Radiology: Imaging Cancer

ORIGINAL RESEARCH

Agreement between Routine-Dose and Lower-Dose CT with and without Deep Learning–based Denoising for Active Surveillance of Solid Small Renal Masses: A Multiobserver Study

Jens Borgbjerg, MD, PhD[1] • *Bendik Stensby Breen, MD*[1] • *Cathrine Helgestad Kristiansen, BSc*[1] • *Nis Elbrønd Larsen, MD*[2] • *Lise Medrud, MD*[2] • *Rasa Mikalone, MD*[3] • *Stig Müller, MD, PhD*[4] • *Gintare Naujokaite, MD*[3] • *Anne Negård, MD, PhD*[1] • *Tommy Kjærgård Nielsen, MD, PhD*[5] • *Ivar Mjåland Salte, MD, PhD*[1] • *Jens Brøndum Frøkjær, MD, PhD*[3]

Author affiliations, funding, and conflicts of interest are listed at the end of this article. See also commentary by Muglia in this issue.

Radiology: Imaging Cancer 2025; 7(4):e240250 • https://doi.org/10.1148/rycan.240250 • Content codes: GU AI OI CT

Purpose: To assess the agreement between routine-dose (RD) and lower-dose (LD) contrast-enhanced CT scans, with and without Digital Imaging and Communications in Medicine–based deep learning–based denoising (DLD), in evaluating small renal masses (SRMs) during active surveillance.

Materials and Methods: In this retrospective study, CT scans from patients undergoing active surveillance for an SRM were included. Using a validated simulation technique, LD CT images were generated from the RD images to simulate 75% (LD75) and 90% (LD90) radiation dose reductions. Two additional LD image sets, in which the DLD was applied (LD75-DLD and LD90-DLD), were generated. Between January 2023 and June 2024, nine radiologists from three institutions independently evaluated 350 CT scans across five datasets for tumor size, tumor nearness to the collecting system (TN), and tumor shape irregularity (TSI), and interobserver reproducibility and agreement were assessed using the 95% limits of agreement with the mean (LOAM) and Gwet AC2 coefficient, respectively. Subjective and quantitative image quality assessments were also performed.

Results: The study sample included 70 patients (mean age, 73.2 years ± 9.2 [SD]; 48 male, 22 female). LD75 CT was found to be in agreement with RD scans for assessing SRM diameter, with a LOAM of ±2.4 mm (95% CI: 2.3, 2.6) for LD75 compared with ±2.2 mm (95% CI: 2.1, 2.4) for RD. However, a 90% dose reduction compromised reproducibility (LOAM ±3.0 mm; 95% CI: 2.8, 3.2). LD90-DLD preserved measurement reproducibility (LOAM ±2.4 mm; 95% CI: 2.3, 2.6). Observer agreement was comparable between TN and TSI assessments across all image sets, with no statistically significant differences identified (all comparisons $P \geq .35$ for TN and $P \geq .02$ for TSI; Holm-corrected significance threshold, $P = .013$). Subjective and quantitative image quality assessments confirmed that DLD effectively restored image quality at reduced dose levels: LD75-DLD had the highest overall image quality, significantly lower noise, and improved contrast-to-noise ratio compared with RD ($P < .001$).

Conclusion: A 75% reduction in radiation dose is feasible for SRM assessment in active surveillance using CT with a conventional iterative reconstruction technique, whereas applying DLD allows submillisievert dose reduction.

Supplemental material is available for this article.

Incidence of kidney cancer, the 14th most common cancer worldwide, is rising, with more than 434 840 new cases reported in 2022 (1). Many of these cancers are detected incidentally as small, localized, asymptomatic masses (2). Contemporary imaging techniques cannot reliably and uniformly distinguish aggressive renal tumors from benign tumors or those with low malignant potential (3). Renal mass biopsy has a relatively high nondiagnostic rate, and benign biopsy does not necessarily indicate that a nonmalignant entity is present (4). As a result, surgical excision is widely considered the first-line choice for managing small renal masses (SRMs) measuring less than 4 cm (cT1a stage), with thermal ablation serving as a minimally invasive alternative (5). To address the limitations of current diagnostic and treatment approaches, active surveillance has emerged as a viable and endorsed management alternative in international guidelines (2,5). Active surveillance entails monitoring renal tumor size with serial imaging with delayed treatment in case of progression (6). This approach has traditionally been reserved for patients who are elderly, comorbid, and frail with limited life expectancy, for whom the risks associated with surgery outweigh the potential oncologic benefits. Nevertheless, recent studies conclude that active surveillance principles can be safely applied in younger patients (ie, ≤ 60 years at diagnosis) (7,8). Evidence suggests that clinicians may overemphasize primary intervention for SRMs, potentially leading to overtreatment (9,10).

The most common imaging metrics prompting crossover from active surveillance to delayed intervention are tumor growth rate based on maximum axial diameter of more than 5 mm/year and absolute tumor size of more than 4 cm (8,11). In this context, contrast-enhanced CT is the standard imaging modality because of its relatively high reliability, speed, and low cost (11). Furthermore, CT is more precise than US and more accessible than MRI, and it is recommended that the same imaging modality be used throughout the active surveillance period (11). However, the radiation dose incurred in serial CT imaging of SRMs is substantial and a cause of concern (12–14). There is an unmet need to reduce radiation dose while preserving the accuracy and reproducibility needed in radiographic measurements such that small changes in renal mass size can be appreciated. Despite recommended use of low-dose CT more than a decade ago (15), few studies have investigated its feasibility for active surveillance of SRMs, and the utility of deep learning–based techniques for the improvement of image quality for active surveillance has yet to be explored.

Abbreviations

DICOM = Digital Imaging and Communications in Medicine, DLD = deep learning–based denoising, LD = lower dose, LD75 = lower dose with 75% radiation dose reduction, LD75-DLD = lower dose with 75% radiation dose reduction with DLD, LD90 = lower dose with 90% radiation dose reduction, LD90-DLD = lower dose with 90% radiation dose reduction with DLD, LOAM = limits of agreement with the mean, RD = routine dose, ROI = region of interest, SRM = small renal mass, TN = tumor nearness to the collecting system or sinus, TSI = tumor shape irregularity

Summary

Interobserver agreement for CT-based assessment of small renal masses during active surveillance was similar between routine-dose and lower-dose images acquired using iterative reconstruction, and deep learning–based denoising enabled dose reduction to submillisievert levels.

Key Points

- Interobserver limits of agreement for CT-based size assessment of 70 small renal masses were similar for routine-dose (limits of agreement with the mean [LOAM] ±2.2 mm), lower-dose (75% reduction) with iterative reconstruction (LOAM ±2.4 mm), and 90% radiation dose-reduced images with deep learning–based denoising (LOAM ±2.4 mm).
- Agreement (Gwet AC2) was similar for tumor nearness to the collecting system (all $P \geq .35$) and tumor shape irregularity (all $P \geq .02$; Holm-corrected threshold, $P = .013$) assessments across the three datasets.
- Subjective image quality was rated higher for lower-dose (75% reduction) CT images with deep learning–based denoising applied compared with routine-dose images ($P < .001$).

Keywords

CT, Urinary, Kidney, Radiation Safety, Observer Performance, Technology Assessment

This multiobserver, multi-institutional study primarily aimed to assess the level of agreement between routine-dose (RD) and lower-dose (LD) contrast-enhanced CT scans, the latter with and without deep learning–based denoising (DLD), for assessing the maximum diameter of SRMs in an active surveillance setting. Secondary aims included evaluating the accuracy and reproducibility of active surveillance-relevant parameters of tumor nearness to the collecting system and shape irregularity and comparing subjective and quantitative image quality, including contrast-to-noise ratio.

Materials and Methods

The study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (reference 285666) and by the institutional data protection officer with a waiver for informed patient consent. ClariPI provided access to the commercial software ClariCT.AI and technical support for simulated low-dose CT. However, the authors maintained complete control of the data and the information submitted for publication at all times.

Patients and Data Acquisition

We queried the picture archiving and communication system at Akershus University Hospital in Lorenskog, Norway, to identify consecutive CT scans of adult patients (≥18 years) undergoing active surveillance for an SRM (≥25% of the mass is composed of enhancing tissue [16]) from January 2015 to December 2021. Exclusion criteria included infiltrative and noncircumscribed masses, predominantly cystic masses with less than 25% enhancing tissue, masses containing macroscopic fat, masses measuring less than 1 cm or greater than 4 cm, severe respiratory artifacts, and noncontrast CT examinations (Fig 1). An abdominal radiologist (J.B., with 12 years of CT experience) assessed patient eligibility.

An upper abdomen CT protocol was used with a body weight–adapted contrast medium volume of iohexol 350 mgI/mL (Omnipaque 350; GE HealthCare) at a dosage of 2 mL/kg. The injection duration was 35 seconds with an 80-second delay. Scans were performed on a Philips Ingenuity Core or Brilliance iCT scanner (Philips Healthcare). The scan parameters were as follows: tube potential, 120 kV; collimation, 64/128 × 0.625 mm; rotation time, 0.5 second; pitch, 1.2; matrix, 512 × 512; section thickness, 0.9 mm; and increment, 0.45 mm. Automatic tube current (DoseRight 3D-DOM; Philips Healthcare) was enabled, and the dose right index was set at 22. All images were reconstructed with 3-mm section axial images using the hybrid iterative reconstruction algorithm iDose (4) (level 2 or 3) with filter B (soft tissue).

Low-Dose Simulation

We used a simulation method for generating low-dose CT images that does not require raw CT data and instead uses sinogram synthesis from RD Digital Imaging and Communications in Medicine (DICOM) images. This choice was motivated by radiation protection considerations, as the method eliminates the need for additional low-dose CT acquisitions, thereby minimizing patient exposure to radiation. This validated vendor-agnostic technique has been shown to provide realistic dose-reduced images concerning noise magnitude and textural appearance, which are indistinguishable from real low-dose CT (17,18). Using the two above CT systems, we obtained multiple phantom (Catphan-600; The Phantom Laboratory) measurements to extract the main features needed to generate a model for noise generation. Based on the reference RD CT examinations, datasets were generated corresponding to examinations at 75% (LD75) and 90% (LD90) dose reductions. This level of dose reduction was based on a prior exploratory study suggesting that 75% dose reduction is feasible in CT-based size assessment of SRMs (19).

Deep Learning–based Image Reconstruction (Denoising)

The low-dose CT images were then reconstructed using a commercially available image-based DLD model (ClariCT.AI; ClariPi) (20). According to the manufacturer, the DLD is based on a U-Net type convolutional neural network that has been trained using more than 1 million pairs of noise-added and original CT images using 24 scanner models from different CT vendors (GE HealthCare, Siemens, Philips Healthcare, and Canon Medical Systems) covering various reconstruction conditions to acquire a vendor-agnostic denoising capability. The DLD focuses exclusively on reducing noise in the image domain and was trained to limit noise reduction to a level that preserves image texture (21,22). The model's performance has been evaluated in several clinically oriented studies (23–28).

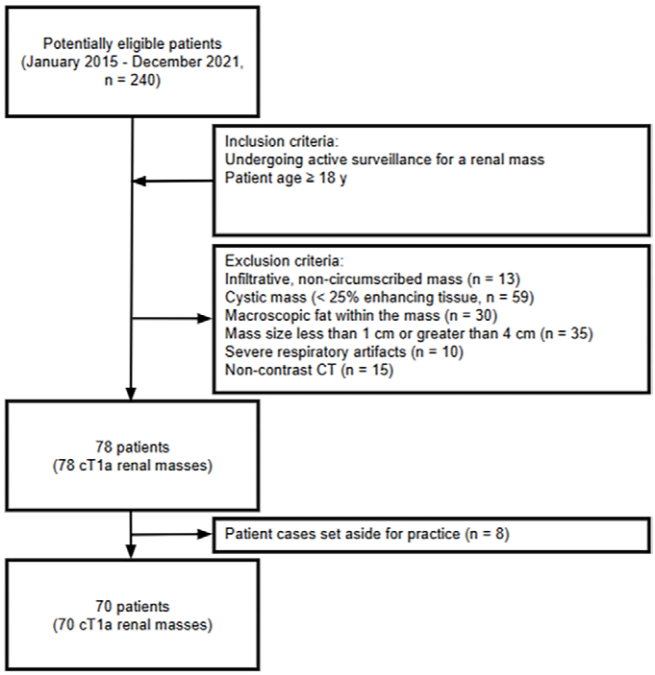

Figure 1: Study flowchart illustrates patient inclusion and exclusion criteria. cT1a = clinical stage, tumor confined to the kidney, less than 4 cm.

Applying a high denoising strength (level 9 of 10) and a standard kernel of the DLD, two additional sets of images were generated: LD75-DLD and LD90-DLD. This addition provided a total of five datasets for the analyses (Fig 2).

Observers

Eight radiology consultants (J.B., N.E.L., L.M., R.M., G.N., A.N., I.M.S., and J.B.F.) and one 4th-year resident radiologist (B.S.B.) from three university-based hospitals (Akershus University Hospital in Norway and Aarhus University Hospital and Aalborg University Hospital in Denmark) with a minimum of 3 years of experience in reading abdominal CT scans were recruited and independently analyzed the images. The radiologists had a mean age of 44.4 years ± 7.4 (SD) (range, 36–61 years) and a mean CT experience of 13.4 years ± 7.8 (range, 3–30 years).

Image Analysis

The anonymized CT datasets were loaded into a web-based platform facilitating observer performance studies in imaging research as used elsewhere (29). This platform offers a setting where image readings can be carried out through an internet browser with automatic data registration. The platform was combined with a multiplanar reconstruction-capable web-based DICOM viewer (Fig 3) (30).

CT Scan Assessments

The 70 reference CT scans were reconstructed into four additional datasets (LD75, LD90, LD75-DLD, and LD90-DLD), for a total of 350 CT scans that were analyzed across five sessions by the nine observers. To minimize memory bias, we used a mixed-order setup in which 70 scans were read at a time with at least a 2-week washout period between sessions (31). CT scans with different dose levels with and without DLD were divided between the five reading sessions, and a CT case of each patient appeared only once per session (Fig 4). Observers were aware that the study was performed to evaluate scans with different dose levels and reconstruction methods but were blinded to the specific dose levels and algorithms as well as any clinical details and scan assessments by the other observers. Written case assessment instructions and videos demonstrating the web-based DICOM viewer functionality were given to the observers. All observers were trained with eight CT test cases encompassing all dose levels to familiarize themselves with the DICOM viewer and the parameters of the case report form, including subjective image quality (Fig 3).

Observers determined the maximal tumor diameter using the available on-the-fly multiplanar reconstruction capability as well as windowing and leveling at the discretion of each observer. However, the caliper for assessing the maximal tumor diameter could be placed in only the axial plane, in accordance with the measurement approach most commonly used for growth assessment in prior studies evaluating active surveillance of SRMs (11). To identify the renal mass for evaluation, the DICOM viewer displayed a green circle above the renal mass's *z* level.

The observers also assessed the ordinal parameters of tumor nearness (TN) and tumor shape irregularity (TSI). TN is defined as the nearness of the tumor to the collecting system or sinus in millimeters in three tiers (≤4, >4 but <7, and ≥7 mm) (32). It is well established as an important predictor of overall complications and postoperative hemorrhage after nephron-sparing surgery, risk of malignancy, and adverse pathology (33). TSI is based on overall tumor shape and contour and has been shown to predict adverse pathologic and oncologic outcomes preoperatively (34). The three TSI categories are defined as 1, a completely elliptical shape with no apparent protrusions along the tumor contour; 2, an approximately elliptical shape with only focal (<50% of the entire circumference) and minor protrusions; and 3, a nonelliptical shape showing extensive (≥50% of the whole circumference) and/or major protrusions.

Finally, subjective image quality was assessed by the observers. Using a five-point scale, overall image quality (1 = poor, 2 = fair, 3 = good, 4 = very good, and 5 = excellent), presence of noise (1 = very noisy, 2 = definitely noisy, 3 = slightly noisy, 4 = minimal noise, and 5 = no perceivable noise), and diagnostic confidence in terms of delineating the contour of the renal mass (1 = poor, 2 = fair, 3 = good, 4 = very good, and 5 = excellent) were rated.

Quantitative Image Quality Assessment

All assessments were performed by a consultant radiologist (J.B.) with 12 years of experience. In the RD datasets at a section thickness of 3 mm, a circular region of interest (ROI) with an area of approximately 0.5 mm^2 was placed in the subcutaneous fat of the anterior abdominal wall and the psoas major muscle at the level of the renal mass. An additional ROI encompassing approximately two-thirds of the mass's area was placed for homogeneous renal masses. For heterogeneous masses, the ROI encompassed approximately two-thirds of the mass's most hyperattenuating area but was required to measure at least 5 mm in diameter. Furthermore, an ROI was placed on the renal cortex ipsilateral to the renal mass (typical

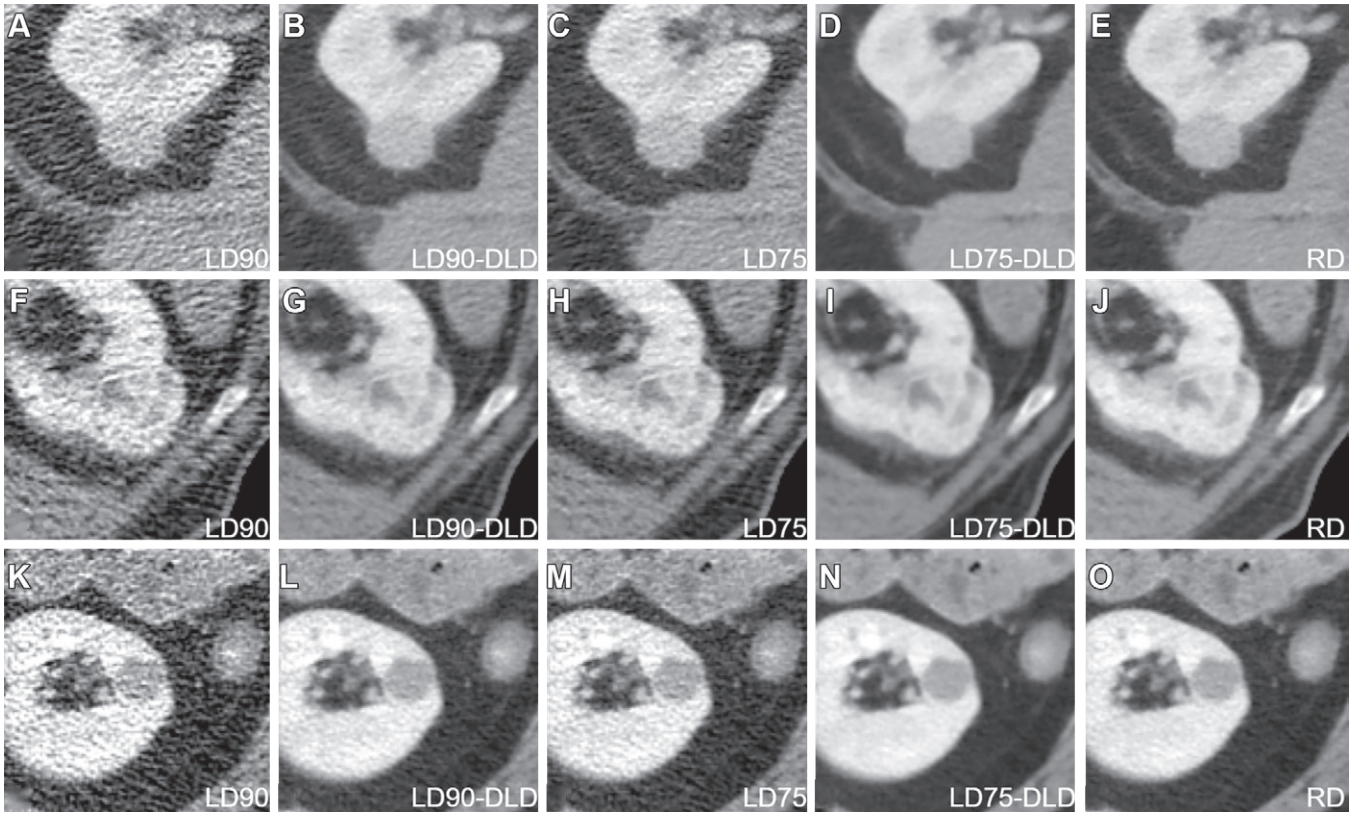

Figure 2: Contrast-enhanced CT images in the axial plane (3-mm section thickness) of three different small renal masses at various dose levels: lower dose with 90% dose reduction (LD90), lower dose with 90% dose reduction with deep learning–based denoising model (LD90-DLD), lower dose with 75% dose reduction (LD75), lower dose with 75% dose reduction with deep learning–based denoising model (LD75-DLD), and routine dose (RD). **(A–E)** A 2.2-cm, right-sided, well-circumscribed, homogeneous, exophytic small renal mass in a 70-year-old female patient. **(F–J)** A 3.2-cm, left-sided, heterogeneous, partly exophytic small renal mass with a contour protrusion in a 65-year-old male patient. **(K–O)** A 1.6-cm, left-sided, homogeneous, endophytic small renal mass in a 67-year-old male patient. Note the increase in image noise at lower-dose levels, whereas the denoising model reduces noise while maintaining image texture.

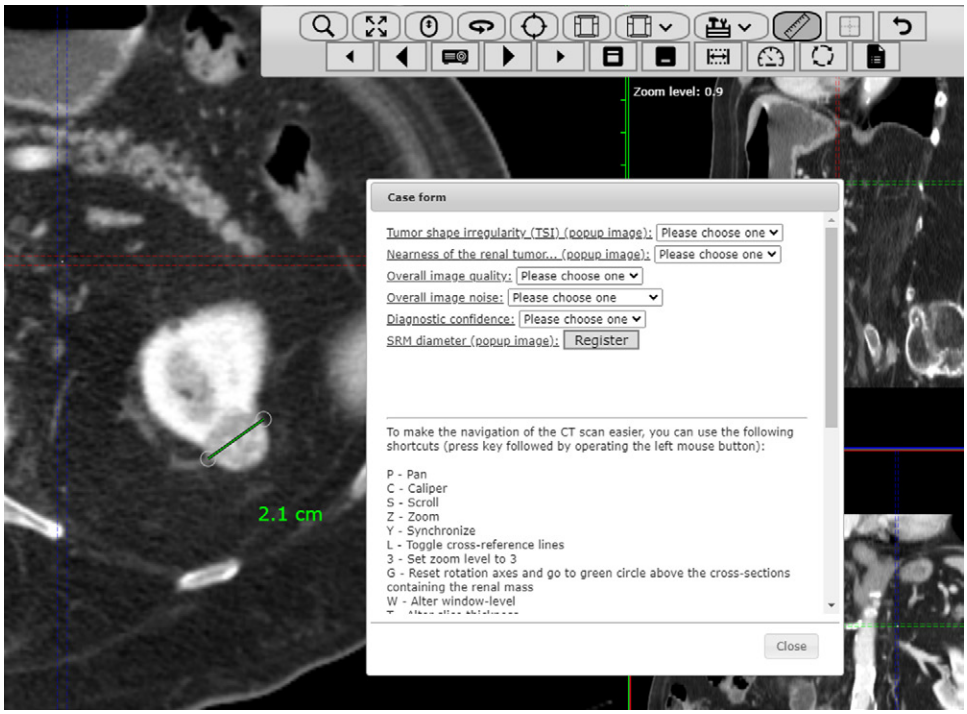

Figure 3: Screenshot of the web-based Digital Imaging and Communications in Medicine viewer equipped with a case report form. A 2.1-cm exophytic small renal mass in the left kidney is depicted on a routine-dose contrast-enhanced CT image.

Session #1		Session #2		Session #3		Session #4		Session #5	
Case 1-14	RD	Case 1-14	LD90-DLD	Case 1-14	LD90	Case 1-14	LD75-DLD	Case 1-14	LD75
Case 15-28	LD75	Case 15-28	RD	Case 15-28	LD90-DLD	Case 15-28	LD90	Case 15-28	LD75-DLD
Case 29-42	LD75-DLD	Case 29-42	LD75	Case 29-42	RD	Case 29-42	LD90-DLD	Case 29-42	LD90
Case 43-56	LD90	Case 43-56	LD75-DLD	Case 43-56	LD75	Case 43-56	RD	Case 43-56	LD90-DLD
Case 57-70	LD90-DLD	Case 57-70	LD90	Case 57-70	LD75-DLD	Case 57-70	LD75	Case 57-70	RD

Figure 4: Overview of the case assessment scheme in a mixed-order setup with at least a 2-week washout period between the five sessions, each encompassing 70 renal tumor scans at different dose levels with and without deep learning–based denoising. The selection sequence of the sessions was randomized for each of the nine observers, and the sequence of scans within each session was also randomized. LD75 = lower dose with 75% dose reduction, LD75-DLD = lower dose with 75% dose reduction with deep learning–based denoising model, LD90 = lower dose with 90% dose reduction, LD90-DLD = lower dose with 90% dose reduction with deep learning–based denoising model, RD = routine dose.

ROI size, 5-mm diameter) using the same axial section used for renal mass attenuation.

Finally, a square ROI of 64 × 64 pixels was placed in the liver parenchyma, carefully avoiding the visible bile ducts and vessels.

These circular ROIs and the square ROIs were copied to the other datasets at the same positions to ensure consistency. From the circular ROIs, the image noise, defined as the SD of pixel values, was obtained in addition to the contrast-to-noise ratio, calculated as follows: contrast-to-noise ratio (renal mass to renal cortex) = $HU_{Renal\ cortex} - HU_{Renal\ mass} / SD_{Fat}$.

To assess variation of noise texture across datasets, noise power spectrum analysis was performed using the square ROIs to calculate the average noise power spectrum spatial frequency (mm^{-1}) for each dataset (35). Additional details about the calculation of the noise power spectrum are provided in Appendix S1.

Statistical Analysis

Statistical analyses were performed using Python (version 3.9.14; *https://www.python.org/*).

The normality of quantitative data was assessed graphically and with the Shapiro-Wilk test and expressed as means ± SDs or medians with IQRs as appropriate. We used the limits of agreement with the mean (LOAM) formulated by Christensen et al (36) to assess observer agreement for tumor diameter measurements in a multiobserver setup. The LOAM represents how much an observer's measurement may plausibly deviate from the mean of all observers' measurements on the specific subject (ie, constituting reproducibility). Based on the data from a prior exploratory study on the reproducibility of size assessment of renal masses using LD CT (19), we found it satisfactory to obtain an approximate width of 0.5 mm of the 95% CIs of the LOAM. Using these data and the method by Christensen et al (36), we estimated that nine observers and at least 40 tumor CT scans would be required.

The repeated measures analysis of variance, paired *t* test, and Wilcoxon signed rank test were performed to examine the null hypothesis of differences between the image sets in relation to continuous values of quantitative image quality. Ordinal assessment was significant at 0.05.

The agreement between all nine observers for ordinal parameters was evaluated using Gwet AC2 with its 95% CI with *z* tests to compare agreement between image sets (37,38). Gwet AC2 accounts for weighted agreement, making it particularly suited for ordinal data by assigning different penalties based on the extent of disagreement between raters. This approach ensures that minor disagreements (eg, one category apart) are weighted less heavily than major ones. An estimate of less than or equal to 0.4 was considered poor agreement, 0.41–0.60 as fair, 0.61–0.80 as good, and 0.81 or greater as excellent (37). The Holm method was used to adjust *P* value thresholds for significance to account for multiple comparisons, resulting in corrected thresholds of .013, .017, .025, and .05 for the four comparisons at an overall significance level of .05 (39).

Results

Patient Characteristics

The study sample consisted of 70 patients (mean age, 73.2 years ± 9.2 [range, 46–88 years]; 48 male, 22 female) (Fig 1). Detailed patient characteristics are summarized in Table 1.

Renal Tumor Assessments

The nine observers completed a total of 3150 CT scan assessments from the five datasets (350 assessments per observer, combined across all observers). The mean and 95% LOAM for the diameter assessments are summarized in Table 2. The mean SRM diameters, averaged across all observers, were highly consistent across the datasets: 19.5 mm (95% CI: 17.6, 21.5) for RD, 19.5 mm (95% CI: 17.6, 21.5) for LD75, 19.4 mm (95% CI: 17.4, 21.3) for LD75-DLD, 19.3 mm (95% CI: 17.4, 21.3) for LD90, and 19.5 mm (95% CI: 17.5, 21.4) for LD90-DLD (analysis of variance, $P > .99$). The 95% LOAM for tumor diameter measurements were ±2.2 mm (95% CI: 2.1, 2.4) for RD, ±2.4 mm (95% CI: 2.3, 2.6) for LD75, ±2.3 mm (95% CI: 2.2, 2.5) for LD75-DLD, ±3.0 mm (95% CI: 2.8, 3.2) for LD90, and ±2.4 mm (95% CI: 2.3, 2.6) for LD90-DLD. Hence, the LOAM values for RD, LD75, LD75-DLD, and LD90-DLD were similar, with overlapping 95% CIs. In contrast, the LOAM for LD90 was significantly wider, as evidenced by its nonoverlapping 95% CI compared with the other datasets. The five agreement plots in Figure 5 give no indication of heteroscedasticity associated with the size of the renal masses. In alignment with the small interobserver variance components observed in Table 2, no observer systematically performed unusually small or large measurements. *P* values for categorical variables and quantitative image quality assessment between RD and the other sets of images are listed in Table S1.

Table 3 provides an overview of the evaluation of ordinal parameters by the nine observers. The median scores of TN and TSI were lower for LD90 compared with RD, with TN medians of 1.0 (IQR, 1.0–3.0) for LD90 and 2.0 (IQR, 1.0–3.0) for RD and TSI medians of 1.0 (IQR, 1.0–1.0) for LD90 and 1.0 (IQR, 1.0–2.0) for RD ($P < .001$ for both comparisons). The agreement

Table 1: Patient Demographics, Tumor Characteristics, and CT Radiation Exposure Data (Routine-Dose Dataset)

Variable	Data (n = 70)
Patient characteristic	
Age (y)	73.2 ± 9.2 (46–88)
Sex	
Female	22
Male	48
Body mass index	27.7 ± 5.4 (13.0–40.5)
Tumor characteristic	
Mean tumor size (mm)	19.5 ± 8.3 (10.0–40.0)
TN (mm)	
≤4	28 (40)
>4 but <7	8 (11)
≥7	34 (49)
TSI	
1	51 (73)
2	14 (20)
3	5 (7)
Radiation dose	
Mean scan length (cm)*	28.1 ± 4.7 (20.4–50.4)
Mean DLP (mGy · cm)	437.0 ± 207.5 (144–1353)
Mean $CTDI_{vol}$ (mGy)	15.3 ± 5.5 (6.3–28.8)
Effective dose (mSv)†	6.4 ± 3.1 (2.2–20.3)

Note.—Data are numbers with percentages in parentheses or means ± SDs with ranges in parentheses. Body mass index is calculated as weight in kilograms divided by height in meters squared. Mean tumor size, tumor nearness to the collecting system or sinus (TN), and tumor shape irregularity (TSI) are based on the routine-dose dataset, with means reflecting the combined assessments of all nine observers. TSI is defined as 1, a completely elliptical shape with no apparent protrusions along the tumor contour; 2, an approximately elliptical shape with only focal (<50% of the entire circumference) and minor protrusions; and 3, a nonelliptical shape showing extensive (≥50% of the whole circumference) and/or major protrusions. $CTDI_{vol}$ = volume CT dose index, DLP = dose length product.

* From the diaphragm to the iliac crest.

† Effective dose = DLP × abdominal weighting factor (0.015 mSv · mGy^{-1} · cm^{-1}).

between all nine observers for TN was good for all image sets, with similar Gwet AC2 values of LD image sets compared with RD (all $P \geq .35$). The agreement for TSI was good for all image sets, with the exception of LD90, which showed excellent agreement, though we found no evidence of a difference (all comparisons $P \geq .02$; Holm-corrected threshold for significance of .013 for RD compared with LD90).

Subjective Image Quality Assessment

Table 3 summarizes the subjective image quality assessment. Regarding overall image quality, observers rated LD75-DLD highest (median score, 4.0; IQR, 3.0–5.0), RD next (median score, 4.0; IQR, 3.0–4.0), and LD90 lowest (median score, 2.0; IQR, 1.0–2.0); scores were significantly different between all image sets ($P < .001$). The same order of ratings for the image sets was observed for image noise ($P < .001$), whereas for diagnostic confidence, LD75-DLD and RD were rated similarly ($P = .68$). Observer agreement ranged from fair to good, with the highest agreement across datasets observed for diagnostic confidence.

Quantitative Image Quality Assessment

The results of quantitative measurements are shown in Table 4. Compared with the RD dataset, both the 75% and 90% dose-reduced datasets reconstructed with hybrid iterative reconstruction showed increased noise across all interrogated tissues—for example, renal mass noise increased from 14.5 HU ± 4.3 for RD to 27.4 HU ± 7.3 for LD75 and 51.1 HU ± 13.5 for LD90, while contrast-to-noise ratio decreased from 10.3 ± 6.4 for RD to 4.9 ± 3.1 for LD75 and 2.6 ± 1.7 for LD90. However, applying DLD to these LD images resulted in more than a twofold noise reduction compared with their iterative-only counterparts: renal mass noise was reduced to 12.8 HU ± 6.0 for LD75-DLD and 20.3 HU ± 6.6 for LD90-DLD. In alignment with the subjective assessment of overall image quality and noise, the LD75-DLD demonstrated superiority in terms of noise level and contrast-to-noise ratio compared with the RD dataset (for all tissues, $P < .001$ for both). Similarly, the LD90-DLD had lower noise and higher contrast-to-noise ratio

Table 2: Mean Renal Mass Diameter and 95% LOAM of 350 Measurements by Nine Observers for Five CT Datasets

Dataset	Mean (mm)	LOAM (mm)	σ_A (mm)	σ_B (mm)	σ_E (mm)
RD	19.5 (17.6, 21.5)	±2.2 (2.1, 2.4)	8.3 (6.9 to 9.7)	0.2 (0.0, 0.3)	1.2 (1.1, 1.3)
LD75	19.5 (17.6, 21.5)	±2.4 (2.3, 2.6)	8.2 (6.9, 9.6)	0.1 (0.0, 0.3)	1.3 (1.2, 1.4)
LD75-DLD	19.4 (17.39, 21.33)	±2.3 (2.2, 2.5)	8.4 (7.0, 9.8)	0.2 (0.0, 0.3)	1.3 (1.3, 1.3)
LD90*	19.3 (17.39, 21.25)	±3.0 (2.8, 3.2)	8.2 (6.8, 9.6)	0.2 (0.0, 0.4)	1.6 (1.5, 1.7)
LD90-DLD	19.5 (17.51, 21.38)	±2.4 (2.3, 2.6)	8.3 (6.9, 9.6)	0.1 (−0.1, 0.3)	1.3 (1.2, 1.4)

Note.—Data in parentheses are 95% CIs. The limits of agreement with the mean (LOAM) integrates the intersubject (σ_A), interobserver (σ_B), and residual (σ_E) variance components in millimeters. LD75 = lower dose with 75% dose reduction, LD75-DLD = lower dose with 75% dose reduction with deep learning–based denoising model, LD90 = lower dose with 90% dose reduction, LD90-DLD = lower dose with 90% dose reduction with deep learning–based denoising model, RD = routine dose.

* The LOAM for LD90 demonstrates a significant widening compared with other datasets, indicating reduced reproducibility at this dose level.

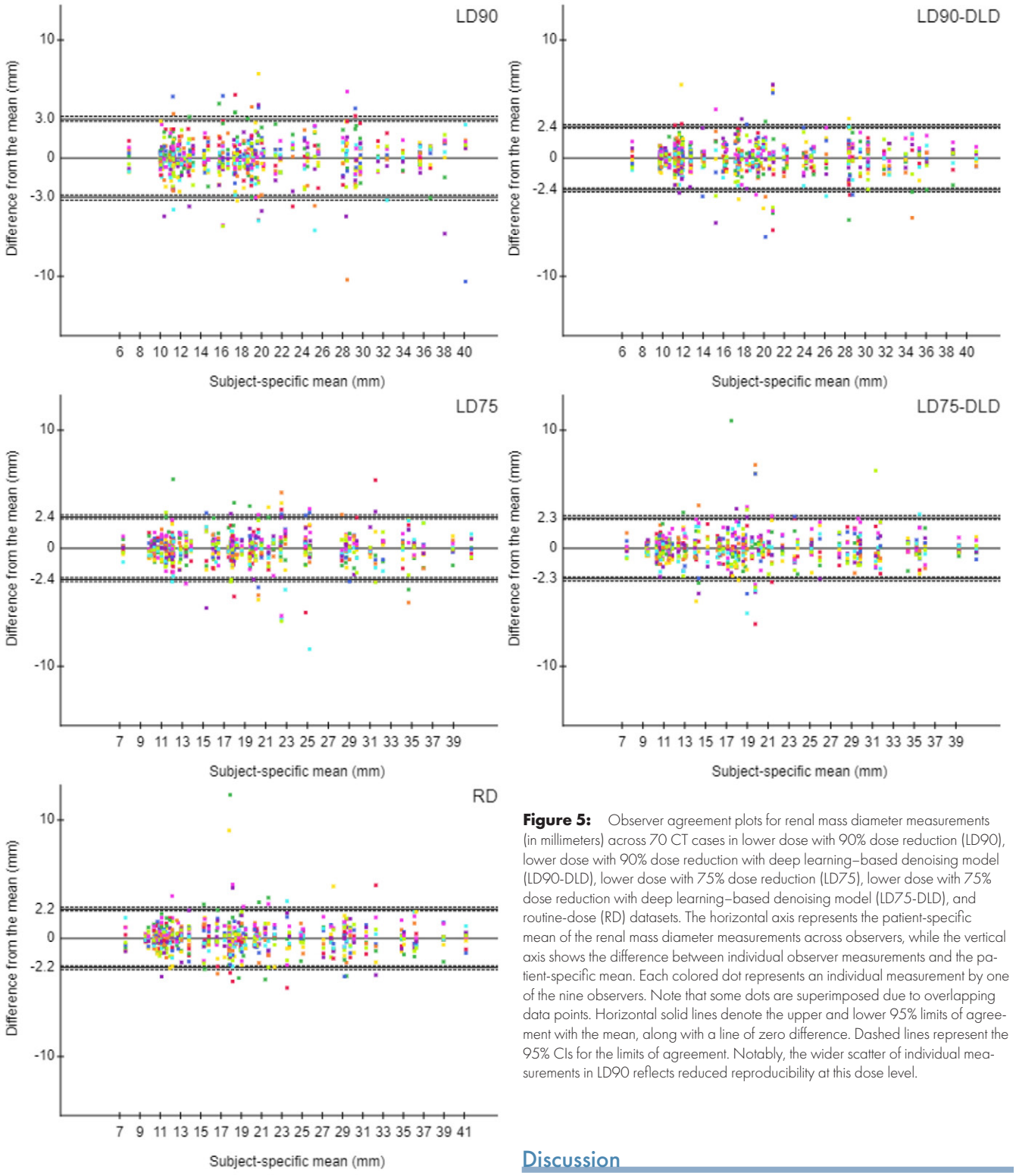

Figure 5: Observer agreement plots for renal mass diameter measurements (in millimeters) across 70 CT cases in lower dose with 90% dose reduction (LD90), lower dose with 90% dose reduction with deep learning–based denoising model (LD90-DLD), lower dose with 75% dose reduction (LD75), lower dose with 75% dose reduction with deep learning–based denoising model (LD75-DLD), and routine-dose (RD) datasets. The horizontal axis represents the patient-specific mean of the renal mass diameter measurements across observers, while the vertical axis shows the difference between individual observer measurements and the patient-specific mean. Each colored dot represents an individual measurement by one of the nine observers. Note that some dots are superimposed due to overlapping data points. Horizontal solid lines denote the upper and lower 95% limits of agreement with the mean, along with a line of zero difference. Dashed lines represent the 95% CIs for the limits of agreement. Notably, the wider scatter of individual measurements in LD90 reflects reduced reproducibility at this dose level.

compared with the LD75 dataset (for all tissues $P < .001$ and $P < .001$). The average noise power spectrum spatial frequency remained similar with the application of the DLD at 75% and 90% dose reductions, showing only minor changes of −0.02 mm^{-1} and 0.02 mm^{-1}, respectively, compared with the RD dataset ($P < .001$).

Discussion

Active surveillance of SRMs, commonly based on CT, has become a widely accepted clinical practice. However, adopting this approach raises concerns regarding the cumulative radiation dose patients receive from repeated CT imaging. This multiobserver agreement study revealed several key findings: 75% radiation dose-reduced CT with hybrid iterative reconstruction demonstrated agreement with RD CT in assessing the maximum diameter of SRMs under active surveillance (RD vs LD75, 95% LOAM ±2.2 mm [95% CI: 2.1, 2.4] vs ±2.4 mm [95% CI: 2.3,

Table 3: Assessment of Tumor Nearness and Tumor Shape Irregularity and Subjective Image Quality by Nine Observers for Five CT Datasets

Variable	RD	LD75	LD75-DLD	LD90	LD90-DLD
TN					
Score	2.0 (1.0–3.0)	2.0 (1.0–3.0)	2.0 (1.0–3.0)	1.0 (1.0–3.0)	2.0 (1.0–3.0)
Gwet AC2	0.78 (0.71, 0.85)	0.77 (0.69, 0.85)	0.79 (0.71, 0.86)	0.76 (0.68, 0.84)	0.73 (0.65, 0.81)
TSI					
Score	1.0 (1.0–2.0)	1.0 (1.0–2.0)	1.0 (1.0–2.0)	1.0 (1.0–1.0)	1.0 (1.0–2.0)
Gwet AC2	0.70 (0.63, 0.77)	0.76 (0.69, 0.83)	0.66 (0.58, 0.74)	0.81 (0.75, 0.87)	0.71 (0.63, 0.78)
Overall image quality					
Score	4.0 (3.0–4.0)	3.0 (2.0–4.0)	4.0 (3.0–5.0)	2.0 (1.0–2.0)	3.0 (3.0–4.0)
Gwet AC2	0.56 (0.53, 0.59)	0.48 (0.44, 0.52)	0.38 (0.3, 0.42)	0.58 (0.55, 0.62)	0.47 (0.43, 0.52)
Noise					
Score	4.0 (4.0–5.0)	3.0 (2.0–3.0)	4.0 (4.0–5.0)	1.0 (1.0–2.0)	3.0 (3.0–4.0)
Gwet AC2	0.52 (0.47, 0.56)	0.55 (0.52, 0.59)	0.53 (0.49, 0.57)	0.71 (0.68, 0.74)	0.34 (0.29, 0.40)
Diagnostic confidence					
Score	4.0 (3.0–4.0)	3.0 (3.0–4.0)	4.0 (3.0–5.0)	3.0 (2.0–3.0)	3.0 (3.0–4.0)
Gwet AC2	0.50 (0.46, 0.54)	0.45 (0.39, 0.51)	0.44 (0.40, 0.49)	0.42 (0.37, 0.46)	0.43 (0.38, 0.47)

Note.—Scores are medians with IQRs in parentheses; Gwet AC2 coefficients are shown with 95% CI in parentheses. LD75 = lower dose with 75% dose reduction, LD75-DLD = lower dose with 75% dose reduction with deep learning–based denoising model, LD90 = lower dose with 90% dose reduction, LD90-DLD = lower dose with 90% dose reduction with deep learning–based denoising model, RD = routine dose, TN = tumor nearness to the collecting system or sinus, TSI = tumor shape irregularity.

Table 4: Quantitative Image Quality Assessment by Nine Observers for Five CT Datasets

Variable	RD	LD75	LD75-DLD	LD90	LD90-DLD
Noise (HU)					
Renal mass	14.5 ± 4.3	27.4 ± 7.3	12.8 ± 6.0	51.1 ± 13.5	20.3 ± 6.6
Renal cortex	14.2 ± 4.5	26.5 ± 7.0	12.3 ± 5.1	50.0 ± 13.7	20.0 ± 5.8
Muscle	12.6 ± 2.8	28.5 ± 5.8	10.0 ± 4.9	56.3 ± 11.7	20.2 ± 5.7
Fat	7.6 ± 2.2	15.8 ± 4.4	6.3 ± 2.1	29.9 ± 8.6	11.5 ± 3.3
CNR	10.3 ± 6.4	4.9 ± 3.1	12.5 ± 7.8	2.6 ± 1.7	6.8 ± 4.2
NPS_{fav} of liver (mm^{-1})	0.22 ± 0.04	0.25 ± 0.03	0.20 ± 0.04	0.26 ± 0.03	0.24 ± 0.04

Note.—Data are presented as means ± SDs. CNR = contrast-to-noise ratio (renal mass to renal cortex), LD75 = lower dose with 75% dose reduction, LD75-DLD = lower dose with 75% dose reduction with deep learning–based denoising model, LD90 = lower dose with 90% dose reduction, LD90-DLD = lower dose with 90% dose reduction with deep learning–based denoising model, NPS_{fav} = noise power spectrum average spatial frequency, RD = routine dose.

2.6]). A 90% dose reduction relative to RD compromised the reproducibility of these measurements (LOAM ±3.0 mm; 95% CI: 2.8, 3.2). However, applying a DLD algorithm at a 90% dose reduction maintained the reproducibility (LOAM ±2.4 mm; 95% CI: 2.3, 2.6). The assessment of the ordinal parameters of TN and TSI demonstrated similar agreement across all image sets (all $P \geq .35$ for TN and $P \geq .02$ for TSI; Holm-corrected threshold for significance of .013 for RD vs LD90, respectively).

Our findings on qualitative and quantitative image quality evaluation corroborate prior research findings that deep learning models can preserve image quality and texture even at significantly reduced dose levels when applied at the CT raw data level or in the image domain (40). However, experience with iterative reconstruction techniques has shown that improved subjective image quality does not always equate to enhanced clinical task performance. Therefore, new imaging techniques, such as deep learning–based reconstruction, require contextual evaluation involving clinically relevant imaging tasks (41). For instance, Jensen et al (42) found that the ability to detect small liver lesions, a low-contrast task, was compromised at a 65% dose reduction with deep learning–based CT reconstruction. Conversely, Zhang et al (43) demonstrated that the high-contrast task of renal stone detection was maintained at a 77% dose reduction using deep learning–based reconstruction compared with low-dose levels reconstructed with hybrid iterative reconstruction. Moreover, a 14-fold dose reduction in noncontrast CT compared with CT angiography has been demonstrated feasible for the task of assessing abdominal aortic diameter (44).

Research by Borgbjerg et al (19) showed that contrast-enhanced low-dose CT yielded similar agreement to RD CT (mean CT dose index-volume, 12.6 mGy) for assessing maximum renal tumor diameter, with comparable reproducibility of diameter measurements to the well-known study by Punnen et al (45). However, the Borgbjerg et al (19) study was limited by a relatively small sample size (n = 40) from Siemens CT systems and used either filtered-back projection or hybrid iterative reconstruction (SAFIRE). Additionally, the study used a low-dose simulation technique based on simple noise addition to DICOM images, approximating a 75% dose reduction. The study did not evaluate renal tumor morphology, which, along with tumor size, is part of the American College of Radiology imaging recommendations (46). Hence, our study adds to the existing literature by validating the findings of Borgbjerg et al (19), demonstrating that the accuracy and reproducibility of size assessment of SRMs has substantial resistance to decreased image quality in LD CT.

Furthermore, we observed resistance in assessing TN and TSI and demonstrated that applying the DLD allows for size assessment with a 90% dose reduction relative to RD levels. The DLD was applied in conjunction with CT images using iterative reconstruction, and it is compatible with both older and newer CT systems without necessitating manufacturer-specific hardware or software, making it potentially widely applicable. The results of our study have considerable implications for the safety of active surveillance in terms of radiation exposure. Based on recommendations from a review of imaging protocols for active surveillance of SRMs (47) and using an average dose of 6.4 mSv per CT scan derived from the RD dataset, the cumulative radiation dose for a 10-year active surveillance regimen (13 CT scans) is estimated at 83 mSv. This total exceeds the suggested 55-mSv threshold, above which cancer risk increases substantially, as a recent review indicates (48). However, applying hypothetical dose reductions of 75% and 90% would lower the cumulative dose to 20.8 mSv and 8.3 mSv, respectively, offering a promising strategy to reduce associated risks.

In terms of growth detection, the observed LOAMs in this study align with the reproducibility reported in the influential study by Punnen et al (45) widely referenced in clinical guidelines, except for LD90, where the slightly wider LOAM introduces additional variability. Importantly, even at the extremes of the CIs, the LOAMs for LD75, LD75-DLD, and LD90-DLD remain within the clinically significant threshold of more than 5 mm/year, commonly used to guide delayed intervention decisions in active surveillance. We recognize that a lack of statistically significant difference between evaluated datasets does not inherently establish equivalence. However, our findings provide evidence that the above dose reductions are feasible for the clinical task of renal mass active surveillance. Hence, the robustness of these findings within established clinical standards implies interchangeability regarding clinical applicability rather than strict equivalence. Yet, regardless of CT dose level, caution is warranted when measurements approach critical intervention thresholds. Variability in such instances may lead to unnecessary referrals for resection or ablation or delay appropriate interventions.

Our study had limitations. First, a minimum 2-week washout period between reading sessions might not entirely eliminate recall bias; however, a memory-mitigating strategy involving a mixed-order reading scheme was used to address this. Second, our study included CT scans from only two systems by a single vendor. However, regarding the generalization of findings, the CT dose reduction potential and image quality in abdominal applications have previously been comparable to iDose in two other hybrid interactive reconstruction algorithms by General Electric and Siemens, suggesting the generalizability of these findings (49). In addition, the DLD model applies to multiple CT systems and is comparable in dose reduction potential to vendor-specific deep learning–based reconstruction (40). Third, the low-dose simulation method has been extensively studied and validated in a real-world pig study, and it has been used in prior CT low-dose studies (17). Still, simulated data may only partially reflect the finesses of real-world clinical imaging situations. Fourth, we did not evaluate cystic renal masses, which are also frequently surveilled. However, surveillance of cystic masses relies on the Bosniak classification, with a more rigorously defined emphasis on changes in mass morphology rather than absolute change in the size of the cystic mass (16). The use of LD CT for this purpose is an avenue for future investigations. Finally, although the maximum axial diameter was used to assess tumor size for consistency and alignment with clinical guidelines and prior studies, we recognize that this approach does not fully account for the complexity of irregularly shaped lesions. Future studies may benefit from exploring volumetric techniques, including segmentation, but these methods were not included here to maintain consistency with prior literature and clinical applicability.

In conclusion, our findings demonstrate that a 75% dose-reduced CT obtained using an iterative reconstruction technique is in agreement with RD CT for the purpose of assessing SRMs under active surveillance. Furthermore, applying a DLD model allows for a 90% reduction in radiation exposure with similar performance. Future clinical guidelines should reflect the feasibility of substantial dose reductions with conventional CT and denoising methods, addressing concerns regarding radiation exposure.

Author affiliations:

[1] Department of Radiology, Akershus University Hospital, Sykehusveien 25, 1478 Lorenskog, Norway

[2] Department of Radiology, Aarhus University Hospital, Aarhus, Denmark

[3] Department of Radiology, Aalborg University Hospital, Aalborg, Denmark

[4] Department of Urology, Akershus University Hospital, Lorenskog, Norway

[5] Department of Urology, Aalborg University Hospital, Aalborg, Denmark

Received August 9, 2024; revision requested October 23; revision received March 29, 2025; accepted April 30.

Address correspondence to: J.B. (email: jens.borgbjerg@ahus.no).

Funding: This study was funded by Akershus University Hospital.

Acknowledgments: We would like to express our gratitude to radiologists Mats Kleivane and Niklas Revold Grønli of Akershus University Hospital for their valuable contributions to this study. Their thorough evaluation of the web-based platform for case evaluation was instrumental in ensuring the robustness and reliability of the observer agreement study.

Author contributions: Guarantors of integrity of entire study, **J.B., J.B.F.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **J.B., R.M., J.B.F.**; clinical studies, **R.M., A.N., J.B.F.**; experimental studies, **J.B., B.S.B., N.E.L., L.M., R.M., S.M., G.N., I.M.S., J.B.F.**; statistical analysis, **J.B., R.M.**; and manuscript editing, **J.B., B.S.B., C.H.K., N.E.L., L.M., R.M., G.N., A.N., T.K.N., I.M.S., J.B.F.**

Data sharing: The data supporting this study's findings are available from the corresponding author, upon reasonable request. However, the data are not publicly available due to privacy and ethical restrictions.

Disclosures of conflicts of interest: J.B. No relevant relationships. **B.S.B.** No relevant relationships. **C.H.K.** Part-time employee of Philips Heathcare as an application specialist. **N.E.L.** No relevant relationships. **L.M.** No relevant relationships. **R.M.** No relevant relationships. **S.M.** No relevant relationships. **G.N.** No relevant relationships. **A.N.** No relevant relationships. **T.K.N.** No relevant relationships. **I.M.S.** No relevant relationships. **J.B.F.** No relevant relationships.

References

1. Kidney cancer statistics. WCRF International. https://www.wcrf.org/cancer-trends/kidney-cancer-statistics/. Accessed March 22, 2025.
2. Wilcox Vanden Berg RN, Basourakos SP, LaRussa S, McClure TD. Management of the Small Renal Mass: A 2020 Update. Curr Oncol Rep 2020;22(7):69.
3. Reynolds AM, Porter KK. Characterizing Indeterminate Renal Masses with Molecular Imaging: The Role of 99mTc-MIBI SPECT/CT. Curr Urol Rep 2017;18(11):86.
4. Patel HD, Johnson MH, Pierorazio PM, et al. Diagnostic Accuracy and Risks of Biopsy in the Diagnosis of a Renal Mass Suspicious for Localized Renal Cell Carcinoma: Systematic Review of the Literature. J Urol 2016;195(5):1340–1347.
5. Jiang P, Ali SN, Peta A, et al. A Review of the Recommendations and Strength of Evidence for Clinical Practice Guidelines on the Management of Small Renal Masses. J Endourol 2023;37(8):903–913.
6. Sebastià C, Corominas D, Musquera M, Paño B, Ajami T, Nicolau C. Active surveillance of small renal masses. Insights Imaging 2020;11(1):63.
7. Cheung DC, Martin LJ, Komisarenko M, McAlpine K, Alibhai SMH, Finelli A. A Matched Analysis of Active Surveillance Versus Nephrectomy for T1a Small Renal Masses. Eur Urol Oncol 2023;6(5):535–539.
8. Metcalf MR, Cheaib JG, Biles MJ, et al. Outcomes of Active Surveillance for Young Patients with Small Renal Masses: Prospective Data from the DISSRM Registry. J Urol 2021;205(5):1286–1293.
9. Shah PH, Alom MA, Leibovich BC, et al. The Temporal Association of Robotic Surgical Diffusion with Overtreatment of the Small Renal Mass. J Urol 2018;200(5):981–988.
10. Sohlberg EM, Metzner TJ, Leppert JT. The Harms of Overdiagnosis and Overtreatment in Patients with Small Renal Masses: A Mini-review. Eur Urol Focus 2019;5(6):943–945.
11. Rebez G, Pavan N, Mir MC. Available active surveillance follow-up protocols for small renal mass: a systematic review. World J Urol 2021;39(8):2875–2882.
12. Borgbjerg J, Bylling T, Andersen G, Thygesen J, Mikkelsen A, Nielsen TK. CT-guided cryoablation of renal cancer: radiation burden and the associated risk of secondary cancer from procedural- and follow-up imaging. Abdom Radiol (NY) 2020;45(11):3581–3588.
13. Mataac M, Rehani MM. Is a one percent occurrence of high-dose patients significant? Eur J Radiol 2024;172:111340.
14. Mataac MT, Li X, Rehani MM. What proportion of CT scan patients are alive or deceased after 10 years? Eur J Radiol 2024;178:111629.
15. Meng MV. Editorial Comment. Urology 2012;79(1):31.
16. Silverman SG, Pedrosa I, Ellis JH, et al. Bosniak Classification of Cystic Renal Masses, Version 2019: an Update Proposal and Needs Assessment. Radiology 2019;292(2):475–488.
17. Brendlin AS, Wrazidlo R, Almansour H, et al. How Real Are Computed Tomography Low Dose Simulations? An Investigational In-Vivo Large Animal Study. Acad Radiol 2023;30(8):1678–1694.
18. Won Kim C, Kim JH. Realistic simulation of reduced-dose CT with noise modeling and sinogram synthesis using DICOM CT images. Med Phys 2014;41(1):011901.
19. Borgbjerg J, Larsen NE, Salte IM, Grønli NR, Klæstrup E, Negård A. Dataset on renal tumor diameter assessment by multiple observers in normal-dose and low-dose CT. Data Brief 2023;51:109672.
20. Kim JH. Apparatus and method for denoising CT images. US Patent (2017). https://patents.google.com/patent/US9852527B2/en. Accessed March 22, 2025.
21. Ahn CK, Kim JH, Yang Z, Heo C, Jin H, Park B. A deep learning-enabled iterative reconstruction of ultra-low-dose CT: use of synthetic sinogram-based noise simulation technique. In: Chen GH, Lo JY, Gilat Schmidt T, eds. Medical Imaging 2018: Physics of Medical Imaging. SPIE, 2018; 112. https://doi.org/10.1117/12.2294013.
22. Choi H, Chang W, Kim JH, et al. Dose reduction potential of vendor-agnostic deep learning model in comparison with deep learning–based image reconstruction algorithm on CT: a phantom study. Eur Radiol 2022;32(2):1247–1255.
23. Lim WH, Choi YH, Park JE, et al. Application of Vendor-Neutral Iterative Reconstruction Technique to Pediatric Abdominal Computed Tomography. Korean J Radiol 2019;20(9):1358–1367.
24. Kolb M, Storz C, Kim JH, et al. Effect of a novel denoising technique on image quality and diagnostic accuracy in low-dose CT in patients with suspected appendicitis. Eur J Radiol 2019;116:198–204.
25. Lee S, Choi YH, Cho YJ, et al. Noise reduction approach in pediatric abdominal CT combining deep learning and dual-energy technique. Eur Radiol 2021;31(4):2218–2226.
26. Choi HU, Cho J, Hwang J, et al. Diagnostic performance and image quality of an image-based denoising algorithm applied to radiation dose-reduced CT in diagnosing acute appendicitis. Abdom Radiol (NY) 2024;49(6):1839–1849.
27. Lee DH, Lee JM, Lee CH, Afat S, Othman A. Image Quality and Diagnostic Performance of Low-Dose Liver CT with Deep Learning Reconstruction versus Standard-Dose CT. Radiol Artif Intell 2024;6(2):e230192.
28. Kapper C, Müller L, Kronfeld A, et al. Value of vendor-agnostic deep learning image denoising in brain computed tomography: a multi-scanner study. Rofo 2025;197(1):65–75.
29. Borgbjerg J, Steinkohl E, Olesen SS, et al. Inter- and intra-observer variability of computed tomography-based parenchymal- and ductal diameters in chronic pancreatitis: a multi-observer international study. Abdom Radiol (NY) 2023;48(1):306–317.
30. Borgbjerg J. Web-based imaging viewer for real-color volumetric reconstruction of human visible project and DICOM datasets. Clin Anat 2021;34(3):470–477.
31. Haygood TM, Smith S, Sun J. Memory bias in observer-performance literature. J Med Imaging 2018;5(3):031412.
32. Alsaikhan N, Alshehri W, Cassidy F, Aganovic L, Vahdat N. Renal tumor structured reporting including nephrometry score and beyond: what the urologist and interventional radiologist need to know. Abdom Radiol (NY) 2019;44(1):190–200.
33. Correa AF, Toussi A, Amin M, et al. Small Renal Masses in Close Proximity to the Collecting System and Renal Sinus Are Enriched for Malignancy and High Fuhrman Grade and Should Be Considered for Early Intervention. Clin Genitourin Cancer 2018;16(4):e729–e733.
34. Tanaka H, Fukuda S, Kimura K, et al. Defining Tumour Shape Irregularity for Preoperative Risk Stratification of Clinically Localised Renal Cell Carcinoma. Eur Urol Open Sci 2023;48:36–43.
35. Samei E, Bakalyar D, Boedeker KL, et al. Performance evaluation of computed tomography systems: summary of AAPM Task Group 233. Med Phys 2019;46(11):e735–e756.
36. Christensen HS, Borgbjerg J, Børty L, Bøgsted M. On Jones et al.'s method for extending Bland-Altman plots to limits of agreement with the mean for multiple observers. BMC Med Res Methodol 2020;20(1):304.
37. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol 2008;61(Pt 1):29–48.
38. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013;13:61.
39. Haynes W. Holm's Method. Encyclopedia of Systems Biology. Springer, 2013; 902. https://doi.org/10.1007/978-1-4419-9863-7_1214.
40. Koetzier LR, Mastrodicasa D, Szczykutowicz TP, et al. Deep Learning Image Reconstruction for CT: Technical Principles and Clinical Prospects. Radiology 2023;306(3):e221257.
41. Toia GV. Editorial Comment: Deep Learning Image Reconstruction—Do Better Images Make a Better Radiologist? AJR Am J Roentgenol 2023;220(3):388.
42. Jensen CT, Gupta S, Saleh MM, et al. Reduced-Dose Deep Learning Reconstruction for Abdominal CT of Liver Metastases. Radiology 2022;303(1):90–98.
43. Zhang X, Zhang G, Xu L, et al. Application of deep learning reconstruction of ultra-low-dose abdominal CT in the diagnosis of renal calculi. Insights Imaging 2022;13(1):163.
44. Borgbjerg J, Christensen HS, Al-Mashhadi R, et al. Ultra-low-dose non-contrast CT and CT angiography can be used interchangeably for assessing maximal abdominal aortic diameter. Acta Radiol Open 2022;11(10):20584601221132461.
45. Punnen S, Haider MA, Lockwood G, Moulding F, O'Malley ME, Jewett MAS. Variability in Size Measurement of Renal Masses Smaller Than 4 cm on Computerized Tomography. J Urol 2006;176(6 Pt 1):2386–2390; discussion 2390.
46. Herts BR, Silverman SG, Hindman NM, et al. Management of the Incidental Renal Mass on CT: A White Paper of the ACR Incidental Findings Committee. J Am Coll Radiol 2018;15(2):264–273.
47. Liaw CW, Winoker JS, Mehrazin R. Imaging Protocols for Active Surveillance in Renal Cell Carcinoma. Curr Urol Rep 2018;19(10):81.
48. Cao CF, Ma KL, Shan H, et al. CT Scans and Cancer Risks: A Systematic Review and Dose-response Meta-analysis. BMC Cancer 2022;22(1):1238.
49. Andrabi Y, Pianykh O, Agrawal M, Kambadakone A, Blake MA, Sahani DV. Radiation Dose Consideration in Kidney Stone CT Examinations: Integration of Iterative Reconstruction Algorithms With Routine Clinical Practice. AJR Am J Roentgenol 2015;204(5):1055–1063.