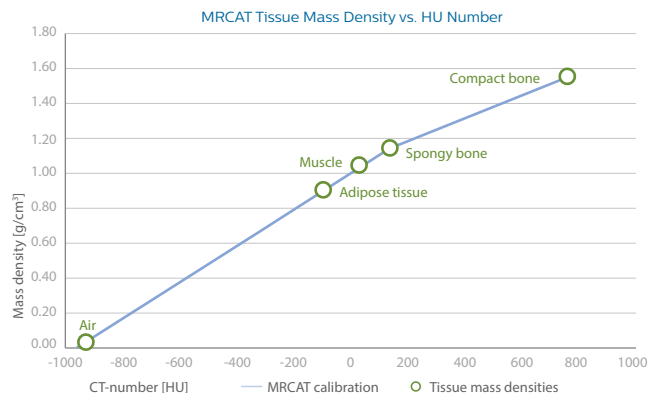

Figure 5. Default calibration table from Hounsfield units to mass densities for use together with MRCAT images.

to air density is required in order to avoid unintended dose mismatch. Such an air cavity can be readily delineated from the in-phase image and used for manual density override in the treatment planning system.

Upgrading to new software releases
The Hounsfield units of the tissue compartments can vary across different software releases. These changes reflect enhancements in stability of the image acquisition and changes in the assignment of voxels at tissue class boundaries. The Philips approach has been that, the discrete calibration values for each of the tissue classes should fall onto the same continuous calibration curve in Figure 5. The need for at least partial re-commissioning may arise upon software upgrades.

Patient positioning verification

Positioning on the linear accelerator
Patient positioning on the linear accelerator is a reoccurring task for each fraction. To support precise therapy delivery, it is crucial to correctly match the patient positions at the planning stage with position verification images. Transition to an MR-only simulation workflow requires patient positioning to be based on MR image sets solely. Here, each facility must find and follow their own approach depending on the technical equipment available and local guidelines.

Cone-beam CT-based positioning
For linear accelerators equipped with a cone-beam CT, this match can be done based on bony anatomy, soft tissue contrast, or fiducial markers in the prostate. The frequency of cone-beam CT-based position verification varies from hospital to hospital. Some centers perform this check for each fraction, others only during the first fractions in order to reduce the overall dose given to the patient. Assessments as to whether the patient's anatomy has changed in response to treatment and re-planning is required are usually carried out at weekly intervals. Positioning for the other fractions is usually done using planar kV or MV images that are matched against digitally reconstructed radiographs (DRRs).

MRCAT accuracy in practice

Several groups have independently performed studies on the dosimetric accuracy of MRCAT. While these differ in terms of the metrics considered and study design, (see Table 1) the general outcome is comparable. Results from MSKCC show a mean difference of less than 0.5% throughout the patient population.[5]

The Utrecht study elaborates on the impact of the Hounsfield unit to density conversion. For matching curves, this site reports a mean dose difference of 0.3% in the CTV.[4] Reported average gamma passing rates for the 3%/3mm and the 2%/2mm criterion range from 99% to 100%.[4][6][7].

		Memorial Sloan–Kettering Cancer Center, USA	University Medical Center Utrecht, The Netherlands	Odense University Hospital, Denmark	Turku University Hospital, Finland
Country		USA	The Netherlands	Denmark	Finland
Reference		[5]	[4]	[6]	[7]
Number of patients		20	14	30	5
HU to density conversion table	MRCAT same as CT	X	X	X	
	Philips provided	–	X	–	X
Registration method	Rigid	–	X	X	
	Deformable	X	–	–	X
	Re-gridding	X	X	X	X
Used evaluation metrics	Abs. dose diff.	X			
	Rel. dose diff.	X	X		X
	Gamma 3% / 3mm		X		
	Gamma 2% / 2mm		X	X	X
	Gamma 1% / 1mm			X	x
Number of volumes for evaluation	Structures from RTSS	15	14	10	11
Treatment planning system		Eclipse	Monaco	Pinnacle³	Pinnacle³
Treatment technique		VMAT	IMRT	VMAT	VMAT

Table 1. Details of published dose comparison studies performed at clinical sites.

DRR-based positioning

Transition to an MR-only workflow is straightforward using DRRs. Treatment planning systems can also create DRRs from the MRCAT images, which can then be used in the position matching (Figure 6).

Positioning accuracy depends on the quality of the DRRs and on how well these are (manually) matched to the kV images on treatment day. Obviously, manual matching is subject to inter-observer variability.
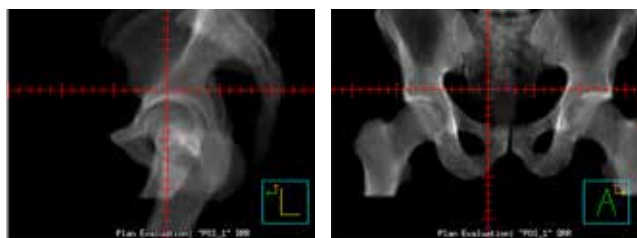
If replacing CT by MRCAT images results in DRRs for which the positioning accuracy does not exceed the inter-observer variability for CT-based DRRs, this step of the commissioning process passes.

Fiducial markers

Both for DRR- and cone-beam CT-based positioning, the presence of fiducial markers in the prostate can increase positioning accuracy. In the MRCAT images, fiducial markers are not depicted as high-density objects, but shown as soft tissue. In order to make them available for positioning, they need to be manually delineated, for instance in the in-phase images. A dedicated seed detection scan provides complementary information for delineation and supports the differentiation of fiducial markers from other signal voids, such as calcifications. The resulting structures can be linked to the DRR and pushed to the linear accelerator. In a similar way, delineations of other anatomical structures can also be used for patient positioning. A contour-based workaround like this is necessary since the on-board imagers at commercial linear accelerators can currently register kV images only to CT images (such as MRCAT), but do not yet provide the possibility of registration with MR images.

Figure 6 . Illustration of DRRs created from MRCAT.

Overcoming workflow limitations in practice

As outlined above, there are many versatile approaches to positioning in an MR-only workflow. In the following, we present how three hospitals chose to address this task and overcome current workflow limitations.

The approach adopted at Turku University Hospital, Finland, in essence, corresponds to the outline above. The center positions of manually delineated fiducial markers are matched against the kV-images from the on-board imager. The hospital has already applied this approach to treat almost 70 prostate cancer patients in an MR-only workflow.[10]

Memorial Sloan-Kettering Cancer Center, USA compared two positioning approaches: a DRR-based matching of MRCAT and pre-treatment kV-images focusing on either the bony anatomy or the location of the fiducial markers.[5] Here, the markers are delineated from the seed detection scan. Prior to generating the DRRs, the locations of the markers are identified in the MRCAT image and the corresponding voxels are overridden with a high HU value. To cover the entire range of densities, an extension of the Hounsfield unit to density calibration table is necessary in this case.

Odense University Hospital, Denmark, manually matched the cone-beam CT visible fiducial markers against delineated marker contours from the in-phase planning images in their first MR-only treatment.[6] With respect to streamlining the workflow, they developed an in-house solution that facilitates an automatic match of the marker positions.

Conclusion

As this paper has demonstrated, there are a number of ways to perform the MR-only simulation commissioning process. The one you select will obviously depend on parameters, system set-up, requirements and goals at your facility. With this document, we aimed to provide a useful overview of best-practice methods, and shed some light on how they can assist in your comparisons MR-only simulation of CT and MRCAT-based simulations. Due to its excellent soft-tissue contrast and functional imaging capabilities, MRI is accepted as a suitable imaging modality to drive delineation accuracy of targets and organs at risk. Capturing this value and implementing a single-modality approach can deliver a range of benefits in the radiation therapy planning process. With the introduction of the first MR-only simulation package for prostate, Philips is paving the way for further adoption of MRI in RT planning.

Further reading

1. M. Kohler, T. Vaara, M. van Grootel, R. Hoogeveen, R. Kemppainen and S. Renisch, "MR-only simulation for radiotherapy planning – Whitepaper: Philips MRCAT for prostate dose calculations using only MRI data".

2. AAPM, RPT 83 for CT, Med. Phys., vol. 30, no. 10, 2003.

3. AAPM, RPT 100 for MR, ISBN:978-1-936366-02-6, 2010.

4. M. Maspero, P. R. Seevinck, G. Schubert, M. A. U. Hoesl, B. van Asselen, M. A. Viergever, J. J. W. Lagendijk, G. J. Meijer and C. A. T. van den Berg, "Quantification of confounding factors in MRI-based dose calculations as applied to prostate IMRT," Phys. Med. Biol., 62(3), 948–965, 2017.

5. N. Tyagi, S. Fontenla, J. Zhang, M. Cloutier, M. Kadbi, J. Mechalakos, M. Zelefsky, J. Deasy and M. Hunt, "Dosimetric and workflow evaluation of first commercial synthetic CT software for clinical use in pelvis," Phys. Med. Biol., 62(8), 2961–2975, 2017.

6. R. L. Christiansen, H. R. Jensen and C. Brink, "Magnetic resonance only workflow and validation of dose calculations for radiotherapy of prostate cancer," Acta Oncol, 56(6), 787–791, 2017.

7. R. Kemppainen, S. Suilamo, T. Tuokkola, P. Lindholm, M. H. Deppe and J. Keyriläinen, "Magnetic resonance-only simulation and dose calculation in external beam radiotherapy: a feasibility study for pelvic cancers," Acta Oncol, 56(6), 792–798, 2017.

8. J. Korhonen, M. Kapanen, J. Keyriläinen, T. Seppälä, M. Tenhunen, "A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer.," Med. Phys., 41(1), 2014.

9. J. Hatton, B. McCurdy and P. B. Greer, "Cone beam computerized tomography: the effect of calibration of the Hounsfield unit number to electron density on dose calculation accuracy for adaptive radiation therapy," Phys. Med. Biol., 54(15), N329–346m 2009.

10. As reported by Turku University Hospital, Finland. March 2017..

MR-only simulation is not available in all countries and for all configurations. Please contact your local Philips representative for further details.

PHILIPS

How to reach us
Please visit www.philips.com/mr-rt
healthcare@philips.com