How to measure CT image quality: Variations in CT-numbers, uniformity and low contrast resolution for a CT quality assurance phantom

Kristine Gulliksrud a, Caroline Stokke a, b, Anne Catrine Trægde Martinsen a, c, *

a The Intervention Centre, Oslo University Hospital, Rikshospitalet, Oslo, Norway
b Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Oslo, Norway
c Department of Physics, University of Oslo, Oslo, Norway

Abstract

Purpose: Quality assurance (QA) phantoms for testing different image quality parameters in computed tomography (CT) are commercially available. Such phantoms are also used as reference for acceptance in the specifications of CT-scanners. The aim of this study was to analyze the characteristics of the most commonly used QA phantom in CT: Catphan 500/504/600.

Methods: Nine different phantoms were scanned on the same day, on one CT-scanner with the same parameter settings. Interphantom variations in CT-number values, image uniformity and low contrast resolution were evaluated for the phantoms. Comparisons between manual image analysis and results obtained from the automatic evaluation software QAlite were performed.

Results: Some interphantom variations were observed in the low contrast resolution and the CT-number modules of the phantoms. Depending on the chosen regulatory framework, the variations in CT-numbers can be interpreted as substantial. The homogenous modules were found more invariable. However, the automatic image analysis software QAlite measures image uniformity differently than recommended in international standards, and will not necessarily give results in agreement with these standards.

Conclusions: It is important to consider the interphantom variations in relation to ones framework, and to be aware of which phantom is used to study CT-numbers and low contrast resolution for a specific scanner. Comparisons with predicted values from manual and acceptance values should be performed with care and consideration. If automatic software-based evaluations are to be used, users should be aware that large differences can exist for the image uniformity testing.

Introduction

International radiation protection authorities like ICRP and IAEA recommend acceptance tests and periodically QA-tests of CT-scanners with respect to radiation dose and image quality. In some countries image quality and radiation dose QA-tests are mandatory [1].

The most important parameters with respect to CT image quality are image noise, CT-numbers, uniformity, spatial resolution and low contrast resolution. Low contrast resolution is the ability to differentiate objects with slightly different densities. The potential to resolve an object depends on the level of contrast in the object and its size, reconstruction algorithm, image noise and window settings used to display the image. Spatial resolution is the ability to differentiate small objects with high contrast compared to the background. The uniformity describes how uniform the image of a homogenous material appears. The uniformity measurements are important to ensure that cupping and beam hardening artifacts are avoided. CT images are generated by measurement of attenuation of x-rays through the tissue of interest. Attenuation is described by CT-numbers or Hounsfield units (HU). Each pixel in the image corresponds to a specific HU [2].

* Corresponding author. Tel.: +47 23070111.
E-mail addresses: krgull@ous-hf.no (K. Gulliksrud), uxnets@ous-hf.no (A.C. Trægde Martinsen).

http://dx.doi.org/10.1016/j.ejmp.2014.01.006

Please cite this article in press as: Gulliksrud K, et al., How to measure CT image quality: Variations in CT-numbers, uniformity and low contrast resolution for a CT quality assurance phantom, Physica Medica (2014), http://dx.doi.org/10.1016/j.ejmp.2014.01.006
uniformity device not to deviate by more than ±4 HU from the nominal values specified by the manufacturer of the CT-scanner for the specific material of the test device [3]. This material is often water, and the specification value zero by definition. The uniformity (the deviation in mean CT-number between central and peripheral regions) must not be greater than 4 HU at acceptance [3,4]. In addition, the difference in uniformity should not vary by more than 2 HU from baseline values [4]. With respect to low contrast resolution, the detectability should meet the specifications of the manufacturer, but because of difficulty in objectively measuring low contrast resolution IEC does not recommend this method as an acceptance test. IPEM suggests a level of ±5 HU from the baseline CT-number of water and ±10 HU from the baseline number of other materials. The suspension levels are ±20 HU and ±30 HU, respectively. According to IPEM the uniformity should not exceed 10 or 20 HU, depending on phantom size [5].

QA phantoms for CT are commercially available. Catphan 500/ 504/600 phantoms (The Phantom laboratory, Salem NY, USA) are used for CT QA-tests worldwide [6–8]. The Catphan phantoms are used as reference for acceptance. To our knowledge there are no publications concerning the potential differences between phantoms, and there is a need to establish that these phantoms are suited for such tests. The aim of this study was to analyze the characteristics of Catphan 500/504/600 phantoms with respect to interphantom variations in low contrast resolution, CT-numbers and uniformity.

Materials and methods

Nine different Catphan phantoms were tested: 3 Catphan 500, 1 Catphan 504 and 5 Catphan 600. The Catphan phantoms varied in age (1–12 years). The Catphan phantoms were scanned on a Toshiba Aquilion One 320 slice CT-scanner (Toshiba medical systems, Zoetermeer, the Netherlands). Air calibration was performed the same day, before initiating the study. The water CT-number was measured to 0.6 HU using the vendor phantom, well within the specifications of the phantoms. The phantoms were scanned subsequently with the same parameter settings (Table 1). The scanning conditions were the same for all phantoms. One of the phantoms was scanned three times to study intraphantom variations. The phantom was repositioned for each scan.

CT-numbers

The sensitometry module (CTP401/CTP404) contains inserts with different known densities for measurements of CT-numbers and linearity of CT-numbers for different mass densities. Catphan 500/504/600 has inserts made from teflon, acrylic, low density polyethylene (LDPE), and air. Catphan 504/600 has in addition inserts made from polymethylpentene (PMP), delrin and polystyrene (Fig. 1).

CT-numbers were measured in all 9 phantoms. The measurements were performed by manually placing ROIs within the inserts in the CT images by three investigators. The placement of the ROIs was aligned between all phantoms and different investigators. The image viewer program DicomWorks 1.3.5 (2002 Philippe Puech – Loic Boussel) was used in the manual analysis. The results were also analyzed with use of the program QAlite (The Institute For Radiological Imaging Sciences, Inc, MD, US) by one investigator. QAlite is a CT automated QA software that supports Catphan 500/600. Phantom 9, a Catphan 504, is excluded from results where QAlite analysis has been used, since QAlite cannot read this specific phantom.

Low contrast resolution

Module CTP515 in the phantoms contains low contrast supra-slice targets with diameters 2–15 mm, and contrast levels of 0.3%, 0.5% and 1.0% (Fig. 1), used to evaluate the ability to differentiate objects with slightly different densities. Low contrast was measured by counting numbers of visible targets, assessed by three independent investigators on the same PACS screen, with the same viewing conditions (same window settings, same day, subdued ambient light). A Fleiss Kappa test [9] was used to evaluate the interobserver differences in low contrast detectability.

CNR between the largest targets of the 0.5% and 1.0% groups and the background were measured for all phantoms. ROIs of the same size were used to measure the mean CT-numbers and the standard deviations in the largest targets and the background in the low contrast module, and the CNR was calculated by the formula:

\[
\text{CNR} = \frac{2(C_0 - C_b)}{\sigma_0 + \sigma_b}
\]

where \(C_0\) and \(C_b\) are the means of the target and background, respectively, and \(\sigma_0\) and \(\sigma_b\) are the standard deviations of the target and background [10]. The Pearson correlation coefficient between the CNR of the 0.5% and 1.0% group of each phantom, respectively, was calculated using the program R (version 2.14.2, 2012 The R Foundation for Statistical Computing).

Uniformity

The consistency of the CT-numbers in the image of a homogenous material was measured in the uniformity module (CTP486) in the phantoms (Fig. 1). Uniformity of the CT-numbers was measured manually by placing ROIs, one in the middle and four in the periphery of the module (clock positions 12, 3, 6 and 9). Measurements were made with the image viewer software DicomWorks 1.3.5 and ImageJ 1.45s (Wayne Rasband National Institutes of Health, USA). The observer used ROIs with a diameter of 10% of the diameter of the image of the uniformity module, in agreement with IEC 61223-3-5. The results were also analyzed with use of the software analyze program QAlite.

Results

CT-number measurements

Measured CT-numbers for the different materials for the 9 phantoms are shown in Table 2. ROIs used were 40 mm². Phantoms 3, 4 and 8 are Catphan 500 phantoms and contain fewer densities. The CT-numbers for each density and phantom are presented.

Table 1

<table>
<thead>
<tr>
<th>Technique</th>
<th>Pitch</th>
<th>Tube voltage (kV)</th>
<th>Tube current (mA)</th>
<th>Rotation time (s)</th>
<th>Collimation (mm)</th>
<th>Filter type</th>
<th>Matrix</th>
<th>Convolution kernel</th>
<th>Slice width (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT-number</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniformity</td>
<td>Volume SS</td>
<td>120</td>
<td>250</td>
<td>1</td>
<td>80&lt;0.5</td>
<td>Large Body</td>
<td>512</td>
<td>FC13</td>
<td>4</td>
</tr>
<tr>
<td>Low contrast</td>
<td>Helical</td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please cite this article in press as: Gulliksrud K, et al., How to measure CT image quality: Variations in CT-numbers, uniformity and low contrast resolution for a CT quality assurance phantom, Physica Medica (2014), http://dx.doi.org/10.1016/j.ejmp.2014.01.006
Median values for all phantoms from both manual analysis and QA lite analysis, and predicted values from the Catphan manual are given.

There were some interphantom variations in CT-numbers for all densities. The variations between minimum and maximum values were 10.3 (Air), 7.3 (Teflon), 1.4 (Delrin), 2.0 (Acrylic), 4.0 (Poly styrene), 3.4 (LDPE) and 1.4 (PMP) HU. In percent the deviations ranged from 1% or less (Air, Teflon, Delrin, PMP), 2% (Acrylic), 4% (LDPE) to 12% (Polystyrene). The measured HU also deviated to some extent from the predicted values from the Catphan manual. There were small variations in CT-numbers between investigators (less than 3% or 3 HU for all materials and phantoms). The manually measured CT-numbers for each phantom deviated little from QA lite analyses. The intraphantom deviation was less than 1% for all inserts.

Low contrast resolution

Low contrast resolution varied between phantoms. The low contrast resolution ranged from 0 – 3, 2 – 6 and 6 – 8 visualized targets for the 0.3%, 0.5% and 1% contrast resolution groups, respectively. The interphantom variation range was largest for the 0.5% contrast group. The mean of three observers for the three contrast groups for each phantom is shown in Fig. 2.

The kappa value for the low contrast resolution for the three observers was 0.26. One observer detected more low contrast objects than the two other observers, contributing to this value. However, the trend in the interphantom resolution was clear; worse or better resolution was demonstrated for the same phantoms for all three observers.

CNR between the largest targets and the background in the module was also measured for all phantoms. ROIs used for CNR measurements were 45 mm³. CNR varied from 1.5 to 5.5 for the 1% contrast group and 0.3–1.8 for the 0.5% contrast group (Table 3). The Pearson correlation coefficient between the 0.5% and 1.0% group of the phantoms was 0.49, and no significant intraphantom correlation between these groups were observed (p = 0.18).

Uniformity

The consistency of CT-numbers in an image of a homogenous material was measured in the uniformity module in the phantoms (Table 4). The periphery values presented are the mean values of the four periphery positions (the deviations between these regions were in general small). The deviations listed to the right in the table represent the differences between the two regions of minimum and maximum CT-number value. As the phantoms were scanned using large body FOV, the calculated uniformity cannot be compared directly with international tolerance levels. Still, the interphantom variation and the differences between manual and automatic analyses can be accurately interpreted. Calculated deviations found using the QA lite program were quite larger than manually measured deviations, except for one phantom. For many phantoms the uniformity worsened by 2–3 HU using QA lite compared to manually analysis, in one case the difference was almost 4 HU. QA lite were found to use small ROIs (diameters approximately 5% of the diameter of the image) opposed to the manual ROIs which in accordance with international recommendations were larger (diameter 10% of the diameter of the image of the test device). Manual measurements using smaller ROIs in agreement with the ROI size used by QA lite were also performed to test the actual CT-number values obtained by QA lite. The CT-number measures by QA lite were confirmed by these additional manual analyses (data not shown).

The intraphantom variability was measured to 0.2 HU by the standard deviation of the three different uniformity scans.

Discussion

The Norwegian radiation protection authority demands periodically QA-tests with respect to image quality and radiation dose, and optimization of all examination protocols with respect to image quality and radiation dose [1]. Image quality parameters like spatial resolution, slice thickness, low contrast detectability,
For objectively comparing image quality parameters of different scanners, the Catphan phantoms are used by physicists worldwide [7,8,11]. Few studies are published on possible interphantom variations and the robustness of using this phantom for QA in CT. It is topical to address the need for a survey on possible interphantom variations and establishment of baselines for the different modules of these phantoms.

The CT-numbers are dependent of several factors like spectral energy, reconstruction algorithms and filtration of radiation. Previous studies have concluded that absolute HU values should not be used for clinical purposes [12,13]. However, HU values are still used by radiologists to characterize and differentiate between different tissue and fluid [14–17]. The Catphan manual lists predicted HU for the different inserts in the sensitometry module in the phantom. Comparison between a specific scanner and these numbers should to a certain degree reflect the scanner’s deviation from CT-numbers found in literature. Earlier studies have demonstrated that these predicted values can deviate from obtained scanner values, and that large interscanner variations can exist [2,18].

Variations in CT-numbers of a scanner over time might indicate lack of consistency in calibration or possible changes in tube filtration. Our results indicate that there are interphantom variations in HU values for the phantoms (Table 2). Measured in absolute values the deviations were observed around 10 HU for materials of uniformity, noise and CT-numbers can be part of the QA-test protocols. However, very little research is concerned by whether these tests ensure that serious errors are detected. Questions about the suitability and robustness of the internationally recommended QA-tests and acceptance criteria of radiation dose and image quality should be raised. For example, if a scanner passes the image QA-tests, will patient pathology actually be revealed?

For objectively comparing image quality parameters of different scanners, the Catphan phantoms are used by physicists worldwide [7,8,11]. Few studies are published on possible interphantom variations and the robustness of using this phantom for QA in CT. It is topical to address the need for a survey on possible interphantom variations and establishment of baselines for the different modules of these phantoms.

The CT-numbers are dependent of several factors like spectral energy, reconstruction algorithms and filtration of radiation. Previous studies have concluded that absolute HU values should not be used for clinical purposes [12,13]. However, HU values are still used by radiologists to characterize and differentiate between different tissue and fluid [14–17]. The Catphan manual lists predicted HU for the different inserts in the sensitometry module in the phantom. Comparison between a specific scanner and these numbers should to a certain degree reflect the scanner’s deviation from CT-numbers found in literature. Earlier studies have demonstrated that these predicted values can deviate from obtained scanner values, and that large interscanner variations can exist [2,18].

Variations in CT-numbers of a scanner over time might indicate lack of consistency in calibration or possible changes in tube filtration. Our results indicate that there are interphantom variations in HU values for the phantoms (Table 2). Measured in absolute values the deviations were observed around 10 HU for materials of uniformity, noise and CT-numbers can be part of the QA-test protocols. However, very little research is concerned by whether these tests ensure that serious errors are detected. Questions about the suitability and robustness of the internationally recommended QA-tests and acceptance criteria of radiation dose and image quality should be raised. For example, if a scanner passes the image QA-tests, will patient pathology actually be revealed?

For objectively comparing image quality parameters of different scanners, the Catphan phantoms are used by physicists worldwide [7,8,11]. Few studies are published on possible interphantom variations and the robustness of using this phantom for QA in CT. It is topical to address the need for a survey on possible interphantom variations and establishment of baselines for the different modules of these phantoms.

The CT-numbers are dependent of several factors like spectral energy, reconstruction algorithms and filtration of radiation. Previous studies have concluded that absolute HU values should not be used for clinical purposes [12,13]. However, HU values are still used by radiologists to characterize and differentiate between different tissue and fluid [14–17]. The Catphan manual lists predicted HU for the different inserts in the sensitometry module in the phantom. Comparison between a specific scanner and these numbers should to a certain degree reflect the scanner’s deviation from CT-numbers found in literature. Earlier studies have demonstrated that these predicted values can deviate from obtained scanner values, and that large interscanner variations can exist [2,18].

Variations in CT-numbers of a scanner over time might indicate lack of consistency in calibration or possible changes in tube filtration. Our results indicate that there are interphantom variations in HU values for the phantoms (Table 2). Measured in absolute values the deviations were observed around 10 HU for materials of uniformity, noise and CT-numbers can be part of the QA-test protocols. However, very little research is concerned by whether these tests ensure that serious errors are detected. Questions about the suitability and robustness of the internationally recommended QA-tests and acceptance criteria of radiation dose and image quality should be raised. For example, if a scanner passes the image QA-tests, will patient pathology actually be revealed?
the highest or lowest CT-numbers. A remedial level of ±10 HU from the baseline CT-number is suggested [5]. Our results show that this deviation can be expected due to interphantom variation alone. Measuring the deviations in percent would reduce the deviations for these CT-numbers. However, for the materials with CT-numbers closer to water the deviations would then exceed 10%. Therefore, regardless of employing a percent or absolute scale, the same phantom should be used for acceptance test and annually QA-tests for one specific scanner. Then, baseline values can be found and results from the annual tests compared to these and followed over time.

The Catphan phantoms vary in age from 1 to 12 years. Since the phantoms were manufactured in different years, the insert materials are probably made from different batches. Our results indicate that the CT-number measurements are robust, as the intraphantom variability was low. There is, however, also a possibility that insert materials’ dilution and their physical properties can drift while aging. Ideally, the different phantoms should be ensured constant in time, but measuring this development is challenging as observed drifting is more likely due to the scanner than the phantom.

Measurements of CT-numbers in this study indicated interphantom variations and showed deviations from predicted values in the manual. One might question precision and accuracy of predicted values as reference values, and should be careful to use the predicted values as absolute thresholds and standard for acceptance criteria. IEC does not have any recommendations for acceptance criteria for CT-numbers. In Table 5 we have summarized our findings.

Our study indicated that low contrast resolution varied between phantoms (Fig. 2). The number of visualized targets also depends on the investigator. In our study, one observer visualized in general more low contrast objects than the two others observers. This might be due to experience in the field. IEC does not recommend this method as an acceptance test [4]. Our results indicate that this test is too subjective, and if the test is included in the QA program, the same physicist should perform the test each time for the same scanner to ensure that interobserver variations do not affect the test results. Ideally, one should use objective measures, however, variability of the phantom and observer’s thresholds and the difficulty of obtaining statistical significant data make it highly difficult to measure low contrast resolution in an objective manner. In our study we also calculated CNR of the 1.0% and 0.5% contrast group in the low contrast module of the phantoms. We found no clear relationship between these values and the numbers of targets visualized by the observers. No statistical significant relationship between the phantoms’ CNR values of the 0.5% and 1.0% group was found. This demonstrates that low contrast resolution is extremely difficult to measure objectively and absolute tolerance levels for low contrast resolution can hardly be made. However, many manufacturers specify the low contrast resolution for their scanners and the QA acceptance program should therefore ideally include such tests. One should aim at developing good objective tests for low contrast resolution. Table 5 shows final results for the low contrast detectability, for three observers and 9 phantoms.

Interphantom, intraphantom and interobserver variations in uniformity were small. The phantom dependence does not seem to have an influence with regard to the recommended standard of a ±4 HU remedial level [3]. QA lite did however yield CT-numbers of the ROIs that diverted from the manual results. Calculated deviations between center and periphery found using this tool were in general much larger than the manually found deviations, in one case almost 4 HU. This is due to the fact that QA lite uses quite small ROIs (diameters approximately 5% of the diameter of the image) compared to manual analysis. In the manual analysis larger ROIs, diameters of 10% of the diameter of the image, were used in agreement with IEC 61223-3-5. The area of the ROIs used by QA lite is hence about four times as small as the IEC recommended size. The results indicate that if one is supposed to follow the IEC standard, one should be aware that the automatic tool QA lite uses smaller ROIs than recommended by IEC, and will not give results in agreement with this standard. One test in the QA protocol is based on comparison with the baseline uniformity, and a level of ±2 HU is suggested [3]. If such precision is required, QA lite analyses of uniformity should be avoided. However, one can discuss which size of the ROIs is appropriate, the QA lite or IEC standard?

Conclusions

The results from this study indicate that there are some interphantom variations for the Catphan phantoms. The variations for low contrast resolution can be large and the evaluation methods are also too subjective. There is a need for development of more objective measurements for low contrast resolution. The interphantom variations in CT-numbers can be interpreted as substantial, and for this test we recommend that one uses the same phantom for acceptance test and the annually QA-tests for one specific scanner. Predicted values from the Catphan manual should be used with care and consideration.

Analyses by the automatic image analysis software QA lite demonstrated that large differences can exist for the image uniformity testing depending on ROI size. One should be aware that QA lite uses smaller ROI’s than recommended by IEC, and will not give results in agreement with this standard. Depending on the chosen regulatory framework one should choose analyzing method with care. Our results are based on tests of 9 phantoms on one specific CT-scanner with one set of scan parameter settings. It is necessary to establish baselines for QA-tests related to the Catphan phantom in general, regardless of scanner models and vendors, for QA-tests in the future.

Conflict of interest

All authors certify that there is no conflict of interest with any financial or personal relationship with other people or organizations.

Acknowledgments

We thank Department of Radiology and Nuclear Medicine, Oslo University Hospital, for use of the Toshiba scanner and assistance with the acquisitions.
References


