Original Investigation

Quantitative Measurements Versus Receiver Operating Characteristics and Visual Grading Regression in CT Images Reconstructed with Iterative Reconstruction: A Phantom Study

Kristin Jensen, MSc, Hilde Kjernlie Andersen, MSc, Örjan Smedby, PhD, Bjørn Helge Østerås, MSc, Anette Aarsnes, MSc, Anders Tingberg, PhD, Erik Fosse, PhD, MD, Anne Catrine Martinsen, PhD

Rationale and Objectives: This study aimed to evaluate the correlation of quantitative measurements with visual grading regression (VGR) and receiver operating characteristics (ROC) analysis in computed tomography (CT) images reconstructed with iterative reconstruction.

Materials and Methods: CT scans on a liver phantom were performed on CT scanners from GE, Philips, and Toshiba at three dose levels. Images were reconstructed with filtered back projection (FBP) and hybrid iterative techniques (ASiR, iDose, and AIDR 3D of different strengths). Images were visually assessed by five readers using a four- and five-grade ordinal scale for liver low contrast lesions and for 10 image quality criteria. The results were analyzed with ROC and VGR. Standard deviation, signal-to-noise ratios, and contrast-to-noise ratios were measured in the images.

Results: All data were compared to FBP. The results of the quantitative measurements were improved for all algorithms. ROC analysis showed improved lesion detection with ASiR and AIDR and decreased lesion detection with iDose. VGR found improved noise properties for all algorithms, increased sharpness with iDose and AIDR, and decreased artifacts from the spine with AIDR, whereas iDose increased the artifacts from the spine. The contrast in the spine decreased with ASiR and iDose.

Conclusions: Improved quantitative measurements in images reconstructed with iterative reconstruction compared to FBP are not equivalent to improved diagnostic image accuracy.

Key Words: CT iterative reconstruction; quantitative measurements; visual grading regression; receiver operating characteristics; diagnostic accuracy.

INTRODUCTION

Iterative reconstruction algorithms decrease image noise in computed tomography (CT) images compared to filtered back projection (FBP) (1–3). In FBP, the image noise is inversely proportional to the square of the radiation dose, but with iterative reconstruction, this relationship is changed. Some iterative algorithms change the image texture, which is shown by the different shape of the noise power spectrum (1,4). The shape of the noise power spectrum can be dose-dependent (5), and thereby influence the relationship between noise and low contrast resolution. Studies have shown that regardless of vendors' claims of dose reduction because of use of iterative reconstruction, low contrast resolution does not benefit from the same improvement as noise (6–10).

Acad Radiol 2017; ■:■■–■■

From the Department of Diagnostic Physics, Oslo University Hospital, P.O. Box, 0454 Oslo, Norway (K.J., H.K.A., B.H.., A.A., A.C.M.); Department of Physics, University of Oslo, P. O. Box 1048 Blindern, N-0316 Oslo, Norway (K.J., A.C.M.); Department of Life Sciences and Health, Oslo and Akershus University College of Applied Sciences, P.O. Box 4 St. Olavs plass, N-0130 Oslo, Norway (K.J.); KTH Royal Institute of Technology / STH, Huddinge, Sweden (.S.); Institute of Clinical Medicine, University of Oslo, Oslo, Norway (B.H.., E.F.); Department of Medical Radiation Physics, Lund University, Skåne University Hospital, Malmö, Sweden (A.T.); The Intervention Centre, Rikshospitalet, Oslo, Norway (E.F.). Received June 8, 2017; revised October 16, 2017; accepted October 26, 2017. **Address correspondence to:** K.J. e-mail: uxjekr@ous-hf.no

https://doi.org/10.1016/j.acra.2017.10.020

In addition to noise, spatial resolution may influence the visibility of small low-contrast objects. Some iterative reconstruction algorithms improve spatial resolution (1,11,12), but there are also studies that show that the spatial resolution can be degraded (1,13). Iterative reconstruction can reduce artifacts such as metal artifacts, beam hardening artifacts, and scattering artifacts (14,15); however, they can also introduce phenomena perceived by viewers as artifacts, like an artificial or blotchy appearance (16,17).

One advantage of iterative reconstruction is that it is easy to implement different models to correct for irregularities in the reconstruction. However, most commercial algorithms appear to the user as black boxes, not making information about models and corrections available to the users. Differences between the algorithms result in differences in noise power spectrum, as well as spatial resolution and artifact reduction (13). It is therefore important to evaluate all the algorithms to ensure diagnostic acceptable images.

Quantitative measurements as standard deviation (SD) of the noise, signal-to-noise ratios (SNR), and contrast-to-noise ratios (CNR) are often chosen to evaluate image quality because of their effectiveness. These evaluations assume that lower noise and better CNR improve diagnostic efficacy. However, these quantitative measures largely ignore changes in noise properties, texture, and spatial resolution introduced by iterative reconstruction. This may influence the relationship between quantitative measurements and diagnostic effectiveness as lesion conspicuity.

Visual grading experiments can be performed on all types of images. Actual diagnostic information present in the images or image quality properties are evaluated visually, and the analysis can be carried out with visual grading regression (VGR) (18,19) or visual grading characteristics (20). This gives information about image quality, but may not necessarily give information about diagnostic effectiveness. The assumption is that visibility of pathology is correlated to the visibility of normal anatomic structures. Careful selection of structures or properties to score is crucial to ensure that the analysis is related to the actual diagnostic outcome.

With access to a reliable reference method ("gold standard"), it is possible to evaluate the diagnostic accuracy of an imaging procedure. Receiver operating characteristics (ROC) analysis (21) is often used for analysis in such studies. However, it may be practically difficult to perform ROC analysis. Even when a "gold standard" is available, studies of this kind are time- and cost-consuming, and still the truthfulness may be questioned. In phantom experiments, however, the ground truth may be available by the construction of the phantoms. However, phantom experiments have other limitations like lack of realistic patient appearance, lack of anatomic details, or that the images always have the same background (if only one phantom is used).

Previous studies have explored the relationship between technical and clinical image quality in general x-ray (22,23) and in CT (24), showing that the correlation is dependent on the imaging task and technical measure. Studies also suggest a nonlinear relationship between technical and visual measures at low doses (25).

The purpose of this study was to evaluate the correlation between quantitative measurements, VGR and ROC, in CT images reconstructed with three different hybrid iterative reconstruction algorithms.

MATERIALS AND METHODS

Phantom

The phantom used in this study was an upper abdominal anthropomorphic phantom custom-made for ROC analysis (St. Bartholomew's Hospital, Clinical Physics Group, London EC1A 7BE, UK) (Figs 1 and 2) (26). The dimensions were 35 cm (lateral direction), 27 cm (anterior-posterior), and 6 cm (superior-inferior). Four cylindrical inserts, each divided into eight sectors, were placed in the liver. Sixteen of these 32 sectors contained lesions with diameters ranging from 2 to 7 mm (mean 4.4 mm, median 4 mm). In this study, the lesions were filled with water, resulting in a contrast difference of about −50 to −30 HU between lesions and liver tissue. In addition to liver inserts, the phantom included structures resembling kidney, pancreas, and spine, three low-contrast objects and two homogenous cylindrical inserts.

Image Acquisition and Reconstruction

Imaging was done using GE LightSpeed VCT (Chicago, IL), Philips Brilliance 64 (Amsterdam, Netherlands), and Toshiba Aquilion One (Otawara, Tochigi, Japan). The phantom was positioned in the isocenter. Scan protocols, shown in Table 1, were derived from the local protocol for liver imaging where mA was adjusted to get $CTDI_{vol}$ of about 5 mGy, 10 mGy,

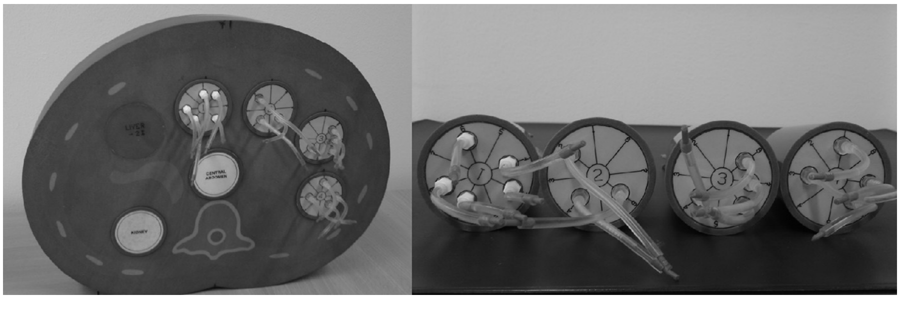

Figure 1. Upper abdominal phantom used in the study (*left*). Four liver inserts with lesions in 16 of 32 sectors (*right*).