EISSN 1305-3612

DIR

Diagnostic and Interventional Radiology

TSR 1924
TURKISH SOCIETY OF RADIOLOGY

dirjournal.org

VOLUME 30
ISSUE 4
JULY 2024

Diagnostic and Interventional Radiology

Official Journal of the Turkish Society of Radiology

E-ISSN: 1305-3612

www.dirjournal.org

Diagnostic and Interventional Radiology

Official Journal of the Turkish Society of Radiology

E-ISSN: 1305-3612
www.dirjournal.org

Diagnostic and Interventional Radiology (Diagn Interv Radiol) is a bimonthly periodical of the Turkish Society of Radiology and the content of the journal is available at https://www.dirjournal.org/. It is peer-reviewed and adheres to the highest ethical and editorial standards. The editors of the journal endorse the Editorial Policy Statements Approved by the Council of Science Editors Board of Directors (https://cse.memberclicks.net/). The journal is in compliance with the Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals published by the International Committee of Medical Journal Editors (updated May 2022, www.icmje.org).

First ten volumes of Diagnostic and Interventional Radiology have been published in Turkish under the name of Tanısal ve Girişimsel Radyoloji (Index Medicus® abbreviation: Tani Girisim Radyol), the current title's exact Turkish translation.

Please refer to the journal's webpage **(https://dirjournal.org/)** for "Aims and Scope", "Instructions to Authors" and "Instructions to Reviewers".

The editorial and publication processes of the journal are shaped in accordance with the guidelines of the ICMJE, WAME, CSE, COPE, EASE, and NISO.
Diagnostic and Interventional Radiology is indexed in **SCI-Expanded, Pubmed/Medline, Pubmed Central, TUBITAK ULAKBIM TR Index, DOAJ, HINARI, EMBASE, CINAHL, Scopus, Gale and CNKI.**

The journal is published online.

Owner: Can Çevikol on behalf of Turkish Society of Radiology

Responsible Manager: Mehmet Ruhi Onur

Contact Information
Diagnostic and Interventional Radiology Turkish Society of Radiology
Hoşdere Cad., Güzelkent Sok., Çankaya Evleri, F/2, 06540
Ankara, Türkiye
E-mail: info@dirjournal.org
Phone: +90 (312) 442 36 53 **Fax:** +90 (312) 442 36 54

Publisher Contact
Address: Molla Gürani Mah. Kaçamak Sk.
No: 21/1 34093 İstanbul, Türkiye
Phone: +90 (530) 177 30 97
E-mail: info@galenos.com.tr/yayin@galenos.com.tr
Web: www.galenos.com.tr **Publisher Certificate Number:** 14521
Online Publication Date: July 2024
EISSN 1305-3612
International scientific journal published bimonthly.

Turkish Society of Radiology is one of the foremost medical specialty organizations in Türkiye. It was formed by the merger of the two main radiology societies of Türkiye, one of which was founded in 1924. The Society is based in Ankara, Türkiye.

TSR 1924 TURKISH SOCIETY OF RADIOLOGY

Diagnostic and Interventional Radiology

Volume 30 • Issue 4

July 2024

www.dirjournal.org

Contents

Full text of these articles can be accessed online at www.dirjournal.org or through PubMed (https://www.ncbi.nlm.nih.gov/pmc/journals/2754/).

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.232601

ABDOMINAL IMAGING

ORIGINAL ARTICLE

Preoperative CT and MRI assessment of the longitudinal tumor extent of extrahepatic bile duct cancer after biliary drainage

Seo-Bum Cho[1]
Yeun-Yoon Kim[1,2]
June Park[1]
Hye Jung Shin[3]

[1]Severance Hospital, Yonsei University College of Medicine, Department of Radiology and Research Institute of Radiological Science, Seoul, Republic of Korea

[2]Samsung Medical Center, Sungkyunkwan University School of Medicine, Department of Radiology and Center for Imaging Sciences, Seoul, Republic of Korea

[3]Yonsei University College of Medicine, Department of Biomedical Systems Informatics, Biostatistics Collaboration Unit, Seoul, Republic of Korea

Corresponding author: Yeun-Yoon Kim

E-mail: cookieks35@gmail.com

Received 22 November 2023; revision requested 11 December 2023; accepted 15 January 2024.

Epub: 20.02.2024

Publication date: 08.07.2024

DOI: 10.4274/dir.2024.232601

PURPOSE

To examine the diagnostic performance for the longitudinal extent of extrahepatic bile duct (EHD) cancer on computed tomography (CT) after biliary drainage (BD) and investigate the appropriate timing of magnetic resonance imaging (MRI) acquisition.

METHODS

This retrospective study included patients who underwent curative-intent surgery for EHD cancer and CT pre- and post-BD between November 2005 and June 2021. The biliary segment-wise longitudinal tumor extent was evaluated according to the 2019 Korean Society of Abdominal Radiology consensus recommendations, with pre-BD CT, post-BD CT, and both pre- and post-BD CT. The performance for tumor detectability was compared using generalized estimating equation (GEE) method. When preoperative MRI was performed, patients were divided into two subgroups according to the timing of MRI with respect to BD, and the performance of MRI obtained pre- and post-BD was compared.

RESULTS

In 105 patients (mean age: 67 ± 8 years; 74 men and 31 women), the performance for tumor detectability was superior using both CT scans compared with using post-BD CT alone (reader 1: sensitivity, 72.6% vs. 64.6%, $P < 0.001$; specificity, 96.9% vs. 94.8%, $P = 0.063$; reader 2: sensitivity, 77.2% vs. 72.9%, $P = 0.126$; specificity, 97.5% vs. 94.2%, $P = 0.003$), and it was comparable with using pre-BD CT alone. In biliary segments with a catheter, higher sensitivity and specificity were observed using both CT scans than using post-BD CT (reader 1: sensitivity, 74.4% vs. 67.5%, $P = 0.006$; specificity, 92.4% vs. 88.0%, $P = 0.068$; reader 2: sensitivity, 80.5% vs. 74.4%, $P = 0.013$; specificity, 94.3% vs. 88.0%, $P = 0.016$). Post-BD MRI (n = 30) exhibited a comparable performance to pre-BD MRI (n = 55) (reader 1: sensitivity, 77.9% vs. 75.0%, $P = 0.605$; specificity, 97.2% vs. 94.9%, $P = 0.256$; reader 2: sensitivity, 73.2% vs. 72.6%, $P = 0.926$; specificity, 98.4% vs. 94.9%, $P = 0.068$).

CONCLUSION

Pre-BD CT provided better diagnostic performance in the preoperative evaluation of EHD cancer. The longitudinal tumor extent could be accurately assessed with post-BD MRI, which was similar to pre-BD MRI.

CLINICAL SIGNIFICANCE

The acquisition of pre-BD CT could be beneficial for the preoperative evaluation of EHD cancer when BD is planned. Post-BD MRI would not be significantly affected by BD in terms of the diagnostic performance of the longitudinal tumor extent.

KEYWORDS

Extrahepatic cholangiocarcinoma, computed tomography, magnetic resonance imaging, biliary tract surgical procedures, drainage

Extrahepatic bile duct (EHD) cancer originates below the intrahepatic secondary biliary confluence and encompasses perihilar and distal bile duct cancers.[1] This type of cancer constitutes the majority of cholangiocarcinoma, with a mortality rate below 2 in 100,000 person-year.[2] Surgical resection is the sole curative treatment for it, underscoring the critical

You may cite this article as: Cho SB, Kim YY, Park J, Shin HJ. Preoperative CT and MRI of the longitudinal tumor extent of extrahepatic bile duct cancer after biliary drainage. *Diagn Interv Radiol.* 2024;30(4):212-219.

need for an accurate assessment of surgical resectability at the initial diagnosis.[3] Despite the pivotal role of the longitudinal extent of the tumor in determining the surgical approach, there is a lack of international radiologic reporting guidelines for EHD cancer. Recently, the Korean Society of Abdominal Radiology (KSAR) published consensus recommendations for the structured radiologic reporting of EHD cancer using computed tomography (CT) or magnetic resonance imaging (MRI).[4] Recent studies have demonstrated the utility of these recommendations in assessing resectability in EHD cancer.[5,6]

Approximately 90% of patients with EHD cancer initially present with cholangitis due to biliary obstruction.[7] Consequently, urgent endoscopic or percutaneous biliary drainage (BD) is necessary when biliary infection is suspected.[7-10] Moreover, BD can induce inflammation in the bile duct, mimicking or obscuring EHD cancer on CT or MRI and posing challenges in the imaging evaluation of the longitudinal tumor extent.[11] A previous study indicated a higher frequency of achieving no residual tumor (R0) resection in perihilar cholangiocarcinoma for patients who underwent CT before BD compared with those who underwent CT after BD. This shows the difficulty in assessing the exact longitudinal tumor extent post-BD.[12] While previous studies have suggested the challenges of post-BD imaging evaluation in perihilar cholangiocarcinoma,[13,14] there is currently no published data examining the impact of BD on the diagnostic performance of imaging studies in an intraindividual manner. Therefore, it is essential to evaluate the diagnostic performance of CT evaluation after BD to devise a more effective strategy for reporting these examinations before the curative resection of EHD cancer. Furthermore, given the limited usage of MRI compared with CT, it is crucial to investigate if BD affects tumor extent evaluation using MRI. A study involving 26 patients observed less accurate performances of MRI after BD in perihilar cholangiocarcinoma, which lacked a statistical comparison and warranted further investigation.[15]

Main points

- Determining surgical resectability is crucial in extrahepatic bile duct cancer.
- Biliary drainage (BD) obscures the longitudinal tumor extent on cross-sectional imaging.
- Evaluation of pre-BD computed tomography improved the diagnostic performance.
- Similar performance was observed in magnetic resonance imaging pre- and post-BD.

This study aims primarily to examine the diagnostic performance of evaluating the longitudinal extent of EHD cancer after BD and improve performance in the preoperative setting. The secondary aim is to determine the appropriate timing of MRI acquisition concerning BD for evaluating the longitudinal extent of EHD cancer.

Methods

Patients

The Severance Hospital Institutional Review Board approved this study (IRB no: 4-2021-1139, date: 10.05.2021) and waived the requirement for patient consent because this study involved a retrospective review of medical records and images. This study identified potentially eligible patients with EHD cancer who underwent curative-intent surgery at our institution between November 2005 and June 2021 (Figure 1). The inclusion criteria were as follows: (a) BD performed before surgery, (b) contrast-enhanced CT scans obtained both pre- and post-BD, and (c) age of patients >18 years. The exclusion criteria were as follows: (a) preoperative embolization of the hepatic artery or portal vein causing metal artifacts on CT scans, (b) preoperative chemotherapy or chemoradiation hampering radiologic–pathologic correlation, (c) palliative surgery, and (d) lack of a reference standard for the longitudinal tumor extent.

Image acquisition

Multiphase or single-portal venous-phase CT images were acquired according to institutional routine protocols. Patients underwent CT using either a 16- or 64-channel scanner (Somatom Sensation 16 or 64, Siemens, Erlangen, Germany; Brilliance iCT or 64, Philips Healthcare, Cleveland, Ohio, U.S.A.; Lightspeed VCT, GE Medical Systems, Milwaukee, Wisconsin, U.S.A.). After obtaining unenhanced images, 120–150 mL of non-ionic contrast medium was administered intravenously at a rate of 2–5 mL/sec. Using the bolus-tracking technique, the arterial phase was obtained 10–25 seconds after the attenuation value reached 100 Hounsfield units at the abdominal aorta. The portal venous and delayed phases were obtained 70–80 seconds and 3 minutes after contrast injection, respectively. The scanning parameters were as follows: beam collimation, 0.625 or 0.75 mm; slice thickness, 3 or 5 mm; reconstruction interval, 3 or 5 mm; rotation time, 0.5 seconds; effective tube current-time charge, 150–250 mAs; and tube voltage, 100–120 kVp. Coronal images were reconstructed with a slice thickness of 3 or 5 mm and a reconstruction interval of 3 or 5 mm. The routine protocol for MRI with MR cholangiopancreatography is described in the Online Resource 1.

Computed tomography analysis

The CT images were retrospectively and independently reviewed by two readers (S.B.C., a radiology resident, and Y.Y.K., a board-certified abdominal radiologist with 5 years of practice experience) who were blinded to the surgical and pathological findings. Pre-BD CT scans, post-BD CT scans, and both pre- and post-BD CT scans were examined in each image review session with a washout period of at least 2 weeks between sessions to reduce recall bias. Image analysis was performed based on the 2019 KSAR consensus recommendations.[4] Moreover, the

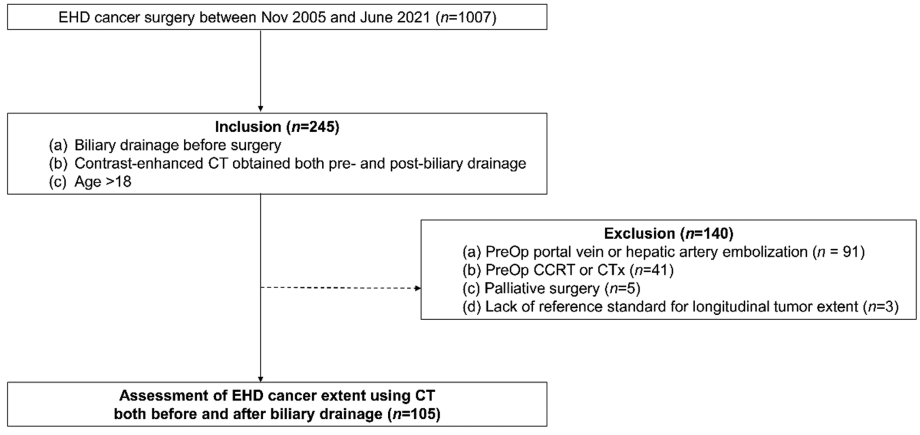


Figure 1. Patient flow diagram. Out of 1,007 consecutive patients who underwent EHD cancer surgery between November 2005 and June 2021 at our institution, 245 met the inclusion criteria. After following the exclusion process, 105 patients who underwent CT both pre- and post-biliary drainage were finally included. EHD, extrahepatic bile duct cancer; Op, operation; CCRT, concurrent chemoradiotherapy; CTx, chemotherapy; CT, computed tomography.

readers recorded the presence or absence of tumors in each segment of the biliary tree (both secondary confluences, both hepatic ducts, primary confluence of the bile duct, common hepatic duct, and supra- and intra-pancreatic common bile duct) and determined the Bismuth–Corlette classification in perihilar cholangiocarcinoma.[4] Distal cholangiocarcinoma was classified as Bismuth–Corlette type 0. Moreover, the Bismuth–Corlette classification on the CT scan was compared with the reference standard and categorized as overestimation, correct assessment, or underestimation of the longitudinal extent of the tumor. When a Bismuth type IIIA tumor was determined to be IIIB on the CT scan or vice versa, it was categorized as incorrect. Furthermore, tumor involvement was evaluated based on the morphology (e.g., wall thickening of the bile duct, intraductal soft tissue mass, and asymmetric stricture) and degree of contrast enhancement (e.g., hyperenhancement to the liver parenchyma).[4,16] Hyperenhancement of the bile duct wall was assessed mainly in the arterial phase, if available, for intrapancreatic extent, and it was assessed in the portal venous phase for extrapancreatic extent.[4,16] Reader confidence on the longitudinal tumor extent was recorded on the 3-point Likert scale as follows: 0, 50%–75% confidence; 1, 76%–90% confidence; and 2, >90% confidence.[17] The location of the biliary stent was also recorded in each segment of the biliary tree to evaluate the effect of the biliary stent on the CT assessment.

Magnetic resonance imaging analysis

When preoperative MRI was performed, the patients were divided into two subgroups according to the timing of MRI with respect to BD to compare the performance of MRI obtained pre- and post-BD. MRI scans were retrospectively and independently reviewed by two readers (S.B.C., a radiology resident, and J.P., a board-certified abdominal radiologist with one year of practice experience) who were blinded to the surgical and pathological findings. Moreover, tumor involvement was evaluated in the same manner as in the CT scan evaluation. Additional consideration was included for mild-to-moderate hyperintensity on T2-weighted images, intraductal filling defects due to a soft tissue mass using MR cholangiopancreatography, and high signal intensity of wall thickening or intraductal mass on high *b* value diffusion-weighted imaging.[4,18] The presence or absence of a tumor in each segment, location of the biliary stent (if obtained after BD), and reader confidence were recorded.

Reference standard

The reference standard was determined by the surgical and pathological reports. Additionally, pathological reports were correlated with preoperative CT or MRI scans when necessary to verify the length of biliary tree segments. The presence or absence of tumors in each segment of the biliary tree and the longitudinal tumor extent were summarized using the Bismuth–Corlette classification.

Statistical analysis

Continuous variables were summarized using either mean with standard deviation or median with interquartile range (IQR), whereas categorical variables were summarized as counts and percentages. Continuous variables were compared using the Wilcoxon signed-rank test and categorical variables by chi-squared test, according to data normality. After pooling all segments of the biliary tree in all patients, the biliary segment-wise sensitivity, specificity, and accuracy of CT scans for detecting tumors were estimated using the generalized estimating equation (GEE) method. The GEE method was used to compare the estimates, considering that the segments were nested within each patient. The comparison was also performed for segments with and without a catheter. The Bismuth–Corlette classification and reader confidence were compared using the McNemar test. Inter-reader agreement for the Bismuth–Corlette classification in each reading session was assessed using Cohen κ statistics. In the subgroup of patients who underwent MRI, the biliary segment-wise performance of MRI for detecting tumors was compared between patients who underwent MRI pre-BD and those who underwent MRI post-BD using the GEE. Inter-reader agreement for the Bismuth–Corlette classification on MRI was assessed using Cohen κ statistics. Moreover, *P* values were adjusted for multiple comparisons using the Bonferroni correction, and a two-sided *P* value of less than 0.05 indicated statistical significance. The R package (version 4.2.2; The R Foundation for Statistical Computing, Vienna, Austria) was used for analyses.

Results

Patient characteristics

A total of 105 patients [mean age: 67 ± 8 years; 74 men (70.5%) and 31 women (29.5%)] were included in this study (Table 1). Among the patients, 64 (61.0%) had distal cholangiocarcinoma, and 41 (39.0%) had perihilar cholangiocarcinoma. Furthermore, BD was performed using endoscopic retrograde cholangiopancreatography in most patients (83.8%). Among the perihilar cholangiocarcinoma cases, Bismuth types I (26.8%) and IV (24.4%) were the most common. Pylorus-preserving pancreaticoduodenectomy was most commonly performed in patients with distal cholangiocarcinoma (98.4%), and hepatectomy was most commonly performed in those with perihilar cholangiocarcinoma (51.2%). Moreover, the median time interval between pre-BD CT and BD was 3 days (IQR, 1–9 days), and that between CT pre- and post-BD was 30 days (IQR, 18–40 days). The median time interval between pre-BD CT and surgery was 38 days (IQR, 28–57 days), and that between post-BD CT and surgery was 7 days (IQR, 2–23 days). Arterial phase CT images were available in 81 (77.1%) pre-BD and in 69 (65.7%) post-BD, with no statistical differences in the frequency ($P = 0.182$).

The MRI subgroup included 85 patients: 55 underwent MRI pre-BD [median time interval between MRI and BD, 1 day (IQR, 0–4 days)], and the remaining underwent MRI post-BD [median time interval between MRI and BD, 1 day (IQR, 1–3 days)]. Most MRIs (90.6%) were performed using contrast media, either a hepatobiliary (63.5%) or extracellular (27.1%) contrast agent. The median time interval between MRI and surgery was 33 days (IQR, 26–52 days), which was shorter than that between pre-BD CT and surgery ($P < 0.001$).

Diagnostic performance of computed tomography

In all biliary segments, the performance for tumor detectability using both pre- and post-BD CTs was superior to post-BD CT and comparable with pre-BD CT for both readers (Table 2). The sensitivity and specificity for detecting the tumor segment were higher using both pre- and post-BD CTs than using post-BD CT alone (reader 1: sensitivity, 72.6% vs. 64.6%, $P < 0.001$; specificity, 96.9% vs. 94.8%, $P = 0.063$; reader 2: sensitivity, 77.2% vs. 72.9%, $P = 0.126$; specificity, 97.5% vs. 94.2%, $P = 0.003$). In biliary segments with a catheter in the lumen, the sensitivity was higher using both pre- and post-BD CTs than using post-BD CT. Moreover, the specificity was higher in reader 2 using both CTs, reducing the overestimation of the longitudinal tumor extent (reader 1: sensitivity, 74.4% vs. 67.5%, $P = 0.006$; specificity, 92.4% vs. 88.0%, $P = 0.068$; reader 2: sensitivity, 80.5% vs. 74.4%, $P = 0.013$; specificity, 94.3% vs. 88.0%,

Table 1. Clinicopathological characteristics of patients with extrahepatic bile duct cancer

Variable	Total (n = 105)	Perihilar (n = 41)	Distal (n = 64)
Age at the time of surgery (y), mean ± SD	67 ± 8	67 ± 7	68 ± 8
Male/female	74 (70.5)/31 (29.5)	33 (80.5)/8 (19.5)	41 (64.1)/23 (35.9)
Biliary decompression method			
PTBD	17 (16.2)	7 (17.1)	10 (15.6)
ERCP	88 (83.8)	34 (82.9)	54 (84.4)
Bismuth–Corlette classification			
I	NA	11 (26.8)	NA
II	NA	9 (22.0)	NA
IIIA	NA	7 (17.1)	NA
IIIB	NA	4 (9.8)	NA
IV	NA	10 (24.4)	NA
Type of surgery			
PPPD	75 (72.4)	12 (29.3)	63 (98.4)
Hepatectomy	21 (20.0)	21 (51.2)	0 (0.0)
Segmental resection of bile duct	5 (4.8)	4 (9.8)	1 (1.6)
Segmental resection of bile duct with hepatectomy	3 (2.9)	3 (7.3)	0 (0.0)
Hepatopancreatoduodenectomy	1 (1.0)	1 (2.4)	0 (0.0)
Resection margin status			
R0	64 (61.0)	20 (48.8)	44 (68.8)
R1	41 (39.0)	21 (51.2)	20 (31.3)
Pathologic grade			
Well differentiated	12 (11.4)	4 (9.8)	8 (12.5)
Moderately differentiated	71 (67.6)	27 (65.9)	44 (68.8)
Poorly or undifferentiated	22 (21.0)	10 (24.4)	12 (18.8)

Data are presented as numbers of patients with percentages in parentheses unless otherwise specified.
SD, standard deviation; PTBD, percutaneous transhepatic biliary drainage; ERCP, endoscopic retrograde cholangiopancreatography; PPPD, pylorus-preserving pancreaticoduodenectomy; NA, not applicable.

P = 0.016) (Figures 2, 3). In biliary tree segments without a catheter, the sensitivity was higher using both pre- and post-BD CTs than using post-BD CT for reader 1 (sensitivity, 67.1% vs. 55.7%, P = 0.001; specificity, 98.9% vs. 97.8%, P = 0.156), yet remained comparable for reader 2 (sensitivity, 67.1% vs. 68.4%, P = 0.763; specificity, 98.9% vs. 96.9%, P = 0.019).

The Bismuth–Corlette classification was comparable among the three reading sessions (all Ps > 0.05 for both readers), but there were fewer cases of overestimating the Bismuth–Corlette classification using pre-BD CT or both pre- and post-BD CTs than using post-BD CT alone (reader 1: 4.8% vs. 11.4% vs. 13.3%; reader 2: 6.7% vs. 12.4% vs. 16.2%) (Table 3). Both readers performed one incorrect classification using post-BD CT but none using pre-BD CT or both CTs (Figure 3). Inter-reader agreement for the Bismuth–Corlette classification in each reading session was substantial (κ = 0.67 using pre-BD CT, κ = 0.71 using both pre- and post-BD CTs, κ = 0.79 using post-BD CT).

Reader confidence was significantly higher using pre-BD CT or both CTs than using post-BD CT, with a higher proportion of >90% reader confidence (reader 1: 64.8–65.7% vs. 13.3%; reader 2: 81.0–87.6% vs. 6.7%, P < 0.001 for both readers) (Table 4).

Diagnostic performance of magnetic resonance imaging

The performance for tumor detectability was not significantly different between pre- and post-BD MRIs (Supplementary Table S1, Supplementary Figures S1, S2). Post-BD MRI scans exhibited a comparable performance with pre-BD MRI scans, but the pre-BD MRI was minimally superior (reader 1: sensitivity, 77.9% vs. 75.0%, P = 0.605; specificity, 97.2% vs. 94.9%, P = 0.256; reader 2: sensitivity, 73.2% vs. 72.6%, P = 0.926; specificity, 98.4% vs. 94.9%, P = 0.068). Inter-reader agreement for the Bismuth–Corlette classification was almost perfect using MRI (κ = 0.85).

Discussion

The accurate evaluation of surgical resectability is crucial in EHD cancer; however, BD poses limitations in the assessment of the longitudinal tumor extent. This study investigated the diagnostic performance of evaluating the longitudinal extent of EHD cancer after BD based on the recent KSAR consensus recommendations. It also explored the best strategy to improve CT performance in the preoperative setting. In 105 patients, reading pre-BD CT scans or the combined reading of pre- and post-BD CT scans showed better diagnostic performance than reading post-BD CT scans alone, which was supported by significantly higher reader confidence. In a subgroup of 85 patients, MRI performance was compared between those who underwent MRI pre- and post-BD, and no significant difference was found in diagnostic performance, which was similar to that of the combined CT reading.

The biliary segment-wise performance for tumor detectability was higher when both pre- and post-BD CT scans were considered compared with post-BD CT alone. These results may be attributed to the overestimation of the longitudinal tumor extent owing to post-BD cholangitis, which contributes to wall thickening and enhancement on CT scans.[19-21] Catheter-related beam-hardening artifacts from BD may also obscure the tumor, causing challenges in the evaluation.[14,22] In a previous study, the acquisition of CT post-BD was associated with decreased R0 resection rates in EHD cancer in comparison with the acquisition of CT pre-BD, which led to poorer survival rates in patients who underwent CT post-BD.[12] The results of that study are reinforced by the limited performance of post-BD CT scans in this study. The head-to-head comparison of the performance between pre-BD CT and post-BD CT corroborates the previous observation that the evaluation of secondary biliary confluence was less accurate in patients who underwent CT after BD than those who underwent CT before BD.[13] Moreover, in the biliary segment where the drainage catheter was located, post-BD CT alone showed a particularly decreased specificity. Therefore, considering pre-BD CT scans is useful for the accurate evaluation of surgical resectability even after BD.

For the Bismuth–Corlette classification evaluated on CT scans, all reading sessions showed accuracies of 66.7%–74.3%. This

Table 2. Diagnostic performance of CT scans for determining the longitudinal extent of the extrahepatic bile duct cancer

Reader	Examined biliary segment	Pre-BD (A)	Pre- and post-BD (B)	Post-BD (C)	*P* (A vs. B)	*P* (B vs. C)	*P* (C vs. A)
Sensitivity (%)							
	All segments	72.6 (67.8–77.5)	72.6 (67.8–77.5)	64.6 (59.4–69.8)	>0.999	<0.001	<0.001
1	Segments with a catheter	NA	74.4 (68.9–79.8)	67.5 (61.6–73.3)	NA	0.006	NA
	Segments without a catheter	NA	67.1 (56.7–77.5)	55.7 (44.7–66.7)	NA	0.001	NA
	All segments	72.9 (68.1–77.8)	77.2 (72.7–81.8)	72.9 (68.1–77.8)	0.114	0.126	>0.999
2	Segments with a catheter	NA	80.5 (75.5–85.4)	74.4 (68.9–79.8)	NA	0.013	NA
	Segments without a catheter	NA	67.1 (56.7–77.5)	68.4 (58.1–78.6)	NA	0.763	NA
Specificity (%)							
	All segments	97.7 (96.4–99.0)	96.9 (95.4–98.4)	94.8 (92.8–96.7)	0.471	0.063	0.015
1	Segments with a catheter	NA	92.4 (88.3–96.5)	88.0 (82.9–93.1)	NA	0.068	NA
	Segments without a catheter	NA	98.9 (97.8–100.0)	97.8 (96.2–99.3)	NA	0.156	NA
	All segments	96.7 (95.2–98.2)	97.5 (96.1–98.8)	94.2 (92.2–96.2)	0.855	0.003	0.069
2	Segments with a catheter	NA	94.3 (90.7–97.9)	88.0 (82.9–93.1)	NA	0.016	NA
	Segments without a catheter	NA	98.9 (97.8–100.0)	96.9 (95.1–98.7)	NA	0.019	NA
Accuracy (%)							
	All segments	88.0 (85.8–90.2)	87.5 (85.3–89.7)	83.1 (80.6–85.6)	1.515	<0.001	<0.001
1	Segments with a catheter	NA	81.4 (77.6–85.2)	75.5 (71.3–79.7)	NA	0.001	NA
	Segments without a catheter	NA	93.1 (90.7–95.5)	90.1 (87.3–92.9)	NA	0.001	NA
	All segments	87.5 (85.3–89.7)	89.6 (87.6–91.7)	86.0 (83.6–88.3)	0.060	<0.001	0.600
2	Segments with a catheter	NA	85.9 (82.5–89.3)	79.7 (75.8–83.6)	NA	0.001	NA
	Segments without a catheter	NA	93.1 (90.7–95.5)	91.7 (89.2–94.3)	NA	0.179	NA

Data are performance measures with 95% confidence intervals in parentheses. CT, computed tomography; BD, biliary drainage; NA, not applicable.

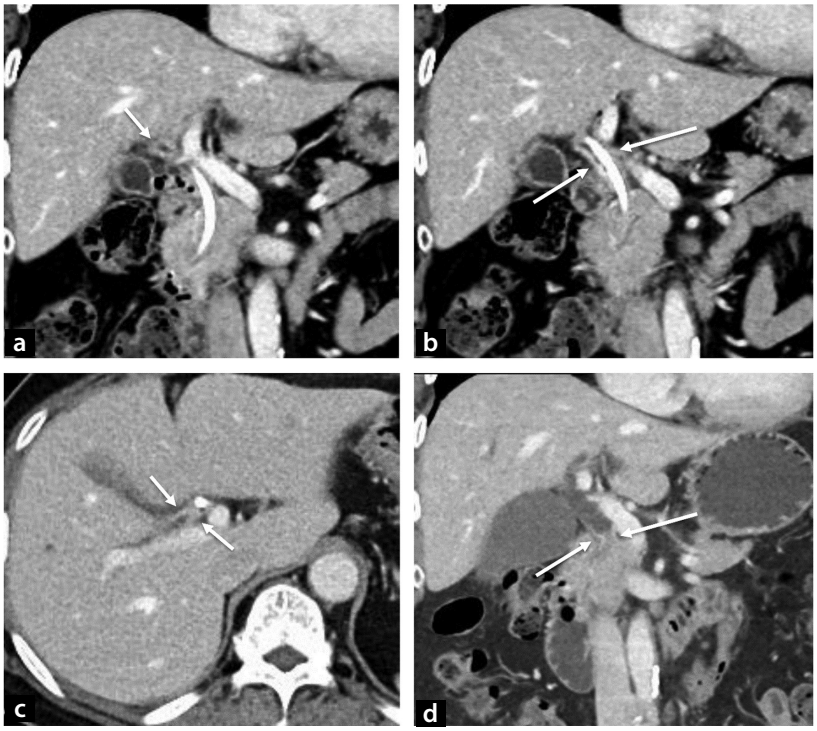


Figure 2. Overestimation of the longitudinal extent of distal cholangiocarcinoma after biliary drainage (BD) in a 68-year-old woman. **(a-c)** Coronal **(a, b)** and axial **(c)** portal venous phase computed tomography (CT) images obtained one month after BD showing diffuse enhancing wall thickening (arrows) of common bile duct (CBD), common hepatic duct, primary biliary confluence, and right hepatic duct. Both readers assessed the tumor as a Bismuth–Corlette type II hilar cholangiocarcinoma using post-BD CT scan. **(d)** Coronal portal venous phase CT image obtained before BD depicting segmental wall thickening and luminal narrowing of the intra- and supra-pancreatic CBD (arrow). Therefore, both readers correctly determined that the tumor was a distal cholangiocarcinoma using both pre- and post-BD CT scans.

result is similar to the previously reported CT accuracy for the longitudinal tumor extent, which ranged between 56.3%–74.1%.[23] When both CT scans were considered, there were fewer cases of overestimation compared with using post-BD CT scans alone. This can be attributed to the limited performance of post-CT scans in differentiating between tumors and inflammation caused by BD.[11] Moreover, reader confidence was higher when both CT scans were considered. Therefore, when evaluating the longitudinal tumor extent using post-BD CT scans, considering pre-BD CT scans, whenever available, may improve the diagnostic performance of CT scans. Furthermore, inter-reader agreement for the Bismuth–Corlette classification was substantial for CT readings, either pre-BD or post-BD, which was comparable with the results of a prior study.[23] These results may be explained by the standardized evaluation of biliary tree segments based on the radiologic consensus guidelines, which supports the structured reporting approach.

A recent study showed that the performance of CT and MRI was comparable before BD for the resectability evaluation of EHD cancer.[6] Of note, MRI scan performance did not significantly decrease after BD in this study in contrast with the CT scan perfor-

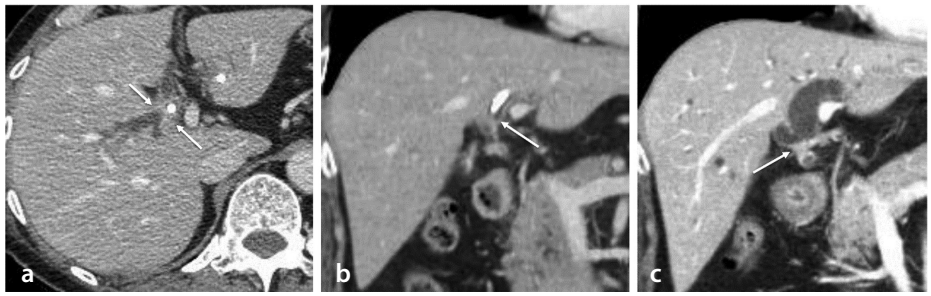


Figure 3. Overestimation of the longitudinal extent of Bismuth–Corlette type IIIA perihilar cholangiocarcinoma after biliary drainage (BD) in a 64-year-old man. **(a, b)** Axial **(a)** and coronal **(b)** portal venous phase computed tomography (CT) images obtained one month after BD showing diffuse enhancing wall thickening (arrows) of the hilar bile duct extending to both hepatic ducts. Both readers assessed the tumor as a Bismuth–Corlette type IIIB hilar cholangiocarcinoma using post-BD CT scan (incorrect assessment), which was probably attributable to cholangitis extending to the left hilar bile duct. **(c)** Coronal portal venous phase CT image obtained before BD depicts segmental wall thickening and luminal narrowing of the hilar bile duct with the proximal end at the primary biliary confluence (arrow). Therefore, both readers decided that the tumor was a Bismuth–Corlette type II using both pre- and post-BD CT scans (underestimation). The involvement of the right secondary biliary confluence was not detected on the pre-BD CT scan, which was probably because of microscopic tumor extension.

Table 3. Results of the Bismuth–Corlette classification using CT scans

	Pre-BD (A)	Pre- and post-BD (B)	Post-BD (C)	*P*		
				A vs. B	B vs. C	C vs. A
Reader 1						
Overestimation	5 (4.8)	12 (11.4)	14 (13.3)	0.390	>0.999	0.864
Correct	78 (74.3)	70 (66.7)	70 (66.7)			
Underestimation	22 (21.0)	23 (21.9)	20 (19.1)			
Incorrect[a]	0 (0.0)	0 (0.0)	1 (1.0)			
Reader 2						
Overestimation	7 (6.7)	13 (12.4)	17 (16.2)	>0.999	>0.999	0.405
Correct	76 (72.4)	72 (68.6)	73 (69.5)			
Underestimation	22 (21.0)	20 (19.1)	14 (13.3)			
Incorrect[a]	0 (0.0)	0 (0.0)	1 (1.0)			

Data are numbers of patients with percentages in parentheses. [a]Bismuth type IIIA determined as IIIB or vice versa. CT, computed tomography; BD, biliary drainage.

Table 4. Comparison of reader confidence

	Pre-BD (A)	Pre- and post-BD (B)	Post-BD (C)	*P*		
				A vs. B	B vs. C	C vs. A
Reader 1						
50%–75%	4 (3.8)	6 (5.7)	57 (54.3)	>0.999	<0.001	<0.001
76%–90%	33 (31.4)	30 (28.6)	34 (32.4)			
>90%	68 (64.8)	69 (65.7)	14 (13.3)			
Reader 2						
50%–75%	1 (1.0)	0 (0.0)	42 (40.0)	0.723	<0.001	<0.001
76%–90%	12 (11.4)	20 (19.1)	56 (53.3)			
>90%	92 (87.6)	85 (81.0)	7 (6.7)			

Data are numbers of patients with percentages in parentheses. BD, biliary drainage.

mance. This is probably because of the higher contrast resolution of the bile duct lumen and wall on MRI scans than on CT scans.[24] Moreover, there was no significant difference in the diagnostic performance between MRI scans obtained pre- and post-BD, and the sensitivity, specificity, and accuracy were similar to those of the combined reading of pre- and post-BD CT scans. The previous study by Chryssou et al.[15] observed a tendency toward overestimating the tumor extent after BD on MRI. However, this study, which recruited a larger number of patients, observed a comparably accurate performance of MRIs pre- and post-BD, approaching the reported near-perfect performance of pre-BD MRI.[25] Hence, it was assumed that even when MRI is performed after an urgent BD, there would be no significant difference in the diagnostic performance. In addition, the almost perfect inter-reader agreement for the Bismuth–Corlette classification on MRI indicates the benefits of radiologic consensus guidelines. However, another study using pre-BD MRI observed only moderate agreement for four readers, which may better reflect the real world.[6]

This study had some limitations. First, a selection bias might have been introduced in the surgical cohort. However, evaluating the longitudinal tumor extent would be more challenging in resectable cancers than in advanced cancers with higher T stages. Second, the validity of surgical and pathological reference standards can be suboptimal in patients with R1 resection margin status. However, the effect would be small in patients where R1 resection margins are attributable to the circumferential margin instead of the ductal margin. Third, multiphasic CT was less frequently performed after BD, albeit statistically insignificant, because single-phase CT was a preferred modality for follow-up imaging due to a lower radiation dose. This might have affected post-BD CT performance for tumors involving intrapancreatic CBD. Fourth, the diagnostic performance of pre- and post-BD MRI could not be compared in the same patient. This was because CT was the preferred imaging modality for EHD cancer, and no patient underwent MRI both pre- and post-BD. Fifth, the interval between MRI and surgery was shorter than that between pre-BD CT and surgery because MRI was the secondary imaging modality performed after CT. Nonetheless, the median interval showed approximately a one-week difference, and the longitudinal tumor extent could not have changed significantly during this interim period.[26] Lastly, the study results might not be generalizable to patients who undergo preoperative chemotherapy or radiotherapy, as preoperative treatment would affect the longitudinal tumor extent.

In conclusion, the consideration of pre-BD CT scans provided better diagnostic performance than reading post-BD CT scans alone. Therefore, the acquisition of pre-BD CT would be beneficial for the preoperative evaluation of EHD cancer when BD is

planned. Moreover, MRI evaluation would not be significantly affected by BD in terms of the diagnostic performance of the longitudinal tumor extent.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

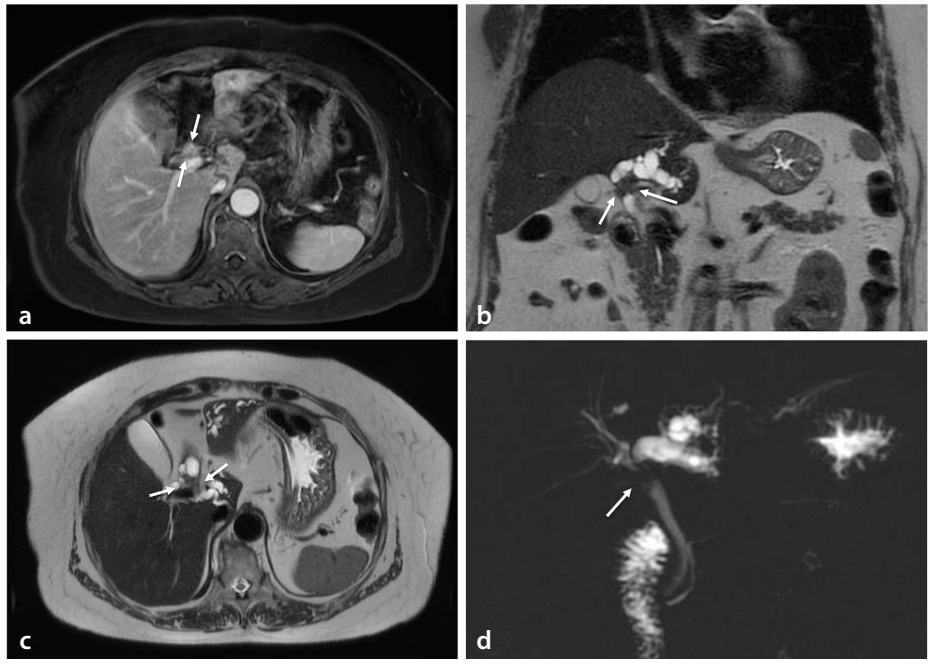
1. Esnaola NF, Meyer JE, Karachristos A, Maranki JL, Camp ER, Denlinger CS. Evaluation and management of intrahepatic and extrahepatic cholangiocarcinoma. *Cancer*. 2016;122(9):1349-1369. [CrossRef]
2. Vithayathil M, Khan SA. Current epidemiology of cholangiocarcinoma in Western countries. *J Hepatol*. 2022;77(6):1690-1698. [CrossRef]
3. Kim BW, Oh CM, Choi HY, Park JW, Cho H, Ki M. Incidence and overall survival of biliary tract cancers in South Korea from 2006 to 2015: using the National Health Information Database. *Gut Liver*. 2019;13(1):104-113. [CrossRef]
4. Lee DH, Kim B, Lee ES, et al. Radiologic evaluation and structured reporting form for extrahepatic bile duct cancer: 2019 consensus recommendations from the Korean Society of Abdominal Radiology. *Korean J Radiol*. 2021;22(1):41-62. [CrossRef]
5. Ke Q, Wang B, Lin N, Wang L, Liu J. Does high-grade dysplasia/carcinoma in situ of the biliary duct margin affect the prognosis of extrahepatic cholangiocarcinoma? A meta-analysis. *World J Surg Oncol*. 2019;17(1):211. [CrossRef]
6. Yoo J, Lee JM, Kang HJ, Bae JS, Jeon SK, Yoon JH. Comparison between contrast-enhanced computed tomography and contrast-enhanced magnetic resonance imaging with magnetic resonance cholangiopancreatography for resectability assessment in extrahepatic cholangiocarcinoma. *Korean J Radiol*. 2023;24(10):983-995. [CrossRef]
7. Akita M, Ajiki T, Matsumoto T, et al. Preoperative cholangitis affects survival outcome in patients with extrahepatic bile duct cancer. *J Gastrointest Surg*. 2017;21(6):983-989. [CrossRef]
8. Celotti A, Solaini L, Montori G, Coccolini F, Tognali D, Baiocchi G. Preoperative biliary drainage in hilar cholangiocarcinoma: systematic review and meta-analysis. *Eur J Surg Oncol*. 2017;43(9):1628-1635. [CrossRef]
9. Rassam F, Roos E, van Lienden KP, et al. Modern work-up and extended resection in perihilar cholangiocarcinoma: the AMC experience. *Langenbecks Arch Surg*. 2018;403(3):289-307. [CrossRef]
10. Cho E, Kim SH, Choi SJ, et al. Diagnosis and treatment of perihilar cholangiocarcinoma: a national survey from the Korean Pancreatobiliary Association. *Gut Liver*. 2023. [CrossRef]
11. Nguyen NT, Khan HA, Abdul-Baki K, et al. CT imaging features of bile duct stent complications. *Clin Imaging*. 2023;103:109986. [CrossRef]
12. Hosokawa I, Shimizu H, Yoshitomi H, et al. Impact of biliary drainage on multidetector-row computed tomography on R0 resection of perihilar cholangiocarcinoma. *World J Surg*. 2018;42(11):3676-3684. [CrossRef]
13. Ni Q, Wang H, Zhang Y, et al. MDCT assessment of resectability in hilar cholangiocarcinoma. *Abdom Radiol (NY)*. 2017;42(3):851-860. [CrossRef]
14. Franken LC, Coelen RJS, Erdmann JI, et al. Multidetector computed tomography assessment of vascular involvement in perihilar cholangiocarcinoma. *Quant Imaging Med Surg*. 2021;11(11):4514-4521. [CrossRef]
15. Chryssou E, Guthrie JA, Ward J, Robinson PJ. Hilar cholangiocarcinoma: MR correlation with surgical and histological findings. *Clin Radiol*. 2010;65(10):781-788. [CrossRef]
16. Choi KS, Lee JM, Joo I, Han JK, Choi BI. Evaluation of perihilar biliary strictures: does DWI provide additional value to conventional MRI? *AJR Am J Roentgenol*. 2015;205(4):789-796. [CrossRef]
17. Obmann MM, An C, Schaefer A, et al. Improved sensitivity and reader confidence in CT colonography using dual-layer spectral CT: a phantom study. *Radiology*. 2020;297(1):99-107. [CrossRef]
18. Ryoo I, Lee JM, Chung YE, et al. Gadobutrol-enhanced, three-dimensional, dynamic MR imaging with MR cholangiography for the preoperative evaluation of bile duct cancer. *Invest Radiol*. 2010;45(4):217-224. [CrossRef]
19. Paik WH, Loganathan N, Hwang JH. Preoperative biliary drainage in hilar cholangiocarcinoma: when and how? *World J Gastrointest Endosc*. 2014;6(3):68-73. [CrossRef]
20. Kawashima H, Itoh A, Ohno E, et al. Preoperative endoscopic nasobiliary drainage in 164 consecutive patients with suspected perihilar cholangiocarcinoma: a retrospective study of efficacy and risk factors related to complications. *Ann Surg*. 2013;257(1):121-127. [CrossRef]
21. Kurahara H, Maemura K, Mataki Y, et al. Preoperative biliary drainage-related inflammation is associated with shorter survival in biliary tract cancer patients. *Int J Clin Oncol*. 2016;21(5):934-939. [CrossRef]
22. Unno M, Okumoto T, Katayose Y, et al. Preoperative assessment of hilar cholangiocarcinoma by multidetector row computed tomography. *J Hepatobiliary Pancreat Surg*. 2007;14(5):434-440. [CrossRef]
23. Lee DH, Kim B, Lee JM, Lee ES, Choi MH, Kim H. Multidetector CT of extrahepatic bile duct cancer: diagnostic performance of tumor resectability and interreader agreement. *Radiology*. 2022;304(1):96-105. [CrossRef]
24. Choi JY, Kim MJ, Lee JM, et al. Hilar cholangiocarcinoma: role of preoperative imaging with sonography, MDCT, MRI, and direct cholangiography. *AJR Am J Roentgenol*. 2008;191(5):1448-1457. [CrossRef]
25. Xin Y, Liu Q, Zhang J, et al. Hilar cholangiocarcinoma: value of high-resolution enhanced magnetic resonance imaging for preoperative evaluation. *J Cancer Res Ther*. 2020;16(7):1634-1640. [CrossRef]
26. Nooijen LE, de Boer MT, Braat AE, et al. National consensus on a new resectability classification for perihilar cholangiocarcinoma - A modified Delphi method. *Eur J Surg Oncol*. 2023:107117. [CrossRef]

Supplementary Table S1. Diagnostic performance of MRI scans in determining longitudinal extent of extrahepatic bile duct cancer

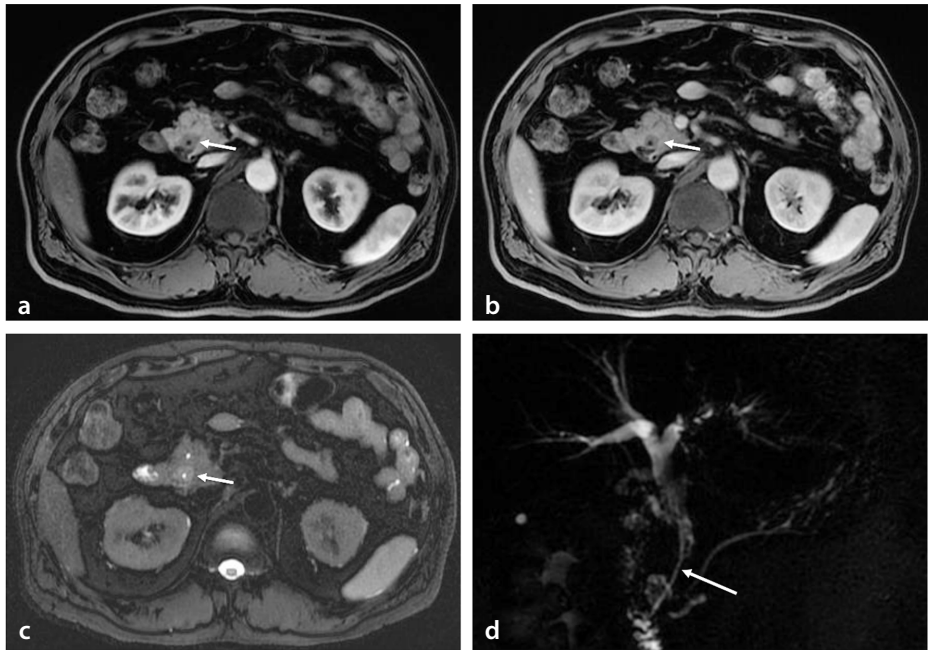
Reader	Sensitivity (%)		*P*	Specificity (%)		*P*	Accuracy (%)		*P*
	Pre-BD MRI	Post-BD MRI		Pre-BD MRI	Post-BD MRI		Pre-BD MRI	Post-BD MRI	
1	77.9 (72.0–83.8)	75.0 (65.7–84.3)	0.605	97.2 (95.2–99.3)	94.9 (91.4–98.3)	0.256	88.9 (85.9–91.8)	87.9 (83.8–92.0)	0.256
2	73.2 (66.9–79.5)	72.6 (63.1–82.2)	0.926	98.4 (96.8–100.0)	94.9 (91.4–98.3)	0.068	87.5 (84.4–90.6)	87.1 (82.8–91.3)	0.876

Data are performance measures in percentages, with 95% confidence intervals in parentheses. MRI, magnetic resonance imaging; BD, biliary drainage.

Online Resource 1. Magnetic resonance imaging (MRI) acquisition. For MRI acquisition, either a 1.5-T (Intera Achieva, Philips Medical Systems, Best, Netherlands) or 3.0-T scanner (Magnetom Trio Tim, Siemens Medical Solutions, Erlangen, Germany; Intera Achieva or Ingenia, Philips Medical Systems, Best, Netherlands; Discovery MR750w MRI unit, GE Medical Systems, Waukesha, Wisconsin, U.S.A.) with a 4-,16-, or 32-channel torso-array coil was used. A breath-hold axial T1-weighted dual-echo gradient-recalled echo sequence was used for pre-contrast T1 images. T2-weighted single- or multi-shot turbo spin-echo with or without spectral fat suppression was performed either before or after contrast use. Magnetic resonance cholangiopancreatography was performed with a two-dimensional thick-slab single-shot turbo spin-echo or three-dimensional T2-weighted respiratory-triggered fast spin-echo sequence using the navigator technique. Dynamic contrast-enhanced T1-weighted imaging was performed after administration of either a hepatobiliary or extracellular contrast agent (Primovist, gadoxetic acid, Bayer Schering Pharma, Berlin, Germany; Dotarem, gadoterate meglumine, Guerbet, France). The contrast was injected as 0.1 mL/kg of Primovist bolus injection at a rate of 1 mL/sec or 0.2 mL/kg of Dotarem at a rate of 1 or 2 mL/sec, followed by 20 mL of saline flush. Arterial phase timing was determined using the test-bolus or bolus tracking method, 2-5 sec after peak aorta enhancement. Portal venous phase (50–60 sec) and delayed or transitional (2–3 min) phase images were obtained. Hepatobiliary phase (15–20 min) images were obtained using gadoxetic acid. Diffusion-weighted imaging was performed at b values of 0 or 50, 400, and 800 s/mm^2.



Supplementary Figure S1. MRI evaluation of perihilar cholangiocarcinoma before biliary drainage in a 80-year-old woman. **(a-c)** Axial (a) T1-weighted fat-suppressed image in portal venous phase, and coronal **(b)** and axial **(c)** T2-weighted images show segmental enhancing wall thickening (arrows) of common hepatic duct, extending to primary biliary confluence, and left secondary biliary confluence. **(d)** Two-dimensional magnetic resonance cholangiopancreatography shows a filling defect (arrow) in biliary tree by the tumor, with dilatation of left intrahepatic bile ducts. Both readers correctly assessed the tumor as Bismuth-Corlette type IIIB perihilar cholangiocarcinoma. MRI, magnetic resonance imaging.



Supplementary Figure S2. MRI evaluation of distal cholangiocarcinoma after biliary drainage in a 61-year-old man. **(a-c)** Axial T1-weighted fat-suppressed image in arterial phase **(a)** and portal venous phase **(b)**, and axial T2-weighted fat-suppressed image **(c)** show segmental enhancing wall thickening (arrow) of intrapancreatic CBD, and biliary stent in the center. **(d)** Two-dimensional magnetic resonance cholangiopancreatography shows biliary stent (arrow) located at common hepatic duct to intrapancreatic CBD, with upstream biliary tree dilatation. Both readers correctly determined the longitudinal tumor extent as distal cholangiocarcinoma. CBD, common bile duct; MRI, magnetic resonance imaging.

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.232586

ABDOMINAL IMAGING

ORIGINAL ARTICLE

Volumetric segmentation analysis of the levator ani muscle using magnetic resonance imaging in pelvic floor function assessment

Ayşenur Buz Yaşar
Rüveyde Begüm Yüzok
Emine Dağıstan

Bolu Abant İzzet Baysal University, Training and Research Hospital, Department of Radiology, Bolu, Turkey

PURPOSE

In this case-control study, we aimed to evaluate how muscle volume affects pelvic floor function by analyzing the levator ani muscle (LAM) using volumetric segmentation in addition to standard magnetic resonance (MR) defecography assessments.

METHODS

We enrolled 85 patients with varying degrees of pelvic floor dysfunction (PFD) and 85 age- and gender-matched controls in this retrospective study. All patients had MR defecography images, while all controls had pelvic MR images obtained for other reasons. Group comparisons were performed using independent samples t-tests and Mann–Whitney U tests. The receiver operating curve (ROC) was constructed to establish a cut-off value for a normal LAM volume. Interrater reliability was assessed by calculating the intraclass correlation coefficient. A *P* value of less than 0.05 was considered statistically significant.

RESULTS

Volumetric measurements revealed that the control group had higher LAM volumes, and the ROC curve analysis indicated a cut-off value of 38934.3 mm^3 with a sensitivity of 0.812 and specificity of 0.8 for PFD assessment using LAM volumetric measurement. Gender did not significantly affect LAM volume in the control group.

CONCLUSION

Alongside the useful structural and functional information acquired from MR defecography images, volumetric analysis, and three-dimensional reconstructions of LAM may help to improve the accuracy of the diagnosis.

KEYWORDS

Segmentation, levator ani muscle, pelvic floor dysfunction, magnetic resonance defecography, volumetric measurement

Pelvic floor dysfunction (PFD) is a comprehensive term that refers to a broad group of medical conditions that can affect the suspensory ligaments, fascial coverings, and muscles supporting the pelvic organs.[1,2] It is a common disorder with an estimated prevalence of 25 percent among women in the United States.[3] The etiological factors of PFD include female gender, a history of vaginal childbirth, chronic constipation, pelvic surgery, obesity, genetic predisposition, menopause, and aging.[1,3,4]

The anatomical structures in the pelvic region include the bladder, prostate, uterus, vagina, and rectum, which are evaluated in three compartments: anterior (bladder and urethra), middle (uterus and vagina), and posterior (rectum, anal canal).[2,4] These structures are attached by the endopelvic fascia, the pelvic diaphragm, and the urogenital diaphragm and function as a single unit.[1] The levator ani muscle (LAM) is a complex funnel-shaped structure consisting of three main components: the pubococcygeus (pubovaginalis, puboprostaticus, puboperineal, puboanal), puborectalis, and iliococcygeus.[4-6] Damage or weakening of the LAM is the most common cause of pelvic organ prolapse (POP), resulting in a distorted shape of muscle that

Corresponding author: Ayşenur Buz Yaşar

E-mail: aysenurbuz@gmail.com

Received 06 November 2023; revision requested 29 November 2023; last revision received 02 January 2024; accepted 18 January 2024.

Epub: 20.02.2024

Publication date: 08.07.2024

DOI: 10.4274/dir.2024.232586

You may cite this article as: Buz Yaşar A, Yüzok RB, Dağıstan E. Volumetric segmentation analysis of the levator ani muscle using magnetic resonance imaging in pelvic floor function assessment. *Diagn Interv Radiol.* 2024;30(4):220-227.

tends to tilt more vertically and widen the levator hiatus. An insult to the pelvic floor muscles or ligaments can lead to urinary or fecal incontinence, dyspareunia, constipation, and pelvic pain.[2,6]

Radiologically, for the assessment of these structures, magnetic resonance (MR) defecography uses sagittal balanced steady-state gradient echo (GRE) (different vendors have similar sequences with different trade names, such as True-FISP, FIESTA, and balanced-FFE) images, and a reference line called the pubococcygeal line (PCL) is drawn from the lower border of the pubic symphysis to the last coccygeal joint.[1,2,7] The distance perpendicular from the posterior wall of the anorectal component to the PCL is called the "M line", which corresponds to a measure of the location of the anorectal junction. The "H line" is the distance from the inferior border of the pubic symphysis to the posterior of the anorectal component and represents the anteroposterior width of the levator hiatus (Figure 1).[7]

MR defecography is a dynamic examination that allows evaluation at rest, during contraction, and defecation.[1,2,4] In the anterior compartment, the position of the urethra and bladder can be assessed for diagnoses such as urethral hypermobility and cystocele. In the middle compartment, uterine or cervical prolapse can be observed. In the posterior compartment, pathologies like rectocele or rectal intussusception can be detected.[4,8] The classic PCL, H, and M lines are utilized for the radiological grading of these pathologies. Additionally, measuring the anorectal angle (ARA) and its dynamic changes are important in diagnosing pelvic floor dyssynergia.[9]

Main points

- Pelvic floor dysfunction (PFD) encompasses various medical conditions affecting the supportive ligaments, fascial coverings, and the muscles in the pelvic region.
- The levator ani muscle (LAM) is a critical component, and damage or weakening of this muscle is a common cause of pelvic organ prolapse (POP) and related conditions like incontinence, dyspareunia, and pelvic pain.
- Our study indicates that individuals with PFD tend to have a lower LAM volume, with a specific cut-off value for muscle volume linked to a higher tendency for PFD.
- However, contrary to initial assumptions, no linear correlation between the severity of POP or pelvic floor relaxation and muscle volume was observed in this study.

The thickness and volume measurement of the LAM have been investigated through ultrasound (US), computed tomography (CT), and MR studies, and the factors influencing the muscle volume and architecture have been explored.[10-12] In our study, we hypothesize that the LAM volume is lower in patients with PFD than in healthy individuals. Alongside routine evaluations in MR defecography, we aimed to assess the impact of muscle volumes on the POP by conducting a volumetric segmentation analysis of the LAM.

Methods

This retrospective study was approved by the Bolu Abant İzzet Baysal University Clinical Research Ethics Committee and written informed consent was waived by it (date of project: 22.08.2023; project decision number: 262). This research study was conducted in accordance with the Declaration of Helsinki. The STARD guidelines were followed for reporting and joint recommendations of the ESUR and ESGAR Pelvic Floor Working Group were followed for patient preparation, image acquisition, and interpretation.[8,13-15]

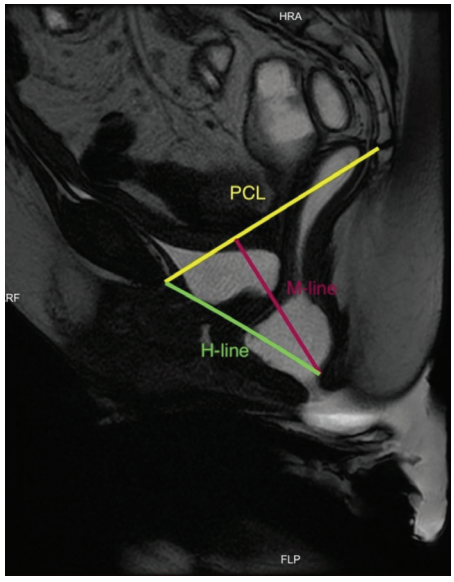


Figure 1. Reference lines of magnetic resonance defecography measurements at rest (pubococcygeal line: yellow line; H-line: green line; M-line: red line). PCL, pubococcygeal line; FLP, foot left posterior; HRA, head right anterior.

Patient selection

A total of 85 patients (M: 16, F: 69), aged between 19 and 92 years old, with varying levels of PFD, as well as 85 controls (M: 16, F: 69) with ages ranging from 20 to 85 years old, were included in this study. The control group consisted of individuals without pelvic floor pathology identified in MR imaging (MRI) assessments at rest but with pelvic MRI taken for other medical reasons. The participants' MR defecography and pelvic MRI were recruited from picture archiving communication systems. All MRI were acquired at the radiology department between September 2020 and November 2023. Patients with poor quality imaging due to artifacts on MR scan, inadequate or incomplete imaging, or in whom volumetric measurements could not be performed were excluded from participation, as were control group participants with findings of PFD on pelvic MRI. There were no patients with a history of pelvic region radiotherapy or oncological surgery in the patient population and control group (Figure 2 shows a flowchart of the study).

Patient preparation

For MR defecography images, patients were asked to empty their bladders and bowels 1–2 hours before the MRI examination. The patient should be trained about the imaging steps (rest, squeeze, strain, Valsalva maneuver, and defecation) prior to the MRI examination. Immediately before imaging, 120–180 mL of rectal gel was injected gently through the anal canal in the decubitus position to fill the rectum. Preparation for a pelvic MRI involves wearing comfortable clothes, removing all metallic accessories, and discussing potential contrast allergies.

Magnetic resonance imaging protocol and image acquisition

The MRI scans of all patients participating in the study were performed on the General

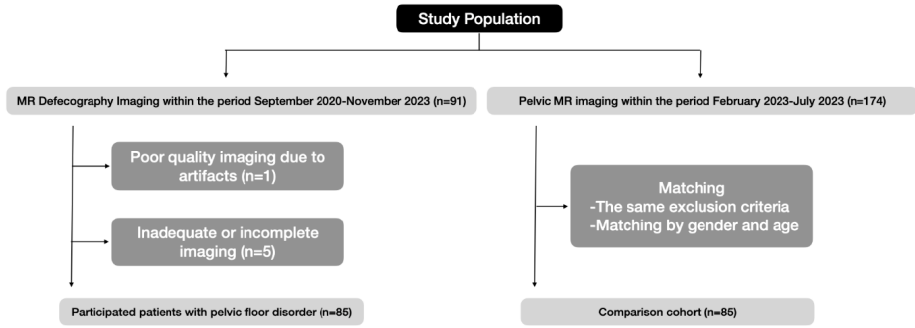


Figure 2. Flowchart of the study. MR, magnetic resonance.

Electric Signa™ Explorer MR 1.5T closed system device (GE Healthcare, Chicago, Illinois, IL, United States) using a phased array body coil without contrast material administration. The patient was positioned lying horizontally with the face and torso facing up with knees elevated on a pillow. To protect the scanner during imaging, adult diapers and disposable sheets were used. MR defecography consists of both statical and dynamical sequences including sagittal, coronal, and axial static T2-weighted images at rest, sagittal and coronal cine balanced steady-state GRE while squeezing and straining, and coronal cine balanced steady-state GRE images during defecation at least three times until the rectum is emptied as much as possible (detailed information is summarized in Table 1). The entire defecography procedure duration varied between 15 to 30 minutes. Pelvic MR protocol includes sagittal and axial T2-weighted, coronal fat-saturated T2-weighted images, axial T1-weighted fast-spin echo images, axial diffusion-weighted images with a *b* value of 50 and 800, and axial liver acquisition with volume acceleration (LAVA) images. In cases where contrast media was required, dynamic sagittal LAVA images (for female patients only) and axial contrast-enhanced LAVA images were obtained. Pelvic MRI lasts approximately 20–30 minutes (Table 2).

Routine magnetic resonance defecography and pelvic magnetic resonance imaging interpretation

Two radiologists (E.D. and A.B.Y.) evaluated the MRI in consensus. An independent radiologist (R.B.Y.), who was blinded to the outcomes, concurrently interpreted a randomly selected subset of 35 cases. All metric measurements including PCL, H, and M-lines at rest and defecation, ARA at rest, squeezing and defecation, levator plate angle (LPA, the angle between the levator plate and PCL) at maximal straining, and urethral angle (the angle between the urethra and PCL) were completed on midline sagittal MRI. The severity of cystocele and uterine prolapse was graded according to the depth of the herniation under the PCL as mild (less than 3 cm), moderate (3 to 6 cm), and severe (greater than 6 cm).[8,16,17] A rectocele is characterized by the rectal wall extending beyond the anticipated typical shape, with grading based on the extent of protrusion: small (<2 cm), moderate (2-4 cm), and large (>4 cm), determined by the depth of the bulge.[18] The assessment of pelvic diaphragm relaxation based on M-line lengths was evaluated using the H-line, M-line, and organ prolapse (HMO) classification system. A normal hiatal position was defined as an M-line measurement between 0 and 2 cm (grade 0), while mild descent was categorized as an M-line measurement ranging from 2 to 4 cm (grade 1). Moderate descent was characterized by an M-line

Table 1. Magnetic resonance defecography protocol

Plane	MR sequence	TR/TE (ms)	Matrix size/NEX	Slice thickness (mm)	Spacing (mm)	FOV (cm x cm)	Information type	Phase
Sagittal	Static, T2 PROPELLER	7570/152.32	288 x 288/4	4	1	25 x 25	Structural	Rest
Coronal	Static, T2 PROPELLER	3859/148.99	288 x 288/4	4	1	34 x 34	Structural	Rest
Axial	Static, T2 PROPELLER	5195/121.3	320 x 320/4	4	1	35 x 35	Structural	Rest
Sagittal midline	Dynamic, Cine FIESTA	4/1.4	256 x 288/4	6	1	25 x 25	Functional	Squeeze (Kegel)
Coronal	Dynamic, Cine FIESTA	5/2.02	256 x 288/4	6	1	25 x 25	Functional	Squeeze (Kegel)
Sagittal midline	Dynamic, Cine FIESTA	4/1.91	256 x 288/4	6	1	25 x 25	Functional	Strain (Valsalva)
Coronal	Dynamic, Cine FIESTA	5/2.01	256 x 288/4	6	1	25 x 25	Functional	Strain (Valsalva)
Sagittal midline	Dynamic, Cine FIESTA	5/1.9	256 x 288/4	6	1	25 x 25	Functional	Defecation (at least 3 times)

MR, magnetic resonance; TR, time of repetition; TE, time of echo; NEX, number of excitations; FOV, field of view.

Table 2. Pelvic magnetic resonance protocol

Plane	MR sequence	TR/TE (ms)	Matrix size/NEX	Slice thickness (mm)	Spacing (mm)	FOV (cm x cm)
Sagittal	T2 PROPELLER	6.121/113.74	288 x 288/4	5	1	25 x 25
Axial	T2 PROPELLER	4.146/96.62	300 x 300/4	5	1	32 x 32
Coronal	T2 FAT-SAT PROPELLER	6.446/96.82	300 x 300/4	5	1	32 x 32
Axial	T1 fast spin echo	552/10.28	320 x 224/4	5	1	32 x 32
Axial	Diffusion-weighted imaging *b* value: 50–800	5.835/78.90	256 x 288/4	5	1	32 x 32
Axial	Apparent diffusion coefficient	5.835/78.90	128 x 128/4	5	1	32 x 32
Axial	LAVA	6/3.15	280 x 192/4	6	1	32 x 32
Sagittal* (female protocol)	Dynamic, contrast-enhanced LAVA	3/1.84	320 x 192/4	4	1	32 x 32
Axial*	Contrast-enhanced LAVA	6/3.15	280 x 192/4	4	1	32 x 32

*Only obtained for contrast-enhanced studies. MR, magnetic resonance; TR, time of repetition; TE, time of echo; NEX, number of excitations; FOV, field of view; LAVA, liver acquisition with volume acquisition.

measurement between 4 and 6 cm (grade 2), and severe descent was indicated when the M-line measurement exceeded 6 cm (grade 3).[19] On the other hand, a normal hiatal width was defined as an H-line measurement between 0 and 6 cm (grade 0), while mild hiatal enlargement was categorized as an H-line measurement ranging from 6 to 8 cm (grade 1). Moderate enlargement was characterized by an H-line measurement between 8 and 10 cm (grade 2), and severe enlargement was indicated when the H-line measurement exceeded 10 cm (grade 3).[19] Urethral hypermobility is a condition of excessive horizontal translation (more than 30°) of the urethra due to a weak pelvic floor.[1]

Levator ani muscle manual segmentation and volumetric measurements

For analyzing the medical image data, a free and open-source imaging package software [three-dimensional (3D) Slicer version 5.2.2 for Mac OS X] was utilized. A radiologist with eight years' experience (A.B.Y.), and a radiology resident with five years' experience (R.B.Y.) manually segmented the LAM from the contiguous axial T2-weighted MRI slices using the "Segment Editor" module in the 3D Slicer software. The anterior boundary of the LAM is defined as the pubic symphysis, whereas the posterior boundary of the LAM is defined as the coccyx. The muscles surrounding the anal canal and rectum were delineated. Quantitative information, including the number of voxels, the volume of the muscle, minimum, maximum, mean, and median values, standard deviation, and surface area, derived using the "Segment Statistics" module, was noted. For each patient and control, 3D reconstruction models of the LAM were also created (Figures 3, 4). The average time to segment the LAM required 10 minutes per patient.

Statistical analysis

All statistical analyses were conducted using the SPSS 24.0 software (IBM Corp., Armonk, NY, USA). Metric measurements and quantitative segmentation results were reported with means and standard deviations. The normality of distribution was assessed using the Shapiro–Wilk test. Normally distributed data were compared using the independent samples t-test, while non-normally distributed data were evaluated using the Mann–Whitney U test. A chi-square test was used to compare the observed frequencies of categorical data. Subgroup comparisons within the patient group were done using the Kruskal–Wallis test. The receiver operator characteristic (ROC) curve was drawn to assess the sensitivity and specificity of the volumetric measurement, and the optimal cut-off value was selected. The intraclass correlation coefficient was used to estimate the interrater reliability of the MR defecography measurements. A *P* value of less than 0.05 was considered statistically significant.

Results

Demographic characteristics

The mean age was calculated as 53.54 ± 15.7 years for the patient group and 51.99 ± 13.4 years for the control group. There was no significant difference between the groups in terms of age. In both groups, the distribution of men and women was equal. Among 69 female participants, 14 in the case and 13 in the control groups had undergone a hysterectomy (*P* value, 0.632).

Routine magnetic resonance imaging of pelvic floor and magnetic resonance defecography findings

Pelvic floor measurements were performed in both the patient and control groups where applicable. The mean PCL length was calculated as 102.24 ± 9.9 mm, the H-line at rest was 49.35 ± 9.8 mm, and the M-line at rest was 15.97 ± 11.7 mm for the patient group. Conversely, the mean PCL length was calculated as 103.06 ± 10.9 mm, the H-line at rest as 31.06 ± 5.4 mm, and the M-line at rest as 6.12 ± 3.2 mm for the control group. The remaining measurements were only performed in the patient group. The H-line at defecation was calculated as 69.07 ± 17.1 mm, and the M-line at defecation was 43.21 ± 21.3 mm. The average ARA angle at rest was 96.11 ± 17.04°, 82.71 ± 17.8° at strain, and 113.01 ± 22.4° at defecation. The mean LPA at maximal straining was 37.86 ± 20.1°; 61.2 percent of the patients (n = 52) had urethral hypermobility. Only 4 patients had peritoneocele (Table 3). The most common pathologies were grade 1 cystocele (n = 36, 42.4%) and mild hiatal enlargement (n = 35, 41.2%) followed by grade 1 anterior rectocele (n = 32, 37.6%), grade 2 anterior rectocele (n = 26, 30.6%), and mild pelvic floor descent (n = 25, 29.4%). Data regarding POP and pelvic floor relaxation are outlined in Table 4.

Interrater reliability of magnetic resonance defecography assessment

To determine interrater reliability, 35 patients were selected randomly and two re-

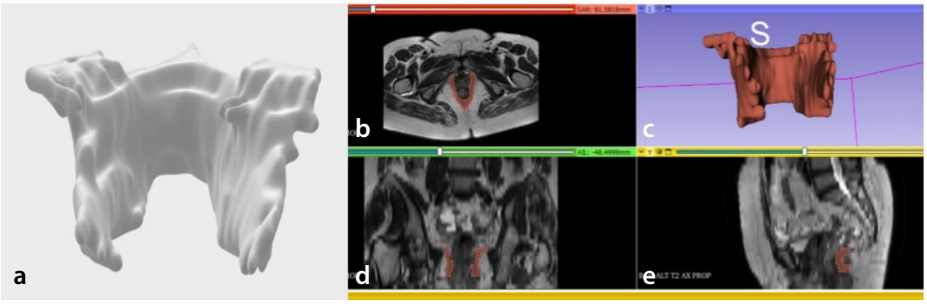
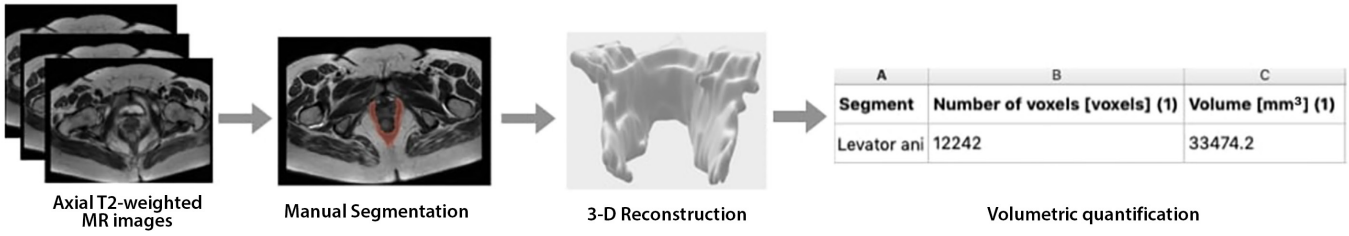


Figure 3. An example of levator ani muscle (LAM) segmentation and three-dimensional (3D) image reconstruction in a healthy individual. Image **(a)** represents a completed volume rendered segmentation model of the LAM. On the right is a screenshot of the 3D-Slicer software "Segment Editor" module. Image **(b)** shows an axial T2-weighted series, which are the source images we loaded for segmentation analysis. Images **(d, e)** show a coronal and sagittal view of the LAM, reconstructed by 3D-Slicer to edit the segment, and image C is a real-time 3D model of LAM.



A	B	C
Segment	Number of voxels [voxels] (1)	Volume [mm³] (1)
Levator ani	12242	33474.2

Figure 4. Pipeline of the study. MR, magnetic resonance.

viewers (blinded to each other) interpreted the PCL line at rest, H-line at rest and defecation, M-line at rest and defecation, ARA at rest, maximal strain and defecation, and LPA. The intraclass correlation analysis revealed excellent agreement (Table 5).[20]

Levator ani muscle volumetric measurement

The mean number of voxels was calculated as 12,896.5 ± 5202.9 for the patient group and 18,778.1 ± 6784.1 for the control group. The mean volume of LAM was quantified as 33,214.6 ± 11,884.6 mm^3 for the patient group and 48,107.9 ± 12,274.2 mm^3 for the control group. The mean surface area of the patient group was 15,425.8 ± 4,022.2 mm^2 and 19.458.4 ± 4,467.9 mm^2 for the control group (Table 3).

Association of levator ani muscle volume and pelvic floor dysfunction

Voxel numbers, LAM volumes, and surface area were higher in the control group. Since the data were not normally distributed, the number of voxels, segment volume, and surface area of the patients and controls' LAM were compared using the Mann–Whitney U test. For each variable, a statistically significant difference was observed (*P* values were <0.001). PFD is defined as the presence of conditions that may affect any compartments, including hiatal enlargement, pelvic floor descent, cystocele, uterine prolapse, rectocele, and peritoneocele. The ROC curve analyses were performed to evaluate the sensitivity and specificity of the volume and surface area measurement of the LAM on PFD. A cut-off value of 38,934.3 mm^3 was set with a 0.812 sensitivity and 0.8 specificity for the LAM volume. The area under the curve (AUC) was computed as 0.834. For surface area measurement, the AUC was calculated as 0.753, and the cut-off value was set as 16,639.4 mm^2 with a sensitivity of 0.753 and specificity of 0.706 (Figure 5). We also compared the mean volume of the LAM in the patient group, depending on the severity of hiatal enlargement, pelvic floor descent, and POP. When the disease worsened, no statistically significant change in the muscle volume was observed (*P* values were 0.440, 0.929, and 0.732, respectively).

Effect of gender on levator ani muscle volume

No statistically significant difference was observed between female (n = 69) and male

Table 3. Magnetic resonance imaging measurements of patient and control groups

	Patient group	Control group	*P* value
Age (year, mean ± SD)	53.54 ± 15.7	51.99 ± 13.4	0.489[a]
Gender distribution (M/F, n)	16/69	16/69	1[b]
Hysterectomy rate of women (n, and percent)	14 (16.5%)	13 (15.3%)	0.632[b]
PCL rest (mm, mean ± SD)	102.24 ± 9.9	103.06 ± 10.9	0.609[a]
H-line rest (mm, mean ± SD)	49.35 ± 9.8	31.06 ± 5.4	*P* < 0.001[c]
M-line rest (mm, mean ± SD)	15.97 ± 11.7	6.12 ± 3.2	*P* < 0.001[c]
H-line defecation (mm, mean ± SD)	69.07 ± 17.1	N/A	N/A
M-line defecation (mm, mean ± SD)	43.21 ± 21.3	N/A	N/A
ARA rest (°, mean ± SD)	96.11 ± 17.04	N/A	N/A
ARA strain (°, mean ± SD)	82.71 ± 17.8	N/A	N/A
ARA defecation (°, mean ± SD)	113.01 ± 22.4	N/A	N/A
LPA maximal straining (°, mean ± SD)	37.86 ± 20.1	N/A	N/A
LAM volume (mm^3, mean ± SD)	33,214.6 ± 11884.6	48,107.9 ± 12274.2	*P* < 0.001[a]
LAM surface area (mm^2, mean ± SD)	15425.8 ± 4022.2	19,458.4 ± 4467.9	*P* < 0.001[a]
LAM number of voxels (mean ± SD)	12896.5 ± 5202.9	18,778.1 ± 6784.1	*P* < 0.001[a]
	n, percentage		
Urethral hypermobility	52 (61.2%)	N/A	N/A
Peritoneocele	4 (4.7%)	N/A	N/A

[a]Independent samples t-test results; [b]Fisher's exact test results; [c]Mann–Whitney U test results. SD, standard deviation; M, male; F, female; n, number; PCL, pubococcygeal line; ARA, anorectal angle; LPA, levator plate angle; LAM, levator ani muscle.

Table 4. Magnetic resonance defecography assessment of patient group

	Absent	Grade 1	Grade 2	Grade 3
	(n, percent)	(n, percent)	(n, percent)	(n, percent)
Cystocele	41 (48.2%)	36 (42.4%)	7 (8.2%)	1 (1.2%)
Uterine prolapse	40 (72.7%)	11 (20%)	4 (7.3%)	0 (0%)
Anterior rectocele	21 (24.7%)	32 (37.6%)	26 (30.6%)	6 (7.1%)
	Healthy	Mild	Moderate	Severe
	(n, percent)	(n, percent)	(n, percent)	(n, percent)
Hiatal enlargement (n, percent)	26 (30.6%)	35 (41.2%)	20 (23.5%)	4 (4.7%)
Pelvic floor descent (n, percent)	15 (17.6%)	25 (29.4%)	23 (27.1%)	22 (25.9%)

n, number.

(n = 16) controls concerning the number of voxels, LAM volumes, and surface areas (*P* values were 0.419, 0.449, and 0.449, respectively).

Effect of aging on levator ani muscle volume

A weak negative correlation was observed between the age and LAM volume in only the patient group (r: −0.227, *P* value, 0.037). All participants (n = 170) were divided into two groups according to their age; individuals older than 50 years comprised the first group (n = 93) and the remaining individuals represented the second group (n = 77). The average volume of the LAM was significantly lower in the first group (*P* value, 0.019).

Effect of history of hysterectomy on levator ani muscle volume

No statistically significant difference was observed among the women controls based on their history of hysterectomy (n = 69, and *P* value is 0.671).

Discussion

In the diagnosis of PFD, clinical examination is generally indefinite in isolation and may lead to the underestimation of pathologies and involved compartments.[2,15] Various imaging modalities are employed to assess the pelvic floor, particularly the LAM, which represents the active component, including translabial–endovaginal US, CT, fluoroscopy, and MRI.[15,21] MR defecography imaging offers exceptional spatial and contrast resolution, enabling the delineation of even small tears or injuries, and providing detailed anatomical and functional information.[3] In this regard, it plays a crucial role in the concurrent assessment of pelvic organs and pelvic floor muscles without radiation exposure and contrast media administration, unlike dynamic fluoroscopic defecography.[22]

In this retrospective case-control study, our primary objective was to investigate the relationship between the volume of the LAM and PFD, and its potential contribution to routine MR defecography measurements. As we assumed, our results confirmed the presence of a correlation between a decreased LAM volume and PFD. Patients with an LAM volume calculated below 38,934.3 mm^3 have a higher tendency toward PFD. However, contrary to our initial hypothesis, we did not observe a linear correlation between the severity of POP or pelvic floor relaxation and muscle volume. The small number of subgroups in the patient group may have affected the reliability of the subgroup comparison results. Further work on larger populations is thus needed to validate our results.

Previous studies focused on LAM segmentation based on transperineal or endovaginal US and MRI.[10-12] Rabbat et al.[23] proposed using deep learning algorithms to automate LAM segmentation as a means to improve the diagnostic ability of the US. Another study utilizing MRI suggests a modified Chan–Vese segmentation model, which uses intensity information and the influence of shape to segment the LAM in axial slices.[24] Compared to manual segmentation, automated segmentation models may shorten the time taken to complete the procedure, which can assist physicians in executing muscle identification, segmentation, 3D reconstruction, and automatic volume measurement.[22]

In the current study, we chose manual segmentation, despite the extended time of the process, because it is the reference standard. Despite employing manual segmentation, the fact that one segmentation could be completed in approximately 10 minutes demonstrates its feasibility and appropriateness for clinical work.

As predicted in prior studies, our study participants mostly included women. One of our secondary objectives was to assess the influence of gender on LAM volume in healthy participants. A publication by Cheung et al.[25] reports that the LAM has extraordinary androgen sensitivity in rodents and humans. We hypothesized that the volume of the LAM in women may be lower than in men, potentially giving rise to vulnerability to pelvic floor disorders. However, we found that LAM volume, LAM surface area, and the number of voxels were similar for both genders. This suggests that it is the difference in processes, such as pregnancy and childbirth, rather than gender, which may be at play. Cheung et al.[25] investigated the LA and walking muscle volumes in patients with prostate cancer receiving androgen deprivation therapy and found that the ther-

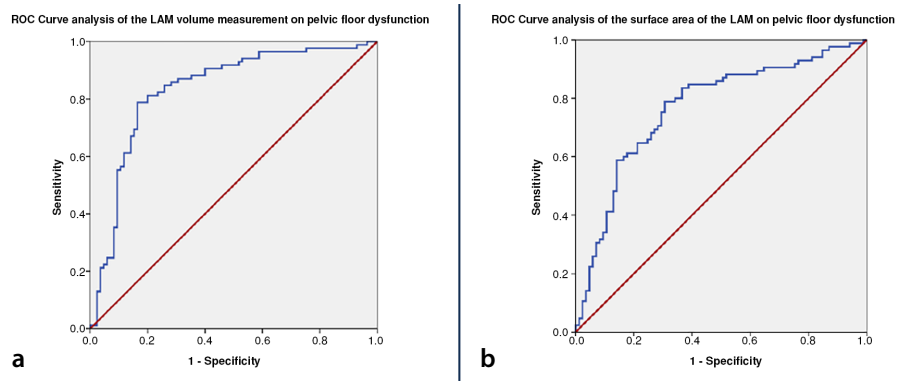


Figure 5. The receiver operating curve analyses of the LAM volume measurement and surface area in the presence of pelvic floor dysfunction. Volumetric measurement: AUC: 0.834, cut-off value: 38,934.3 mm^3 with a sensitivity of 0.812 and specificity of 0.8 **(a)**. Surface area measurement: AUC: 0.753, cut-off value 16,639.4 mm^2 with a sensitivity of 0.753 and specificity of 0.706 **(b)**. LAM, levator ani muscle; AUC, area under the curve.

Table 5. Interobserver correlations of routine magnetic resonance defecography measurements

	Observer 1	Observer 2	ICC	*P* value
Measurements (n = 35)	(Mean ± SD)	(Mean ± SD)	(95% CI)	
PCL rest (mm)	100.49 ± 10.7	102.09 ± 10.1	0.985 (0.970–0.992)	<0.001
H-line rest (mm)	50.9 ± 9.9	51.2 ± 10.1	0.992 (0.985–0.996)	<0.001
M-line rest (mm)	19.5 ± 12.2	18.6 ± 12.1	0.976 (0.952–0.988)	<0.001
H-line defecation (mm)	71.9 ± 18.5	74.2 ± 18.5	0.955 (0.914–0.977)	<0.001
M-line defecation (mm)	47.5 ± 21.5	48.8 ± 21.2	0.967 (0.936–0.983)	<0.001
ARA rest (°)	99.8 ± 15.7	98.7 ± 14.6	0.948 (0.9–0.973)	<0.001
ARA strain (°)	84.9 ± 18.9	83.4 ± 17.2	0.943 (0.891–0.971)	<0.001
ARA defecation (°)	112.3 ± 22.8	113.7 ± 22.4	0.991 (0.982–0.995)	<0.001
LPA maximal straining (°)	35.6 ± 19.3	37.3 ± 19.6	0.967 (0.936–0.983)	<0.001

ICC, intraclass correlation coefficient; CI, confidence interval; n, number; SD, standard deviation; PCL, pubococcygeal line; ARA, anorectal angle; LPA, levator plate angle.

apy process caused muscle volume loss. The limitation of their study is that patients with prostate cancer often concurrently receive radiation therapy (RT); it should thus be kept in mind that the outcomes may have been influenced by RT rather than the androgen sensitivity of the muscle. There is therefore a need to elucidate the molecular mechanism of androgen sensitivity of the LAM.

Wyman et al.[26] conducted a study to evaluate the relationship between LAM volume, age, and body mass index (BMI). Interestingly, the results showed that an increased age in female participants correlated with an elevated LAM volume; however, there was no correlation between BMI and muscle volume. The study authors assumed these results to be related to a reduction in the strength and integrity of the LAM resulting from sarcopenia.[26] In their previous paper, the authors evaluated whether the estimated LA subtended volume (eLASV) could predict the success of POP surgical treatment. Their results indicated that patients with a higher eLASV had an increased risk of surgical failure.[12] The main question concerning this study is the absence of a clear definition of the volumetric measurement process described in their paper; accordingly, these findings must be interpreted with caution. In contrast to earlier findings presented by Wyman et al.[26], our study reveals that older individuals have a lower LAM volume.

A similar study to the current research was carried out by Nandikanti et al.[27] to evaluate the LA bowel volume variation between resting and straining states in patients with POP and healthy controls. The results indicate that hiatus size and bowel volume change during straining.[27] To the best of our knowledge, this is the first study comparing LAM volume and routine MR defecography measurements.

This study has some limitations. Due to the retrospective design of the research, we were unable to gather selected information, including BMI, history of pregnancy/vaginal birth, and abortion, which indicate a constant relationship with pelvic floor insufficiency. Additionally, we only assessed LAM volume; however, anal and urethral sphincters, internal obturator muscle, coccygeus muscle, and perineal muscles also play roles in PFD. Finally, the technique for the axial T2 pelvis MRI for the controls was not exactly matched to what was used for MR defecography.

In conclusion, a lower LAM volume appears to show a direct correlation with an increased probability of PFD. Our results did not reveal a linear correlation between the severity of POP or pelvic floor relaxation and muscle volume. Future research with a larger participant pool is warranted to further investigate this matter.

Acknowledgments

We would like to thank our MRI team.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Salvador JC, Coutinho MP, Venâncio JM, Viamonte B. Dynamic magnetic resonance imaging of the female pelvic floor-a pictorial review. Insights Imaging. *Insights Imaging*. 2019;10(1):4. [Crossref]
2. Kanmaniraja D, Arif-Tiwari H, Palmer SL, et al. MR defecography review. *Abdom Radiol (NY)*. 2021;46(4):1334-1350. [Crossref]
3. Wu JM, Vaughan CP, Goode PS, et al. Prevalence and trends of symptomatic pelvic floor disorders in U.S. women. *Obstet Gynecol*. 2014;123(1):141-148. [Crossref]
4. Maccioni F, Alt CD. MRI of the pelvic floor and MR defecography. 2018 Mar 21. In: Hodler J, Kubik-Huch RA, von Schulthess GK, editors. Diseases of the Abdomen and Pelvis 2018-2021: Diagnostic Imaging - IDKD Book [Internet]. Cham (CH): Springer; 2018. Chapter 2. [Crossref]
5. Margulies RU, Hsu Y, Kearney R, Stein T, Umek WH, DeLancey JO. Appearance of the levator ani muscle subdivisions in magnetic resonance images. Appearance of the levator ani muscle subdivisions in magnetic resonance images. *Obstet Gynecol*. 2006;107(5):1064-1069. [Crossref]
6. Gowda SN, Bordoni B. Anatomy, abdomen and pelvis: levator ani muscle. [Updated 2022 Oct 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. [Crossref]
7. Li M, Wang B, Liu X, Qiao P, Jiao W, Jiang T. MR defecography in the assessment of anatomic and functional abnormalities in stress urinary incontinence before and after pelvic reconstruction. *Eur J Radiol*. 2020;126:108935. [Crossref]
8. El Sayed RF, Alt CD, Maccioni F, et al. Magnetic resonance imaging of pelvic floor dysfunction - joint recommendations of the ESUR and ESGAR Pelvic Floor Working Group. *Eur Radiol*. 2017;27(5):2067-2085. [Crossref]
9. Haliloglu N, Erden A. Magnetic resonance defecography findings of dyssynergic defecation. *Pol J Radiol*. 2022;87(1):181-185. [Crossref]
10. Zhang X, Xiang Y, Yao J, et . Automatic segmentation of the female pelvic floor muscles on MRI for pelvic floor function assessment. *Quant Imaging Med Surg*. 2023;13(7):4181-4195. [Crossref]
11. Asif Z, Tomashev R, Peterkin V, et al. Levator ani muscle volume and architecture in normal vs. muscle damage patients using 3D endovaginal ultrasound: a pilot study. *Int Urogynecol J*. 2023;34(2):581-587. [Crossref]
12. Wyman AM, Tanner JP, Bassaly R, Hoyte L, Greene K. Levator ani muscle (LAM) anatomical volume compared to age and body mass index (BMI) of patients. *Int Urogynecol J*. 2020;31(7):1457-1461. [Crossref]
13. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015;277:826-832. [Crossref]
14. https://www.ohsu.edu/school-of-medicine/diagnostic-radiology/mr-defecography-wo-body-protocol [Crossref]
15. Lalwani N, Khatri G, El Sayed RF, et al. MR defecography technique: recommendations of the society of abdominal radiology's disease-focused panel on pelvic floor imaging. *Abdom Radiol (NY)*. 2021;46(4):1351-1361. [Crossref]
16. Chamié LP, Ribeiro DMFR, Caiado AHM, Warmbrand G, Serafini PC. Translabial US and dynamic mr imaging of the pelvic floor: normal anatomy and dysfunction. *Radiographics*. 2018;38(1):287-308.
17. Yoon I, Gupta N. Pelvic prolapse imaging. [Updated 2023 May 1]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. [Crossref]
18. Thapar RB, Patankar RV, Kamat RD, Thapar RR, Chemburkar V. MR defecography for obstructed defecation syndrome. *Indian J Radiol Imaging*. 2015;25(1):25-30. [Crossref]
19. Picchia S, Rengo M, Bellini D, et al. Dynamic MR of the pelvic floor: influence of alternative methods to draw the pubococcygeal line (PCL) on the grading of pelvic floor descent. *Eur J Radiol Open*. 2019;6:187-191. [Crossref]
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174. [Crossref]
21. Guo M, Li D. Pelvic floor images: anatomy of the levator ani muscle. *Dis Colon Rectum*. 2007;50(10):1647-1655. [Crossref]
22. Boyadzhyan L, Raman SS, Raz S. Role of static and dynamic MR imaging in surgical pelvic floor dysfunction. *Radiographics*. 2008;28(4):949-967. [Crossref]
23. Rabbat N, Qureshi A, Hsu KT, et al Automated segmentation of levator ani muscle from 3D endovaginal ultrasound images. *Bioengineering (Basel)*. 2023;10(8):894. [Crossref]
24. Ma Z, Jorge RN, Tavares JM. A shape guided C-V model to segment the levator ani muscle in axial magnetic resonance images. *Med Eng Phys*. 2010;32(7):766-774. [Crossref]

25. Cheung AS, Cunningham C, Ko DD, et al. Selective loss of levator ani and leg muscle volumes in men undergoing androgen deprivation therapy. *J Clin Endocrinol Metab*. 2019;104:2229-2238. [Crossref]

26. Wyman AM, Rodrigues Jr AA, Hahn L, et al. Estimated levator ani subtended volume: a novel assay for predicting surgical failure after uterosacral ligament suspension. *Am J Obstet Gynecol*. 2016;214:611. [Crossref]

27. Nandikanti L, Sammarco AG, Chen L, Ashton-Miller JA, DeLancey JO. Levator bowl volume during straining and its relationship to other levator measures. *Int Urogynecol*. 2019;30(9):1457-1463. [Crossref]

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.232460

ABDOMINAL IMAGING

ORIGINAL ARTICLE

Value of perfusion parameters from golden-angle radial sparse parallel dynamic contrast-enhanced magnetic resonance imaging in predicting pathological complete response after neoadjuvant chemoradiotherapy for locally advanced rectal cancer

Yu-Ning Pan[1,2]
Meng-Yin Gu[3]
Quan-Liang Mao[2]
Yu-Guo Wei[4]
Lin Zhang[1]
Guang-Yu Tang[1]

[1]Shanghai Tenth People's Hospital of Tongji University, Department of Radiology, Shanghai, China

[2]The First Affiliated Hospital of Ningbo University, Department of Radiology, Ningbo, China

[3]Ningbo University, Department of Medical College, Ningbo, China

[4]GE Healthcare, Global Medical Service, Advanced Analytics, Hangzhou, China

PURPOSE

Non-invasive methods for predicting pathological complete response (pCR) after neoadjuvant chemoradiotherapy (nCRT) can provide distinct leverage in the management of patients with locally advanced rectal cancer (LARC). This study aimed to investigate whether including the golden-angle radial sparse parallel (GRASP) dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) perfusion parameter (K_{trans}), in addition to tumor regression grading (TRG) and apparent diffusion coefficient (ADC) values, can improve the predictive ability for pCR.

METHODS

Patients with LARC who underwent nCRT and subsequent surgery were included. The imaging parameters were compared between patients with and without pCR. Receiver operating characteristic (ROC) curve analysis was used to evaluate the predictive ability of these parameters for pCR.

RESULTS

A total of 111 patients were included in the study. A pCR was obtained in 32 patients (28.8%). MRI-based TRG (mrTRG) showed a negative correlation with pCR ($r = -0.61$, $P < 0.001$), and the average ADC value showed a positive correlation with pCR ($r = 0.62$, $P < 0.001$). Before nCRT, K_{trans} in the pCR group was significantly higher than in the non-pCR group (1.30 ± 0.24 vs. 0.88 ± 0.34, $P < 0.001$), but no difference was identified after nCRT. Following ROC curve analysis, the area under the curve (AUC) of mrTRG (level 1–2), average ADC value, and K_{trans} value for predicting pCR were 0.738 [95% confidence interval (CI): 0.65–0.82], 0.78 (95% CI: 0.69–0.86), and 0.84 (95% CI: 0.77–0.92), respectively. The model combining the three parameters had significantly higher predictive ability for pCR (AUC: 0.94, 95% CI: 0.88–0.98).

CONCLUSION

The use of a combination of the GRASP DCE-MRI K_{trans} with mrTRG and ADC can lead to a better pCR predictive performance.

KEYWORDS

Rectal cancer, locally advanced, magnetic resonance imaging, neoadjuvant chemoradiotherapy, tumor regression grading, complete response

Corresponding authors: Lin Zhang, Guang-Yu Tang

E-mail: lynn122500@hotmail.com, tgy17@tongji.edu.cn

Received 18 August 2023; revision requested 13 September 2023; last revision received 05 February 2024; accepted 10 February 2024.

Epub: 18.03.2024

Publication date: 08.07.2024

DOI: 10.4274/dir.2024.232460

The prevalence of colorectal cancer is projected to rise by 60% in 2030,[1] with morbidity and mortality rates rapidly increasing in many low- and middle-income countries. Rectal cancer (RC) accounts for approximately 30% of all cases of colorectal cancer.[2] Neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision is the standard treatment for locally advanced RC (LARC).[3]

You may cite this article as: Pan YN, Gu MY, Mao QL, Wei YG, Zhang L, Tang GY. Value of perfusion parameters from golden-angle radial sparse parallel dynamic contrast-enhanced magnetic resonance imaging in predicting pathological complete response after neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Diagn Interv Radiol.* 2024;30(4):228-235.

Approximately 50%–60% of patients with LARC experience tumor regression after nCRT, and 15%–30% of these patients achieve pathological complete response (pCR),[4] which is defined as the absence of cancer cells in the surgically resected samples. Therefore, the pathological stage for a pCR specimen is T0 N0 M0.[5] Achievement of pCR does not guarantee long-term survival;[6,7] however, the local recurrence rate and distant metastasis rate of patients achieving pCR are lower than those of patients not achieving pCR, and the 5-year survival rate is higher than that of patients who do not achieve pCR.[4,8] Therefore, pCR has remained the objective of nCRT.

The optimal treatment approach for patients who achieve pCR after nCRT is an important issue. Instead of the traditional radical surgery, some surgeons recommend non-operative treatment to avoid these complications.[7-9] Before choosing the therapeutic method, it is crucial to develop an accurate and non-invasive strategy for identifying individuals who could have a pCR.

Rectal magnetic resonance imaging (MRI) has become the standard method to evaluate the efficacy of nCRT in the treatment of LARC. In 2011, Patel et al.[10] proposed tumor regression grading (TRG) based on the proportion of lesion fibrosis and residual tumor on MRI (mrTRG). However, the traditional morphological qualitative assessment based on a T2-weighted (T2W) sequence has suboptimal performance in observing and distinguishing residual tumors and treatment-related changes. As a result, radiologists may over-stage the tumor after nCRT,[11] particularly since it is not effective in predicting pCR,[12] and the diagnostic accuracy is approximately 50%.[13]

Main points

- A non-invasive method to identify individuals who achieved a pathological complete response (pCR) after neoadjuvant chemoradiotherapy (nCRT) is important to avoid excessive medical treatment.
- The value of golden-angle radial sparse parallel (GRASP) dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) in predicting the therapeutic effect of nCRT in tumors is still controversial.
- By comparing with final pathological outcomes, the diagnostic accuracy of imaging parameters obtained from GRASP DCE-MRI was assessed.
- Combining the GRASP DCE-MRI perfusion parameter value with tumor regression grading and apparent diffusion coefficient values can lead to a better pCR predictive performance.

Recently, there has been a need for integrating multiple imaging evaluation methods to enable a more comprehensive characterization of tumor biology and therapeutic response.[14] Dynamic contrast-enhanced MRI (DCE-MRI) can reflect blood vessel permeability by displaying hemodynamic changes and can enable an assessment of tissue perfusion and oxygen levels at the macro level.[15] However, due to the influence of respiratory movement and temporal resolution, its value in predicting the therapeutic effect of nCRT in tumors is still controversial.[16-18] The golden-angle radial sparse parallel (GRASP) MRI sequence has recently been applied in clinical settings. This technique integrates the advantages of StarVIBE and TWISTIBE sequences and combines motion-insensitive, golden-angle, star-stacked acquisition and compressed sensing reconstruction to improve temporal resolution. The artifacts caused by patient and intestinal motion are reduced by radial acquisition.[19] The GRASP technique has been shown to have high accuracy in imaging motion-sensitive organs such as kidneys, liver, and prostate,[20-22] as well as the rectum.[19] However, the use of GRASP DCE-MRI perfusion parameters (K_{trans}) to predict pCR has not yet been investigated.

This study investigated whether the additional GRASP DCE-MRI K_{trans} value, based on the mrTRG and apparent diffusion coefficient (ADC) values of T2W imaging (T2WI), can enable a more accurate prediction of pCR after nCRT for LARC.

Methods

Study population

This was a retrospective study. The study protocol was approved by the Ethics Committee of The First Affiliated Hospital of Ningbo University (approval number: 2022-R01025). The informed consent of patients was waived due to the nature of the study. Clinicopathologic data of patients with RC who were admitted to the hospital between January 2020 and August 2022 were retrospectively analyzed. The inclusion criteria were as follows: (1) RC was pathologically confirmed by colonoscopy, and LARC was confirmed by preoperative MRI (cT3-4 and/or cN+), and all patients underwent nCRT followed by radical total mesorectal resection; (2) the distal margin of the lesion was <12 cm from the anus; (3) there were no distant metastases. The exclusion criteria were as follows: (1) incomplete nCRT; (2) total mesorectal excision was not performed; (3) the time interval from nCRT to operation was >16 weeks; (4) there was a lack of complete MRI or postoperative pathological data.

Neoadjuvant chemotherapy protocol

All patients received conventional long-term concurrent chemoradiotherapy. Gross tumor volume included the primary rectal mass and metastatic lymph nodes, and clinical target volume included the mesenteric region, anterior sacral lymph nodes, internal iliac lymph nodes, and obturator lymph node drainage area. External iliac lymph nodes were irradiated if T4 tumors invaded the anterior structures (male prostate or female vagina) and/or obturator lymph node metastasis occurred. The total dose of radiation was 45.5–50.4 Gy (25–28 times), and the single dose was 1.8–2.0 Gy. Radiotherapy was administered in combination with oral capecitabine (825 mg/m^2) twice a day. One cycle of XELOX (capecitabine and oxaliplatin) consolidation chemotherapy was administered 3–4 weeks after the completion of radiotherapy. Radical surgery was performed 8–12 weeks after radiotherapy.

Magnetic resonance imaging examinations

All patients underwent MRI examinations twice. The first examination was 1 week before nCRT, and the second examination was 8 weeks after nCRT. A Siemens Vida 3.0 T scanner (Erlangen, Germany) and 16-channel abdominal coil were applied. The patient was placed in the supine position, and the foot was scanned first. Scanning protocols included high-resolution T2WI, diffusion-weighted imaging (DWI), and GRASP DCE-MRI. Scanning directions included axial, coronal, sagittal, and oblique planes (Table 1). The GRASP DCE-MRI contrast agent was Gd-DTPA (0.1 mmol/kg, 2 mL/s, Hengrui Medicine), and a star K-space trajectory of a golden-angle stack using a 3D gradient echo sequence was implemented. The minimum sampling time was 150 s, and a total of 1,586 radial spokes were obtained consecutively within an interval of 185 s.

Tumor regression grading

Pathologic TRG (pTRG) grading was performed according to the criteria proposed by Mandard et al.[23], as follows: pTRG0 (pCR): no tumor cells; pTRG1: single or small clusters of tumor cells; pTRG2: fibrosis more than tumor residual; pTRG3: fibrosis less than tumor residual; pTRG4: free of fibrosis with extensive tumor residue.

Image analysis

All image analyses and measurements were performed at the post-processing workstation (Siemens, Germany) using the measurement tools provided by the workstation. Measurements of ADC values and DCE-MRI parameters were performed by two senior radiologists (YN Pan and L Zhang) with more than 10 years of experience in this field. The radiologists selected three regions of interest (ROIs) in the plane of maximum tumor size on the original T2W image. The same ROI was then automatically overlaid on the DWI, ADC, and GRASP DCE-MRI K_{trans} images. Each ROI had an area ≥4 mm^2. The average value of the three ROI areas was taken as the final result. When obtaining the ROI, the intestinal lumen, artifacts, and blood vessels were not included. Notably, the radiologists who performed the ROI measurements were blinded to the pathological outcomes to minimize the likelihood of selection bias during the analysis. If the boundary of the residual tumor could not be determined clearly, the ROI was placed in the region corresponding to the tumor area before nCRT. Since the ADC map had fewer pixels and the ROI area after treatment was small, only the ADC_{mean} obtained from the ROI placement was calculated. In accordance with the Mercury Group's definition,[10] the mrTRG grading was performed on post-treatment T2WI images. Subsequently, mrTRG grading was performed using the following criteria: grade 1-linear or crescent-shaped body, mucosa or submucosa with a 1–2 mm scar or rectum wall clearly normalized; grade 2-dense fibrosis, no significant residual tumor; grade 3-more than 50% fibrosis or mucous, residual tumor signals can be seen; grade 4-small areas of fibrosis or mucus, but mostly tumors; grade 5-identical in appearance to the primary tumor or tumor progression. The mrTRG grades 1 and 2 were defined as a clinically complete response (cCR) (Figure 1). All scanned images were transferred to a Siemens workstation running syngo.via for post-processing. The Tofts two-compartment model was used for the calculation. The artery input function was selected in "fast" mode to obtain permeability-related parameters in the ROI through measurements. These parameters included the volume transfer constant (K_{trans}), extracellular extravascular space volume fraction (V_e), and rate constant (K_{ep}). Pre-treatment values of these parameters were utilized as primary measures.

Statistical analysis

The SPSS 22.0 and R (4.1.3) software packages were used for statistical analyses. Student's t-test and the Wilcoxon rank sum test were used to compare the values of ADC, mrTRG, K_{trans}, K_{ep}, and V_e between patients with and without pCR after nCRT. The intraclass correlation coefficient (ICC) was used to assess the consistency between the two radiologists in evaluating the various parameters. Receiver operating characteristic (ROC) curve analysis was performed to evaluate the predictive value of the cCR (mrTRG1–2 level), average ADC value, and pre-nCRT K_{trans} value for pCR. The optimal threshold was determined by the Youden index, and the sensitivity, specificity, positive predictive value, and negative predictive value were calculated. Multivariate logistic regression was used to construct a model to predict pCR with backward stepwise selection. The Delong test was used to analyze the difference in diagnostic performance among ROC curves, and *P* values of <0.05 were considered indicative of statistical significance.

Table 1. Patient characteristics and pathological outcomes of the study cohort

	Overall (n = 111)	pCR (n = 32)	Non-pCR (n = 79)	*P* value
Age (years) ± SD	62.3 ± 10.6	62.9 ± 9.6	61.8 ± 11.2	0.49
Sex, n				
M	66	19	47	0.35
F	45	13	32	
The pathological types, n				
Canalicular adenocarcinoma	81	23	58	0.37
Papillary adenocarcinoma	19	5	14	
Mucinous adenocarcinoma	11	4	7	
Tumor differentiation, n				
Well-differentiated	21	13	8	0.03
Moderately differentiated	68	15	53	
Poorly differentiated	22	4	18	

SD, standard deviation; pCR, pathological complete response; M, male; F, female.

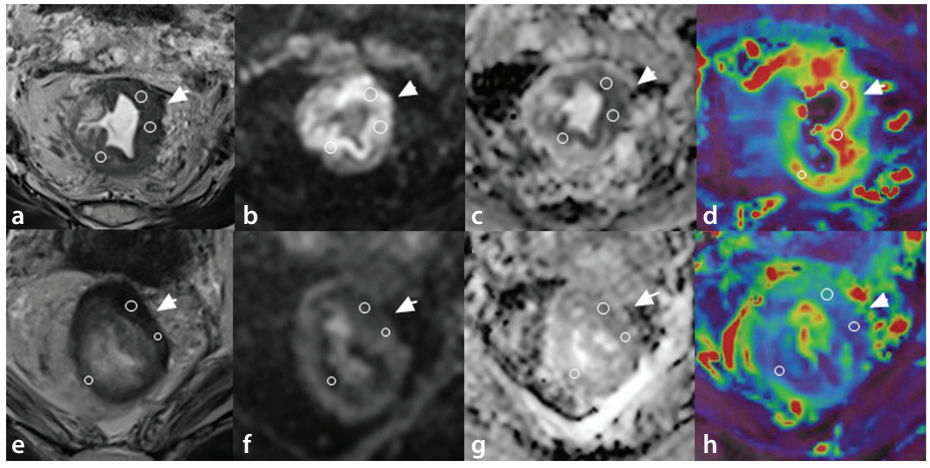


Figure 1. Magnetic resonance imaging of a 68-year-old patient with rectal cancer who underwent neoadjuvant chemoradiotherapy (nCRT) and postoperative pathological specimen identified achievement of complete response. **(a)** Before treatment, the tumor was mainly located on the left side, involving 3/4 perimeter of the rectum, and the outer membrane was involved (cT3N0, white arrow). **(b)** Before treatment, the tumor area was significantly limited in diffusion (white arrow). **(c)** Apparent diffusion coefficient (ADC) image before treatment showed a low signal in the tumor area (white arrow). **(d)** Golden-angle radial sparse parallel dynamic contrast-enhanced magnetic resonance imaging (GRASP DCE-MRI) perfusion parameter (K_{trans}) image: the tumor area is dominated by high signals with red on presentations (white arrow). **(e)** The mass was significantly reduced and fibrotic after nCRT (white arrow). **(f)** After nCRT, the diffusion restriction on the diffusion-weighted imaging disappeared, leaving only a few high signal areas (white arrow). **(g)** No obvious low signal area was found in ADC values after nCRT (white arrow). **(h)** GRASP DCE-MRI K_{trans} image; the tumor area is dominated by a blue signal (white arrow).

Results

Baseline characteristics

A total of 285 patients with RC were admitted during the study reference period. Of

these, 111 patients (45 women, 66 men, mean age: 62.3 ± 10.6 years) met the criteria for inclusion in the study. The patient selection flowchart is shown in Figure 2. The distribution of pathological types in the cohort was as follows: 81 cases of canalicular adenocarcinoma, 19 cases of papillary adenocarcinoma, and 11 cases of mucinous adenocarcinoma. Of the 111 cases, 21 were well-differentiated, 68 were moderately differentiated, and 22 were poorly differentiated. According to postoperative pathological specimens, pCR was achieved in 32 cases (28.8%).

Results of imaging evaluation

The grade of tumor regression was evaluated on T2WI. Five patients had mrTRG grade 1, 18 patients had mrTRG grade 2, 68 patients had mrTRG grade 3, 17 patients had mrTRG grade 4, and 3 patients had mrTRG grade 5. In total, 23 patients (20.7%) experienced cCR.

Post-treatment ADC values ranged from (0.83 ± 0.12) × 10^{-3} mm^2 to (2.6 ± 0.25) × 10^{-3} mm^2. After nCRT, the values of K_{trans} (0.96 ± 0.40 vs. 0.44 ± 0.25, $P < 0.001$) and K_{ep} (0.69 ± 0.54 vs. 0.55 ± 0.38, $P = 0.02$) were both significantly decreased in all patients. However, V_e showed no significant decrease after treatment (0.59 ± 0.36 vs. 0.54 ± 0.26, $P = 0.12$) (Table 2).

The two radiologists showed good consistency in evaluating mrTRG (ICC: 0.81, 95% CI: 0.77–0.88), ADC (ICC: 0.86, 95% CI: 0.84–0.97), K_{trans} (ICC: 0.88, 95% CI: 0.76–0.87), K_{ep} values (ICC: 0.68, 95% CI: 0.62–0.78), and V_e (ICC: 0.69, 95% CI: 0.63–0.76).

Correlation between imaging evaluation parameters and pathological outcomes

The pathological results showed that 32 patients (28.8%) achieved pCR (pTRG: 0) (Supplementary Table 1). The relationship between mrTRG and pTRG is presented in Table 3. Before nCRT, K_{trans} in the pCR group was significantly higher than in the non-pCR group, but there was no significant difference in K_{ep} or V_e between the two groups. After treatment, there was no significant difference in the above parameters between the two groups (Table 2 and Figure 1). Univariate logistic regression was performed to investigate the relationship between the mrTRG grading, post-treatment ADC value, and pathological outcomes. The results suggested that mrTRG grade 3 patients had a significantly lower probability of achieving pCR compared with grade 1 patients ($B = -2.56$, $P = 0.032$). The post-treatment ADC value was significantly correlated with the outcome ($B = 4.91$, $P < 0.001$).

Predictive performance of golden-angle radial sparse parallel dynamic contrast-enhanced magnetic resonance imaging parameters for pathological complete response

According to the ROC curve analyses, the area under the curve (AUC) of mrTRG (level 1–2), average ADC value (optimal threshold 1.05 × 10^{-3} mm^2), and K_{trans} value (optimal threshold 0.95/min) for predicting pCR were 0.738 (95% CI: 0.646–0.817), 0.782 (95% CI: 0.692–0.855), and 0.844 (95% CI: 0.772–0.916), respectively. The model combining the three parameters had the highest AUC (0.942, 95% CI: 0.881–0.977) (Figure 3 and Table 4). The DeLong test showed that the ability of the model to predict pCR when combining all three parameters was better than that of mrTRG, ADC value, and K_{trans} value alone ($P = 0.015$, 0.023, and 0.030, respectively) but not better than the model combining mrTRG and K_{trans} ($P = 0.099$).

Discussion

In the current study, we investigated the use of GRASP DCE-MRI for predicting pCR in patients with RC who underwent nCRT. The parameters (mrTRG, ADC, and K_{trans}) obtained from GRASP DCE-MRI imaging were used to quantify the predictive ability of the technique. Our results demonstrated that GRASP DCE-MRI imaging can predict pCR well. The predictive ability of the model combining the three parameters was ideal, with an AUC as high as 0.942. To the best of our knowledge, this is the first study to explore the role of GRASP DCE-MRI in predicting pCR for patients with RC.

For evaluating the efficacy of nCRT, MRI has the advantage of being non-invasive, and the mrTRG is a reliable parameter to evaluate the efficacy of nCRT.[24,25] However, the accuracy of mrTRG has been contested. Tumor re-

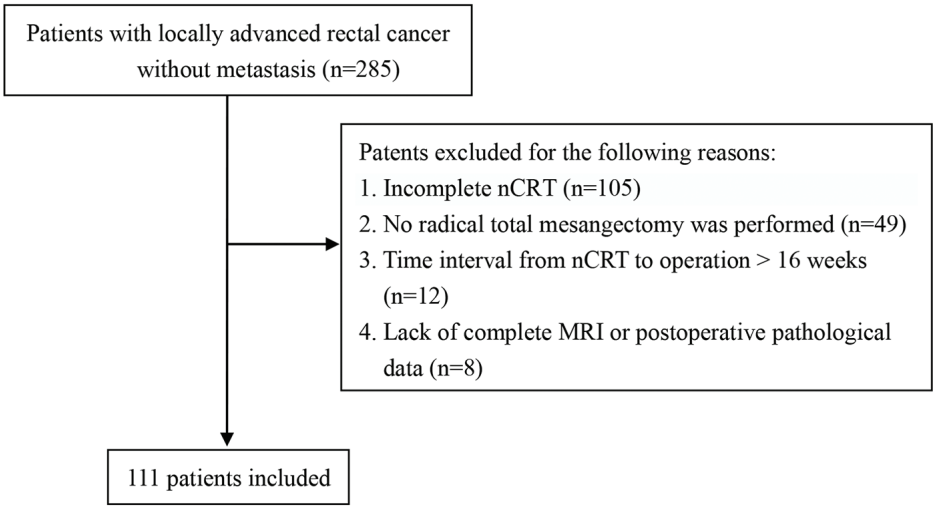


Figure 2. Patient selection flowchart. nCRT, neoadjuvant chemoradiotherapy; MRI, magnetic resonance imaging.

Table 2. Comparison of GRASP DCE-MR parameters before and after nCRT in pCR and non-pCR group

Group	Overall (n = 111)			pCR (n = 32)			Non-pCR (n = 79)		
	K_{trans} (/min)	K_{ep} (/min)	V_e	K_{trans} (/min)	K_{ep} (/min)	V_e	K_{trans} (/min)	K_{ep} (/min)	V_e
Pre-treatment (mean ± SD)	0.96 ± 0.40	0.69 ± 0.54	0.59 ± 0.36	1.30 ± 0.24	1.49 ± 0.39	0.57 ± 0.22	0.88 ± 0.34	1.37 ± 0.34	0.47 ± 0.23
Post-treatment (mean ± SD)	0.44 ± 0.25	0.55 ± 0.38	0.54 ± 0.26	0.55 ± 0.38	0.54 ± 0.26	0.35 ± 0.18	0.43 ± 0.25	0.75 ± 0.25	0.38 ± 0.27
t value	10.36	12.04	1.55	7.56	1.4	1.91	0.84	1.65	0.44
P value	<0.001	0.02	0.12	<0.001	0.08	0.06	0.40	0.10	0.67

nCRT, neoadjuvant chemoradiotherapy; K_{trans}, volume transfer constant; K_{ep}, rate constant; V_e, extracellular extravascular space volume fraction; pCR, complete response according to pathological outcome; GRASP, golden-angle radial sparse parallel; DCE-MR, dynamic contrast-enhanced magnetic resonance; SD, standard deviation.

Table 3. Relationship between mrTRG and pTRG after nCRT

mrTRG	pTRG					
	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	In total
Grade 1	5	0	0	0	0	5
Grade 2	16	1	0	1	0	18
Grade 3	11	28	28	1	0	68
Grade 4	0	0	2	15	0	17
Grade 5	0	1	0	2	0	3
In total	32	30	30	19	0	111

TRG, tumor regression grading; pTRG, TRG according to pathological outcomes; mrTRG, TRG according to magnetic resonance imaging evaluations; nCRT, neoadjuvant chemoradiotherapy.

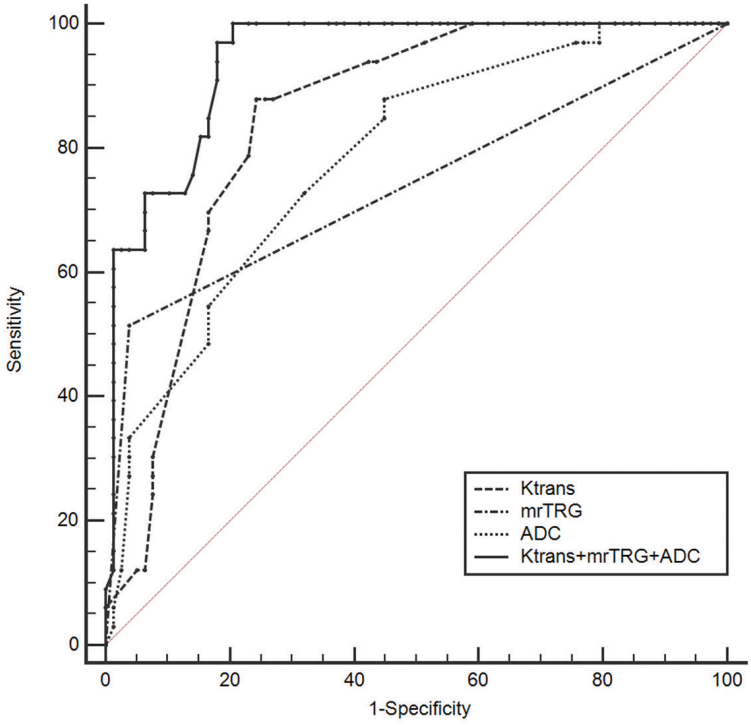


Figure 3. Ability of the magnetic resonance imaging-based tumor regression grading, apparent diffusion coefficient, and perfusion parameter in predicting pathological complete response after neoadjuvant chemoradiotherapy for rectal cancer.

gression after nCRT is a continuous process, with the peak usually occurring 8–11 weeks after the completion of treatment. This may explain the difference between mrTRG and pTRG.[26] In this study, MRI was performed 8 weeks after nCRT, and the median interval between MRI and radical surgery was 1 week. The sensitivity of mrTRG (71.5%) in our study was comparable with that (74.4%) reported by Sclafani et al.[27] In that study, the median interval between MRI and surgery was 2.7 weeks. There is still no evidence to standardize the selection of MR examination and operation time, and its influence on pCR prediction results is still unknown. Therefore, further studies are required to clarify this aspect.

However, this study defined mrTRG1–2 as cCR, and the sensitivity of mrTRG1–2 in the study by Bhoday et al.[28] was 66.7%. If mrTRG3 is included in the category of cCR, its sensitivity is greatly improved to 94%. This would further enhance the value of GRASP DCE-MRI. Whether mrTRG3 can be defined as a cCR also needs further study. Moreover, our results showed that mrTRG had good specificity (96.2%) for pCR, suggesting a higher diagnostic ability of mrTRG for patients with poor therapeutic efficacy (pTRG 2–4 grade).

The post-nCRT occurrence of necrosis and fibrosis in the tumor results in a decrease in the T2WI signal. However, there may still be a small number of tumor cells in the scarred and fibrotic tissue, which is not accurately distinguished by mrTRG. The addition of DWI can evaluate residual tumor activity to compensate for the deficiency of mrTRG in pCR prediction.[29] The ADC value is a quantitative index of the DWI sequence, and its increase is related to tumor necrosis. In one study, the average ADC value in patients achieving pCR was significantly higher than that in patients

Table 4. Ability of ADC, mrTRG, K_{trans}, and combined models to predict pCR after nCRT

	ADC	mrTRG	K_{trans}	ADC + mrTRG	ADC + K_{trans}	mrTRG + K_{trans}	ADC + mrTRG + K_{trans}
AUC (95% CI)	0.782 (0.694–0.85)	0.738 (0.646–0.817)	0.844 (0.763–0.906)	0.877 (0.801–0.932)	0.893 (0.82–0.944)	0.919 (0.851–0.962)	0.942 (0.881–0.977)
P values for Delong test*	0.023	0.015	0.030	0.039	0.049	0.099	
Accuracy	0.775	0.829	0.784	0.82	0.784	0.848	0.865
Specificity	0.962	0.962	0.833	0.936	0.872	0.897	0.923
Sensitivity	0.733	0.715	0.667	0.546	0.576	0.697	0.727
Positive predictive value	0.686	0.85	0.629	0.783	0.655	0.742	0.8
Negative predictive value	0.773	0.824	0.855	0.83	0.829	0.875	0.889

*The Delong test results were obtained from the comparison results of the combined model (three parameters included) with other parameters or models. ADC, apparent diffusion coefficient; TRG, tumor regression grading; pCR, complete response according to pathological outcome; nCRT, neoadjuvant chemoradiotherapy; K_{trans}, volume transfer constant; mrTRG according to magnetic resonance evaluations; AUC, area under the curve; CI, confidence interval.

who did not.[30] However, another study by Chandramohan et al.[30] found no significant association between the ADC value and pCR,[31] which may be due to the small sample size in their study (n = 22). The present study had a larger sample size, and we observed that the ADC value was significantly correlated with the outcome (B = 4.91, P < 0.001). Furthermore, ROC curve analysis revealed a moderate predictive ability of ADC (AUC: 0.78).

In this study, the RESOLVE sequence is affiliated with readout-segmented echo-planar imaging (readout-RS-EPI). Readout RS-EPI is characterized by small deformation and high resolution, which has little influence on the generated ADC value, thus reducing the impact of ADC value measurement bias.[32] Factors such as mucin pools in tumors, tiny residual tumor cell nests, low spatial resolution of DWI, radiation proctitis, and intestinal wall fibrosis may limit the predictive ability of ADC.[30] Nine patients in this study showed mucoid changes after treatment, increasing the average ADC value; thus, false positives may occur.

Angiogenesis is essential for tumor growth. In this study, the K_{trans} value of patients in the pCR group was higher than that in the non-pCR group, suggesting that chemotherapy drugs were more likely to enter the blood vessels with high permeability, and the blood vessels with high permeability had better oxygenation capacity and radiosensitivity.[33] The K_{trans} and K_{ep} values showed a significant decrease after treatment, which may be related to CRT-induced tumor necrosis and interstitial fibrosis. In this study, the results showed a high specificity but suboptimal sensitivity of K_{trans} for predicting pCR. Therefore, the use of K_{trans} alone may have low accuracy in predicting pCR. In addition, there was no significant difference in K_{trans}, K_{ep}, or V_e between the pCR group and non-pCR group after treatment, which is similar to the study by Kim et al.[34] but differs from the study by Gollub al.[16] The difference in results may be related to the non-standard combined cytotoxic and anti-angiogenic nCRT regimen adopted by Gollub et al.[16], whereas the standard nCRT regimen was adopted in our cohort. It may also be related to the GRASP DCE-MRI acquisition adopted in this study, in which 21 spokes were combined in each image, resulting in a time resolution of 3.45 s. This single reconstruction is well-balanced because it has a sufficiently high spatial resolution to compute perfusion maps and morphological assessments.[19] The combined model (mrTRG + ADC + K_{trans}) had the highest ability in predicting pCR (AUC: 0.942). However, the predictive ability was not superior to that of the combination model (mrTRG + K_{trans}; P = 0.099). This may be related to the small tumor parenchyma of pCR patients, which is difficult to measure. Moreover, the measurement error of the ADC value discussed above may also play a role. However, due to the lack of more detailed criteria and interobserver agreement, the current evaluation results based on imaging modalities were not consistent among centers and showed poor reproducibility. The pTRG may still play an irreplaceable role in the evaluation of nCRT treatment efficacy for patients with RC.

Some limitations of this study should be considered when interpreting the results. First, the retrospective nature of the study may have introduced an element of selection bias. Second, tumor regression and diffusion limitation due to tissue edema, fibrosis, and radiation enteritis after treatment all cause difficulties and biases in the measurement of mrTRG, ADC, and GRASP DCE-MRI parameters. Finally, the change in ADC value before and after nCRT was not analyzed in this study due to the lack of data. The change in ADC value may be a more accurate predictor of pCR.

In conclusion, the results of this study indicated that for patients with LARC who underwent nCRT, the K_{trans} values obtained from GRASP DCE-MRI, mrTRG, and ADC can be used as non-invasive indicators to evaluate the treatment efficacy of nCRT, and adding the K_{trans} value to mrTRG and ADC can lead to a better pCR predictive performance. Our findings may help inform individualized treatment planning.

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This study was supported by the Ningbo Science and Technology Public Welfare Project (grant no: 2022S26, grant no: 2022S133), the HwaMei Research Foundation of Ningbo No. 2 Hospital (grant no.: 2023HMKY50), and the Ningbo Clinical Research Center for Medical Imaging (grant no: 2022LYKEYB03).

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. [CrossRef]
2. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(3):145-164. [CrossRef]
3. Valentini V, Aristei C, Glimelius B, et al. Multidisciplinary rectal cancer management: 2nd European Rectal Cancer Consensus Conference (EURECA-CC2). *Radiother Oncol*. 2009;92(2):148-163. [CrossRef]
4. Martin ST, Heneghan HM, Winter DC. Systematic review and meta-analysis of outcomes following pathological complete response to neoadjuvant chemoradiotherapy for rectal cancer. *Br J Surg*. 2012;99(7):918-928. [CrossRef]
5. de Campos-Lobato LF, Stocchi L, da Luz Moreira A, et al. Pathologic complete response after neoadjuvant treatment for rectal cancer decreases distant recurrence and could eradicate local recurrence. *Ann Surg Oncol*. 2011;18(6):1590-1598. [CrossRef]
6. Bedrosian I, Rodriguez-Bigas MA, Feig B, et al. Predicting the node-negative mesorectum after preoperative chemoradiation for locally advanced rectal carcinoma. *J Gastrointest Surg*. 2004;8(1):62-63. [CrossRef]
7. Habr-Gama A, Perez RO, Nadalin W, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg*. 2004;240(4):717-718. [CrossRef]
8. Renehan AG, Malcomson L, Emsley R, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol*. 2016;17(2):174-183. [CrossRef]
9. Rullier E, Vendrely V, Asselineau J, et al. Organ preservation with chemoradiotherapy plus local excision for rectal cancer: 5-year results of the GRECCAR 2 randomised trial. *Lancet Gastroenterol Hepatol*. 2020;5(5):465-474. [CrossRef]
10. Patel UB, Taylor F, Blomqvist, et al. Magnetic resonance imaging-detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience. *J Clin Oncol*. 2011;29(28):3753-3760. [CrossRef]
11. Dresen RC, Beets GL, Rutten HJ, et al. Locally advanced rectal cancer: MR imaging for restaging after neoadjuvant radiation therapy with concomitant chemotherapy. Part I. Are we able to predict tumor confined to the rectal wall? *Radiology*. 2009;252(1):71-80. [CrossRef]
12. Pham TT, Liney GP, Wong K, Barton MB. Functional MRI for quantitative treatment response prediction in locally advanced rectal cancer. *Br J Radiol*. 2017;90(1072):20151078. [CrossRef]

13. Memon S, Lynch AC, Bressel M, Wise AG, Heriot AG. Systematic review and meta-analysis of the accuracy of MRI and endorectal ultrasound in the restaging and response assessment of rectal cancer following neoadjuvant therapy. *Colorectal Dis*. 2015;17(9):748-761. [CrossRef]
14. Just N. Improving tumour heterogeneity MRI assessment with histograms. *Br J Cancer*. 2014;111(12):2205-2213. [CrossRef]
15. Li Z, Huang H, Wang C, et al. DCE-MRI radiomics models predicting the expression of radioresistant-related factors of LRP-1 and survivin in locally advanced rectal cancer. *Front Oncol*. 2022;12881341. [CrossRef]
16. Gollub MJ, Gultekin DH, Akin O, et al. Dynamic contrast enhanced-MRI for the detection of pathological complete response to neoadjuvant chemotherapy for locally advanced rectal cancer. *Eur Radiol*. 2012;22(4):821-831. [CrossRef]
17. Kim SH, Lee JM, Gupta SN, Han JK, Choi BI. Dynamic contrast-enhanced MRI to evaluate the therapeutic response to neoadjuvant chemoradiation therapy in locally advanced rectal cancer. *J Magn Reson Imaging*. 2014;40(3):730-737. [CrossRef]
18. Fusco R, Granata V, Sansone M, et al. Validation of the standardized index of shape tool to analyze DCE-MRI data in the assessment of neo-adjuvant therapy in locally advanced rectal cancer. *Radiol Med*. 2021;126(8):1044-1054. [CrossRef]
19. Attenberger UI, Liu J, Riffel P, et al. Quantitative perfusion analysis of the rectum using golden-angle radial sparse parallel magnetic resonance imaging: initial experience and comparison to time-resolved angiography with interleaved stochastic trajectories. I*nvest Radiol*. 2017;52(12):715-724. [CrossRef]
20. Riffel P, Zoellner FG, Budjan J, et al. "One-Stop Shop": free-breathing dynamic contrast-enhanced magnetic resonance imaging of the kidney using iterative reconstruction and continuous golden-angle radial sampling. *Invest Radiol*. 2016;51(11):714-719. [CrossRef]
21. Feng L, Grimm R, Block KT, et al. Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magn Reson Med*. 2014;72(3):707-717. [CrossRef]
22. Rosenkrantz AB, Geppert C, Grimm R, et al. Dynamic contrast-enhanced MRI of the prostate with high spatiotemporal resolution using compressed sensing, parallel imaging, and continuous golden-angle radial sampling: preliminary experience. *J Magn Reson Imaging*. 2015;41(5):1365-1373. [CrossRef]
23. Mandard AM, Dalibard F, Mandard JC, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer*. 1994;73(11):2680-2686. [CrossRef]
24. Tan Z, Cheng L, Xie L, et al. Comparison of the diagnostic performance of changes in signal intensity and volume from multiparametric MRI for assessing response of rectal cancer to neoadjuvant chemoradiotherapy. *Asia Pac J Clin Oncol*. 2023;19(3):327-336. [CrossRef]
25. Boubaddi M, Fleming C, Vendrely V, et al. Feasibility study of a response surveillance program in locally advanced mid and low rectal cancer to increase organ preservation. *Eur J Surg Oncol*. 2023;49(1):237-243. [CrossRef]
26. Sun Z, Adam MA, Kim J, Shenoi M, Migaly J, Mantyh CR. Optimal timing to surgery after neoadjuvant chemoradiotherapy for locally advanced rectal cancer. Optimal timing to surgery after neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *J Am Coll Surg*. 2016;222(4):367-374. [CrossRef]
27. Sclafani F, Brown G, Cunningham D, et al. Comparison between MRI and pathology in the assessment of tumour regression grade in rectal cancer. *Br J Cancer*. 2017;117(10):1478-1485. [CrossRef]
28. Bhoday J, Smith F, Siddiqui MR, et al. Magnetic resonance tumor regression grade and residual mucosal abnormality as predictors for pathological complete response in rectal cancer postneoadjuvant chemoradiotherapy. *Dis Colon Rectum*. 2016;59(10):925-933. [CrossRef]
29. Horvat N, Veeraraghavan H, Khan M, et al. MR Imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology*. 2018;287(3):833-843. [CrossRef]
30. Chandramohan A, Siddiqi UM, Mittal R, et al. Diffusion weighted imaging improves diagnostic ability of MRI for determining complete response to neoadjuvant therapy in locally advanced rectal cancer. *Eur J Radiol Open*. 2020;7100223. [CrossRef]
31. Jacobs L, Intven M, van Lelyveld N, et al. Diffusion-weighted MRI for early prediction of treatment response on preoperative chemoradiotherapy for patients with locally advanced rectal cancer: a feasibility study. *Ann Surg*. 2016;263(3):522-528. [CrossRef]
32. Tang C, Lin MB, Xu JL, et al. Are ADC values of readout-segmented echo-planar diffusion-weighted imaging (RESOLVE) correlated with pathological prognostic factors in rectal adenocarcinoma? *World J Surg Oncol*. 2018;16(1):138. [CrossRef]
33. Cooper RA, Carrington BM, Loncaster JA, et al. Tumour oxygenation levels correlate with dynamic contrast-enhanced magnetic resonance imaging parameters in carcinoma of the cervix. *Radiother Oncol*. 2000;57(1):53-59. [CrossRef]
34. Kim SH, Lee JM, Hong SH, et al. Locally advanced rectal cancer: added value of diffusion-weighted MR imaging in the evaluation of tumor response to neoadjuvant chemo- and radiation therapy. *Radiology*. 2009;253(1):116-125. [CrossRef]

Supplementary Table 1. Parameters for MRI scans

Parameters	T2 weighted imaging	Resolve DWI	Grasp DCE-MRI
Plane	Axial position (perpendicular to the long axis of the tumor)	Axial position	Axial position
Repeat time/echo time (ms)	6770/104	5800/78	4.09/1.95
Number of layer	35	35	24
Layer thickness (mm)	3	3	3
Layer distance (mm)	0.6	0.6	0.6
Field of view (mm)	220 x 220	220 x 220	240 x 240
Matrix of scanning	384 x 384	114 x 114	256 x 256
Pixel	0.33 x 0.33 x 3.0	1.0 x 1.0 x 3.0	0.9 x 0.9 x 3.0
Fat inhibition	No	Yes	Yes
B value (s)	-	0,50,1000	-
GRAPPA acceleration factor	1	1	1
Acquisition time	2 min 15s	3 min 13s	6 min 06s

MRI, magnetic resonance imaging; DWI, diffusion weighted imaging; Grasp DCE-MRI, Golden-angle radial sparse parallel dynamic contrast-enhanced magnetic resonance imaging.

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2023.232386

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE

Cystic renal mass screening: machine-learning-based radiomics on unenhanced computed tomography

Lesheng Huang[1*]
Yongsong Ye[2*]
Jun Chen[1,2]
Wenhui Feng[3]
Se Peng[4]
Xiaohua Du[5]
Xiaodan Li[6]
Zhixuan Song[7]
Tianzhu Liu[1]

[1]Guangdong Provincial Hospital of Chinese Medicine, Department of Radiology, Zhuhai, China

[2]Guangdong Provincial Hospital of Chinese Medicine, Department of Radiology, Guangzhou, China

[3]Zhuhai People's Hospital, Department of Radiology, Zhuhai, China

[4]Guangdong Provincial Hospital of Chinese Medicine, Department of Laboratory Medicine, Zhuhai, China

[5]Guangdong Provincial Hospital of Chinese Medicine, Department of Pathology, Guangzhou, China

[6]Guangdong Provincial Hospital of Chinese Medicine, Department of Gynaecology, Zhuhai, China

[7]Philips Healthcare, Clinical and Technical Support, Guangzhou, China

PURPOSE

The present study compares the diagnostic performance of unenhanced computed tomography (CT) radiomics-based machine learning (ML) classifiers and a radiologist in cystic renal masses (CRMs).

METHODS

Patients with pathologically diagnosed CRMs from two hospitals were enrolled in the study. Unenhanced CT radiomic features were extracted for ML modeling in the training set (Guangzhou; 162 CRMs, 85 malignant). Total tumor segmentation was performed by two radiologists. Features with intraclass correlation coefficients of >0.75 were screened using univariate analysis, least absolute shrinkage and selection operator, and bidirectional elimination to construct random forest (RF), decision tree (DT), and k-nearest neighbor (KNN) models. External validation was performed in the Zhuhai set (45 CRMs, 30 malignant). All images were assessed by a radiologist. The ML models were evaluated using calibration curves, decision curves, and receiver operating characteristic (ROC) curves.

RESULTS

Of the 207 patients (102 women; 59.1 ± 11.5 years), 92 (41 women; 58.0 ± 13.7 years) had benign CRMs, and 115 (61 women; 59.8 ± 11.4 years) had malignant CRMs. The accuracy, sensitivity, and specificity of the radiologist's diagnoses were 85.5%, 84.2%, and 91.1%, respectively [area under the (ROC) curve (AUC), 0.87]. The ML classifiers showed similar sensitivity (94.2%–100%), specificity (94.7%–100%), and accuracy (94.3%–100%) in the training set. In the validation set, KNN showed better sensitivity, accuracy, and AUC than DT and RF but weaker specificity. Calibration and decision curves showed excellent and good results in the training and validation set, respectively.

CONCLUSION

Unenhanced CT radiomics-based ML classifiers, especially KNN, may aid in screening CRMs.

KEYWORDS

Cystic renal mass, diagnosis, radiomics, machine-learning

Corresponding author: Tianzhu Liu

E-mail: hadesfantasy012@21cn.com

Received 27 Jun 2023; revision requested 30 July 2023; last revision received 23 November 2023; accepted 30 November 2023.

Epub: 02.01.2024

Publication date: 08.07.2024

DOI: 10.4274/dir.2023.232386

Cystic renal masses (CRMs) are defined as renal lesions with <25% enhancing tissue, and they are often identified incidentally on abdominal computed tomography (CT) scans.[1] The majority of CRMs are benign, but a minority are diagnosed as renal cell carcinoma or other rare malignant renal tumors.[2,3] The proposed 2019 version of the Bosniak classification stratifies CRMs according to their risk of malignancy;[1] however, the diagnostic accuracy of this classification is low when applied to unenhanced CT scans because of the poor ability to visually judge gray-scale features with the naked eye.[4] Unfortunately, plain CT scans are commonly used in many situations, such as renal insufficiency, night-time emergencies, and especially annual CT examinations. Thus, a technique that enables the use of unenhanced CT scans for the accurate stratification of CRMs could assist radiologists and surgeons in screening to differentiate between malignant and benign CRMs.

You may cite this article as: Huang L, Ye Y, Chen J, et al. Cystic renal mass screening: machine-learning-based radiomics on unenhanced computed tomography. *Diagn Interv Radiol*. 2024;30(4):236-247.

Radiomic features have the potential to aid in the classification of lesion characteristics.[5] This quantitative approach to analyzing microscopic differences represents an emerging method in the pursuit of better understanding and identifying tumor phenotypes, although further research is required to establish specific feature-to-property correlations and standardize methodologies. Multiple supervised machine learning (ML) classifiers, such as the support vector machine, random forest (RF), decision tree (DT), and k-nearest neighbor (KNN), can be used to build diagnostic models based on radiomic features. Numerous studies on renal cell carcinoma have confirmed the excellent diagnostic efficacy of radiomics-based ML methods.[6-10] Recently, several ML algorithms were applied to classify CRMs into benign or malignant masses by using CT-based radiomic features.[11-13] While these studies are important and indispensable, further research on CRMs and ML is required for a number of reasons. First, previous algorithms were trained with arterial-phase (AP) and venous-phase (VP) scans; unenhanced CT features were either not used at all or only used as a supplementary part during model construction.[11-13] Second, some studies[11] lacked external data validation or validation in other centers to verify the diagnostic effectiveness of the models constructed. Finally, the above studies did not compare the diagnostic effectiveness of the ML-based models with that of manual diagnosis by experienced radiologists. To overcome the above shortcomings, the present authors aimed to build diagnostic ML models of CRMs based on unenhanced CT radiomic features; these models were verified with external data from a different center, and the diagnostic efficiency of the ML classifiers was compared with that of manual diagnosis.

Main points

- Several machine learning (ML) algorithms have been used to classify cystic renal masses (CRMs) into benign or malignant masses using computed tomography (CT)-based radiomic features, but previous algorithms were trained with arterial-phase and venous-phase scans.
- The present study showed that ML algorithms with unenhanced-CT radiomics features also presented acceptable diagnostic efficiency. The k-nearest neighbor (KNN) model presented satisfactory sensitivity and accuracy and was similar to the radiologist's performance, and the decision tree and random forest models presented satisfactory specificity.
- Due to its satisfactory sensitivity, the KNN model could be a potential screening method for patients with CRMs.

Methods

Ethics approval and case selection

This retrospective study was approved by the Medical Ethics Committee of Guangdong Provincial Hospital of Chinese Medicine (no: ZE2023-090-01), and the requirement for written informed consent was waived. Patients with CRMs who were treated at Guangdong Provincial Hospital of Chinese Medicine in either Guangzhou or Zhuhai (Center 1: Guangzhou and Center 2: Zhuhai) between January 2018 and February 2022 were eligible for this study. The inclusion criteria were as follows: (a) unenhanced and enhanced CT scans, including AP and VP images, were completed for the stratification of CRMs with the Bosniak classification; (b) complete clinical data were available, including age, sex, location of the lesions, intact operation and/or biopsy records, and histopathological results (obtained from the pathological retrieval systems of the two centers); and (c) good-quality CT images were stored in the Picture Archiving and Communications System. The exclusion criteria were (a) low-quality or incomplete CT data and (b) masses belonging to category II or lower according to the Bosniak classification. After the application of the above selection criteria, a total of 207 cases (92 benign and 115 malignant CRMs) were included in the study. The cases from Center 1 (77 benign and 85 malignant CRMs) were allocated to the training set, while the cases from Center 2 (15 benign and 30 malignant CRMs) were assigned to the validation set for external validation. The workflow of the ML approach is shown in Figure 1, and a flow chart of the case selection is shown in Figure 2.

Computed tomography examinations

All patients underwent unenhanced and dual-phase contrast-enhanced CT. The CT scanning was performed using three CT scanners: Definition Flash (Siemens, Forchheim, Germany) and IQon Spectral (Philips Healthcare, Amsterdam, Netherlands) in Center 1, and Aquilion One 750 W (Canon, Tokyo, Japan) in Center 2. Images obtained in three phases (unenhanced, AP, and VP) were used for the Bosniak classification, and unenhanced images were used for radiomic-feature extraction. The following scanning parameters were applied for all images: tube voltage = 120 kV; tube current = 250 mA; section interval = 5 mm; section thickness = 5 mm; and matrix size = 512 × 512 mm. After conventional unenhanced scanning, 100–120 mL of the contrast medium, iopromide (Ultravist 370, Bayer Schering Pharma, Germany) was injected into the median cubital vein via a pump injector (MEDRAD Stellant CT, Ulrich Medical, Ulm, Germany) at a flow rate of 3–4 mL/s. The AP was scanned using an aortic monitoring trigger, and the VP was scanned after approximately 60 s of delay after the contrast medium injection.

A single radiologist (J.C.) with 18 years of experience analyzed all the CT images to (a) check that all cases met the standard of <25% enhancing tissue, (b) confirm the Bosniak class (version 2019), and (c) measure the size of the CRMs.

Mass segmentation and radiomic-feature extraction

The open-source software platform, 3D Slicer (version 5.2.1, www.slicer.org), was applied for mass segmentation and calculation of radiomic features. Masses were delineated on the original CT images using 3D Slicer. Segmentation of whole masses was performed by associate chief radiologists (T.L. and L.H.) with more than 15 years of experience in abdominal radiography; to outline the shape and edges of the masses more accurately, the radiologists were allowed to observe the enhanced CT images. In each case, the entire CRM was carefully and manually segmented to avoid beyond-boundary or insufficient filling. Following tumor segmentation, 855 radiomic features were extracted using the "PyRadiomics" package with 3D Slicer. The extracted features were classified into seven categories as follows: first-order features, two-dimensional features, gray-level co-occurrence matrix, gray-level dependence matrix, gray-level size-zone matrix, gray-level run-length matrix, and neighboring gray tone difference matrix. Additionally, the following 14 filters were applied to the original images: exponential, gradient, square, square root, logarithm, lbp2D, wavelet-HLH, wavelet-HLL, wavelet-LHL, wavelet-LLL, wavelet-LHH, wavelet-LLH, wavelet-HHL, and wavelet-HHH. The images thus derived were analyzed for each patient. All classes of features were computed on both the original images and the derived images.

To ensure the stability of the radiomic features extracted from the CT images, the segmentation and feature-extraction process was repeated in 80 randomly selected patients with CRMs from the training set.

Intraclass correlation coefficients (ICCs) were used to evaluate consistency across the radiomic features; features with ICCs >0.75 were considered stable and were included in this analysis.

After meeting the standard of consistency, the features were further selected to avoid overfitting. The least absolute shrinkage and selection operator (LASSO) method was applied to select the most suitable radiomic features to develop a radiomic signature with the "glmnet" package. First, 10-fold cross-validation was performed to obtain the optimal parameter λ[14] by 1,000 iterations. Second, the LASSO method based on the optimal parameter λ was used to calculate the coefficient of each feature, and features with non-zero coefficients were selected.[14] Finally, bidirectional elimination was used to further filter the radiomic features selected using the LASSO method;[15] the "mass" package in the R software (version 4.2.2) was used for bidirectional elimination (Figure 1).

Statistical analysis

The χ^2 test was used to compare categorical data, and the independent-samples t-test was used to compare inter-group differences in clinical data. Statistical analysis was conducted using SPSS (version 26.0, IBM, Armonk, NY, USA) and R (version 4.2.2). A two-sided *P* value of <0.05 was considered statistically significant.

Machine learning algorithms

The radiomic features selected using the above steps were standardized to a mean of 0 and an standard deviation of 1 before ML algorithm construction. Supervised learning was achieved using three supervised learning classifiers: RF, DT, and KNN. A 10-fold cross-validation strategy was applied to assess the performance of the classification models. Under this strategy, the data were divided into 10 parts; nine parts were used for training in turn, and the remaining part was used to estimate the efficacy of the models. During the process of fine-tuning the models, the grid search method was employed to select the best combination of hyperparameter values.

Patients from Center 1 (77 benign and 85 malignant CRMs) were allocated to the training set, and patients from Center 2 (15 benign and 30 malignant CRMs) were allocated to the validation set for external validation to estimate the performance of the models. The discriminative performance of different models was quantified using area under the [receiver operating characteristic (ROC)] curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The AUCs of the ML models were also compared using the Delong test. The SHapley Additive exPlanations (SHAP) values, which indicate the importance of radiomic features, were derived for the RF and DT models (SHAP values are not suitable for the KNN model).

The ML algorithm creation was performed using the "Caret" package. Calibration curves were plotted using the "rms" package. Decision curve analysis (DCA) was performed using the "rmda" and "ggDCA" packages. The "pROC" package was used for ROC curve analysis. The "reportROC" package was used to present the sensitivities, specificities, accuracies, PPVs, NPVs, and 95% confidence intervals (CIs) of the AUCs obtained using ROC curve analysis.

Manual diagnosis by radiologist

The study authors also assessed the diagnostic performance of an attending radiologist (W.F.) with more than 7 years of experience in radiology diagnoses. This radiologist used the open-source DICOM viewer MicroDicom (https://www.microdicom.com/) for image evaluation. The radiologist was from a hospital not involved in this study and was blinded to the patient demographic and clinical characteristics. The radiologist independently reviewed the unenhanced CT images and established a diagnosis according to the Bosniak classification, based on the morphological features of the lesions.

Quality evaluation of research

To evaluate the quality of research, the study authors used the CheckList for EvaluAtion of Radiomics research (CLEAR)[16] and the radiomics quality score (RQS).[17] The datasets and source code generated and/or analyzed during the current study are available on GitHub (https://github.com/elliiesong/CRM-screening-with-machine-learning-unenhanced-CT).

Results

Patient characteristics

This study included 207 patients (105 men, 102 women; mean age: 59.1 ± 11.5 years) with CRMs. Of these, 92 patients (51 men, 41 women; mean age: 58.0 ± 13.7 years) had benign CRMs, and 115 patients (54 men, 61 women; mean age: 59.8 ± 11.4 years) had malignant CRMs (Figure 2). There were no significant differences in age, sex, or mass location or size between patients with benign or malignant CRMs (Table 1). All benign CRMs were simple kidney cysts, except for one case of angiomyolipoma. All malignant CRMs were clear cell carcinoma, except for one case of mixed epithelial and stromal tumor of the kidney.

Radiomic-feature selection

Following univariate analyses, 216 radiomic features were extracted from the unenhanced CT images, and LASSO and 10-fold cross-validation were used to screen and select radiomic features. Finally, the following four features screened out from unenhanced CT images were selected: Original_glcm_Maximum_Probability, Wavelet.LHH_firstorder_Median, Wavelet.LLL_firstorder_90Percentile, and Wavelet.LLL_firstorder_Median.

Diagnostic performance of machine learning algorithms

Four features (Original_glcm_Maximum_Probability, Wavelet.LHH_firstorder_Median, Wavelet.LLL_firstorder_90Percentile, and Wavelet.LLL_firstorder_Median) were used to construct the ML models. The diagnostic efficiencies of the ML classifiers are summarized in Table 2 and Figure 3. In the training set, the accuracy, specificity, sensitivity, and AUC of RF, DT, and KNN (*k*-value: 4) were satisfactory and similar to each other. A confusion matrix was prepared from the verification set, and the accuracy of RF, DT, and KNN in this set was 77.3% (95% CI: 76.5%–78.1%), 79.5% (95% CI: 78.8%–80.3%), and 84.1% (95% CI: 83.5%–84.7%), respectively. The specificity of KNN (73.3%, 95% CI: 51.0%–95.7%) was significantly weaker than that of RF (80.6%, 95% CI: 60.7%–100%) and DT (80.0%, 95% CI: 59.8%–100%). The sensitivity of KNN (89.7%, 95% CI: 78.6%–100%) was significantly better than that of RF (65.5%, 95% CI: 48.2%–82.8%) and DT (79.3%, 95% CI: 64.6%–94.1%). The AUC of KNN (0.86, 95% CI: 0.74–0.98) was slightly better than that of RF (0.77, 95% CI: 0.61–0.92) and DT (0.80, 95% CI: 0.67–0.93). None of the ML classifiers significantly differed from manual diagnosis (Supplemantary Table S1). The results of the Delong test showed that there was no statistical difference between the ML classifiers (KNN and RF: $P = 0.205$; KNN and DT: $P = 0.061$; RF and DT: $P = 0.586$). The SHAP values of DT and RF (Supplemantary Figure S1) showed that the feature Wavelet.LLL_firstorder_Median held absolute weight in the two models, especially in the DT model.

Calibration curve analysis and DCA of the ML classifiers were performed in the training and validation sets (Figure 3c-f). The calibration curves were excellent and close to the ideal line in the training set but showed some degree of deviation from the ideal line in the validation set. The KNN and DT lines were above the ideal line but became close to and intersected the ideal line in the latter half, and the RF line was below the ideal line in the first half and above it in the second half. The DCA showed excellent results in the training set and revealed a greater net benefit than all positive and negative lines when the risk threshold was more than approximately 0.3 in the validation set; the KNN, DT, and RF lines were similar.

Efficiency of manual diagnosis

The manual diagnosis results are summarized in Table 2 and Figure 3b. The radiologist's diagnoses using unenhanced CT images presented an accuracy, sensitivity, and specificity of 85.5%, 84.2%, and 91.1%, respectively, with an AUC of 0.866.

Radiomics quality score

The quality of this study was evaluated using CLEAR[16] and RQS.[17] The results of the CLEAR evaluation were 43/9/6 (Yes/No/n/a, total: 58), and the RQS was 47.22% (17/36). The details of the RQS and CLEAR are summarized in Supplementary Tables S2, S3.

Discussion

In this bicentric study, the authors attempted to create multiple ML classifiers to distinguish between benign and malignant CRMs on unenhanced CT images. The results indicated that the accuracy and AUC of the ML classifiers were satisfactory (accuracy: 77.3%–84.1%; AUC: 0.77–0.86) and similar to that of the radiologist's diagnoses. The KNN

Table 1. Clinical and pathological characteristics of the included CRM patients

	Training set (n = 162)			Validation set (n = 45)		
Characteristic	Benign (n = 77)	Malignant (n = 85)	*P* value	Benign (n = 15)	Malignant (n = 30)	*P* value
Age (years), mean ± SD	57.4 ± 9.8	60.3 ± 12.6	0.746	58.2 ± 10.6	61.6 ± 14.0	0.633
Gender			0.281			0.831
Male	41 (19.8%)	37 (17.87%)		9 (4.35%)	17 (8.21%)	
Female	36 (12.56%)	48 (23.18%)		6 (2.90%)	13 (6.28%)	
Mass size (cm), mean ± SD	4.80 ± 1.32	5.06 ± 1.85	0.790	4.77 ± 1.69	6.10 ± 1.22	0.509
Location						
Right kidney	41 (19.80%)	53 (25.60%)	0.311	6 (2.90%)	13 (6.28%)	0.831
Left kidney	36 (17.39%)	32 (15.46%)		9 (4.35%)	17 (8.21%)	
Histological subtype			<0.0001			<0.0001
Simple kidney cyst	77 (37.19%)	0 (0%)		14 (6.76%)	0 (0%)	
Clear cell carcinoma	0 (0%)	84 (40.57%)		0 (0%)	30 (14.50%)	
Other	0 (0%)	1 (0.48%)		1 (0.48%)	0 (0%)	
Bosniak classification			<0.0001			<0.0001
IIF	63 (30.43%)	16 (7.73%)		12 (5.80%)	3 (1.45%)	
III	14 (6.76%)	21 (10.14%)		3 (1.45%)	10 (4.83%)	
IV	0 (0%)	48 (23.18%)		0 (0%)	17 (8.21%)	

Data are expressed as mean ± SD, median (interquartile range), or frequency (constituent ratio). CRM, cystic renal mass; SD, standard deviation.

Table 2. Diagnostic efficiency of three computed tomography radiomic feature–based machine learning algorithms in differentiating benign from malignant cystic renal masses (n = 207) in the training and validation sets

Machine learning algorithm/ manual analysis	Sensitivity (%), (95% CI)	Specificity (%), (95% CI)	Accuracy (%), (95% CI)	PPV (%), (95% CI)	NPV (%), (95% CI)	AUC (95% CI)
Training set						
RF	100 (99.2–100)	100 (98–100)	100 (98.9–100)	100 (99.4–100)	100 (99.4–100)	1.00 (0.98–1.00)
DT	94.2 (89.2–99.1)	94.7 (89.7–99.8)	94.4 (94.4–94.5)	95.3 (90.8–99.8)	93.5 (88.0–99.0)	0.95 (0.91–0.98)
KNN	94.2 (89.2–99.1)	95.0 (90.1–100)	94.3 (94.0–94.7)	95.3 (91.0–99.8)	93.5 (88.0–99.0)	0.97 (0.95–0.99)
Validation set						
RF	65.5 (48.2–82.8)	80.6 (60.7–100)	77.3 (76.5–78.1)	87.4 (73.2–100)	57.2 (36.9–78.3)	0.77 (0.61–0.92)
DT	79.3 (64.6–94.1)	80.0 (59.8–100)	79.5 (78.8–80.3)	88.5 (76.2–100)	66.7 (44.9–88.4)	0.80 (0.67–0.93)
KNN	89.7 (78.6–100)	73.3 (51.0–95.7)	84.1 (83.5–84.7)	86.7 (74.5–98.8)	78.6 (57.1–100)	0.86 (0.74–0.98)
Radiologist	84.2	91.1	85.5	90.9	83.6	0.87

CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; RF, random forest; DT, decision tree; KNN, k-nearest neighbor.

presented the highest sensitivity and accuracy, and the DT and RF presented the highest specificity.

The Bosniak classification is the standard stratification method used to estimate the risk of malignancy in CRMs; however, this classification does have some limitations. First, ambiguous definitions, such as "cystic," "solid," "walls," and "septa," are difficult to quantify.[18-23] Second, the Bosniak classification is limited by considerable variability between radiologists,[24] especially for Bosniak classes II, IIF, and III, for which absolute disagreement ranges from 6% to 75%.[25] Finally, most CRMs are found incidentally, owing to which the scanning procedure is not planned for imaging the entire mass and may not include enhanced CT scans; hence, the Bosniak classification often cannot be applied.[26]

Compared with visual analysis, ML classifiers of radiomic features could more comprehensively and objectively reflect the phenotypic properties of masses, which may represent the underlying microscopic pathological changes and heterogeneity of the disease. The ML classifiers have potential benefits in screening CRMs: first, they are objective and not subject to reader interpretation, although segmentation by readers can still be needed; however, automatic segmentation has been used in some situations. Second, unlike the Bosniak classification, which depends on enhanced scanning, the ML classifiers can be applied to single-phase CT scans and may obviate additional radiological examinations.

Other diagnostic models based on radiomic features have also been studied. A decision algorithm used by Dana et al.[12] was built by combining consensus radiological readings of Bosniak categories and radiomics-based risks; the results showed excellent diagnostic performance (AUC: 0.96). He et al.[13] applied deep learning and a radiomic feature-based blending ensemble classifier to predict the malignancy risk of CRMs and obtained satisfactory diagnostic performance (AUC: 0.934). However, both these models were based on CT images obtained in the three phases or in the arterial phase. The following inferences can be drawn from the above findings: first, radiomic features play a valuable role in the diagnosis of CRMs; second, unenhanced CT scan-based radiomic features of CRMs were underappreciated in previous studies. Unlike other studies, the present study focused on unenhanced CT scan-based radiomic features and presented acceptable diagnostic efficiency (RF: AUC = 0.77; DT: AUC = 0.80; KNN: AUC = 0.86) in the absence of other CT phases.

Building on prior studies,[6,8,11] this study applied unenhanced CT-based ML classifiers independent of the Bosniak classification and compared their performance in the diagnosis of pathologically proven masses. Each of the three ML classifiers (RF, DT, and KNN) showed a similar high accuracy in distinguishing between benign and malignant CRMs. Although prior work has demonstrated the ability of ML classifiers to differentiate between benign and malignant

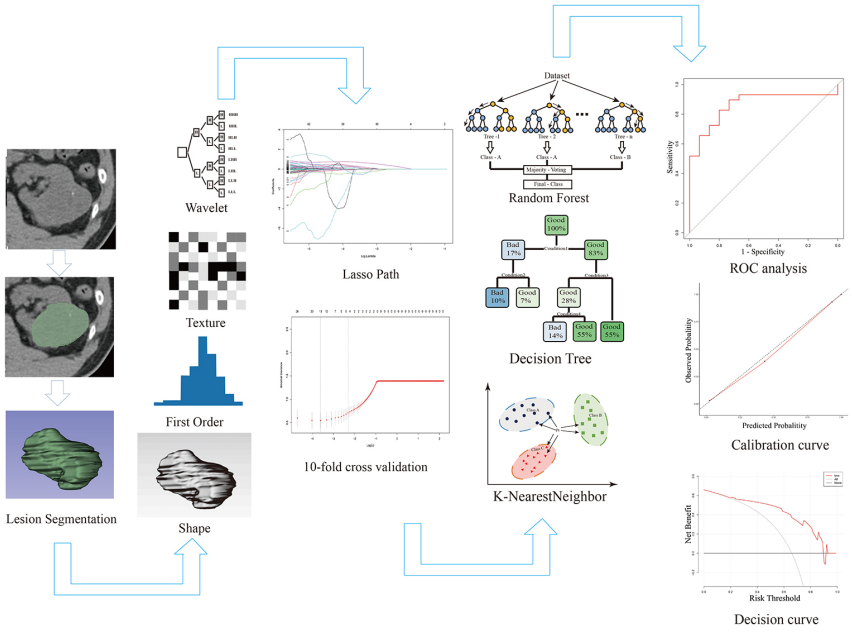


Figure 1. Workflow of the machine learning approach.

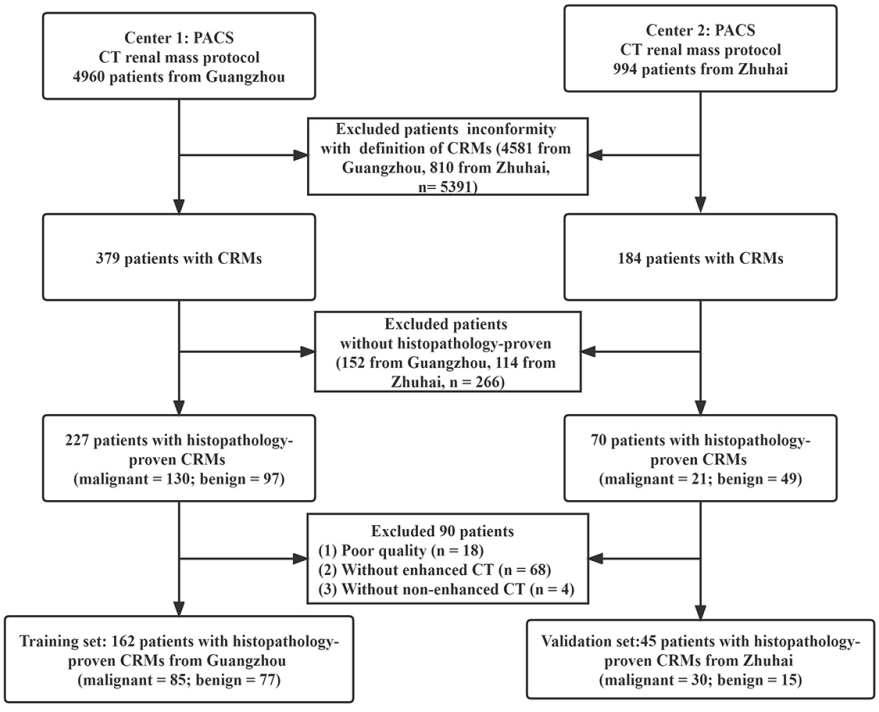


Figure 2. Flow chart of patient inclusion and exclusion criteria. PACS, Picture Archiving and Communications System; CT, computed tomography; CRMs, cystic renal masses.

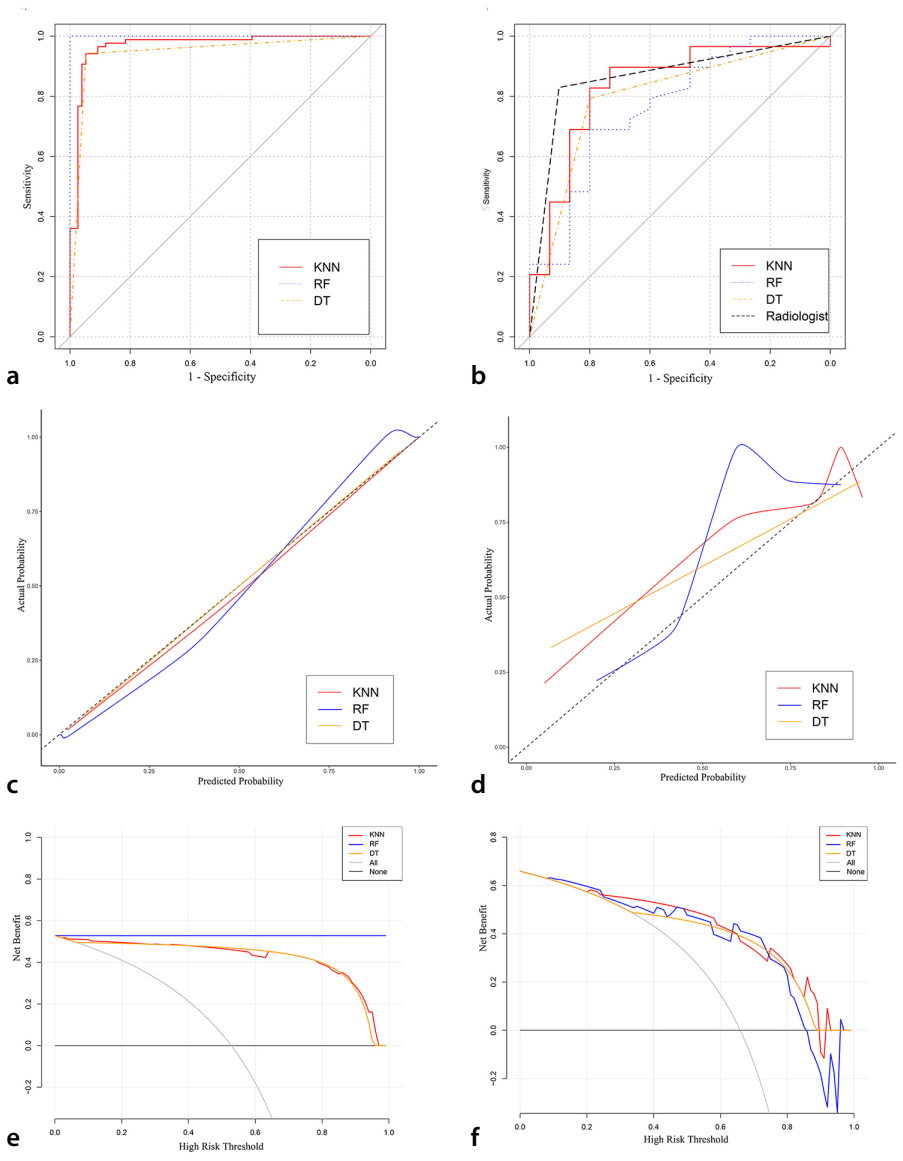


Figure 3. (a, b) Receiver operating characteristic curves of the machine learning (ML) classifiers for cystic renal masses (CRMs) in the training **(a)** and validation **(b)** sets. **(c, d)** Calibration curves of the ML classifiers for CRM prediction in the training **(c)** and validation **(d)** sets. **(e, f)** Decision curve analysis of the ML classifiers for CRMs in the training **(e)** and validation **(f)** sets.

solid or CRMs,[11] to the best of the authors' knowledge, this study is the first to develop ML classifiers to distinguish between benign and malignant CRMs based on unenhanced CT images, as well as compare the diagnostic effectiveness of ML classifiers with that of manual diagnosis by a radiologist.

The ML classifiers showed acceptable-to-high sensitivity (65.5%–89.7%) and specificity (73.3%–80.6%) in the validation set in this study. The authors considered satisfactory sensitivity of single-phase radiomics models, especially unenhanced models, important for clinical application because most CRMs are found incidentally, and an unenhanced model could provide a preliminary diagnosis to help clinicians make the next decision. In this study, KNN presented the highest sensitivity among the ML classifiers, which was better than that of manual diagnosis (KNN vs. radiologist: 89.7% and 84.2%, respectively). This indicates that KNN could screen malignant CRMs at a greater probability. Compared with the increased detection of suspected malignant masses that need further examination, such as enhanced CT or MR scanning, the misdiagnosis of malignant CRMs is a greater disadvantage and may cause patients to miss the optimal time window for treatment. An unenhanced CT-based KNN classifier could be a valuable diagnostic method for CRMs in clinical and radiological practice. Compared with the linear pattern of the DT line and the sigmoid pattern of the RF line, the KNN line in the calibration curve analysis was close to the ideal line in the second half. This may mean that the KNN classifier exhibited more adaptability in the positive diagnosis of CRMs. On the other hand, the composition and importance of features are also noteworthy points. In this study, three of the four radiomic features used for model predictions were computed with wavelet filters. Thus, radiomic features derived using wavelet filters dominated the models and may have had a significant impact on the predictive performance of the models.[27]

The drawbacks of ML classifiers need to be acknowledged. The ML classifiers used in this study are supervised methods that require a reader to segment the masses and extract the features; thus, the performance of the models may be affected by the segmentation process, unless an automatic segmentation is applied.

There are several limitations to this study. First, although this study is a bicentric study, the two hospitals share a set of CT scanning and image-reconstruction standards, although the CT scanning equipment is different; hence, the images can still have relatively high consistency. Verification with scans from other hospitals with different scanning parameters is required to confirm the diagnostic efficiency of the ML models from this study. Second, the composition of the validation set was not balanced (15 benign and 30 malignant CRMs), which may have led to potential risks and affected the validation results. Third, KNN is a simple classifier and has the potential risk of overfitting; hence, even though the diagnostic efficiency of the models was satisfactory in both the training and validation sets, more data are needed for verification. Fourth, the majority of patients were pathologically diagnosed with renal cysts and clear cell carcinomas; the diagnostic performance of the models on other pathological types of CRMs, such as papillary and tubular renal cell carcinomas, remains unconfirmed. To truly understand the models' capabilities across all pathological types, further comprehensive research is essential. Fifth, the radiologists were allowed to observe the enhanced CT images to delineate the boundaries of the masses, which may have led to bias in practical applications. Sixth, although identical CT acquisition and reconstruction settings were used in both centers, there is still a concern that the radiomic feature values may have been

affected by the use of different scanners (two scanners in the training cohort center, and one scanner in the validation cohort center). Thus, it may be necessary to apply a data harmonization procedure, such as ComBat and modified ComBat, for non-single center radiomics studies. Finally, there were some unusual findings for the RF model, such as the widening gap between the AUCs of this model in the training and validation sets and the parallel line in the DCA in the training set. The authors consider the RF model to possibly have the risk of overfitting.

In conclusion, ML classifiers based on unenhanced CT scans showed acceptable diagnostic efficiencies in the diagnosis of CRMs. Furthermore, KNN may be used as a potential screening method in patients with CRMs.

Acknowledgement

We thank Philips Healthcare Inc. and Medjaden Inc. for their technical support and assistance in the preparation of this manuscript.

Data sharing statement

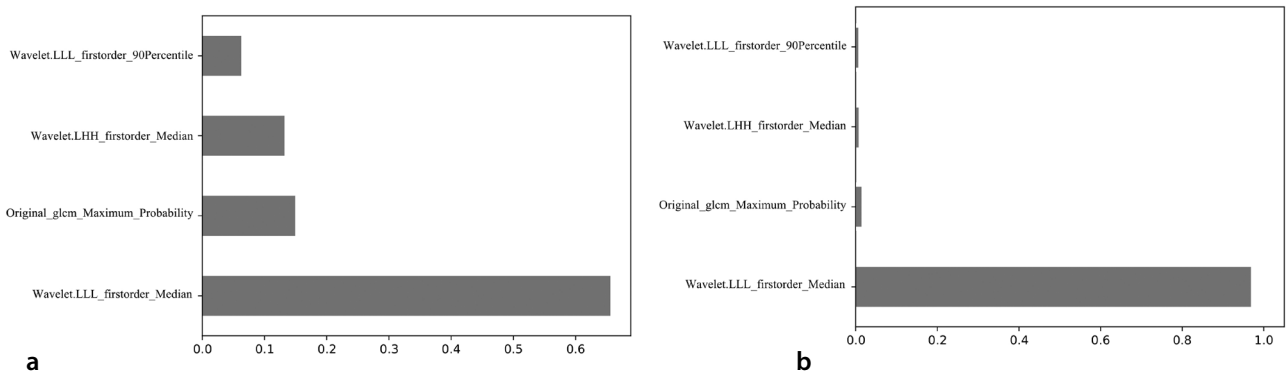
The source code, datasets, and models generated and/or analyzed during the current study are available on GitHub (https://github.com/elliiesong/CRM-screening-with-machine-learning-unenhanced-CT).

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. 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. [CrossRef]
2. Terada N, Arai Y, Kinukawa N, Yoshimura K, Terai A. Risk factors for renal cysts. *BJU Int.* 2004;93(9):1300-1302. [CrossRef]
3. Carrim ZI, Murchison JT. The prevalence of simple renal and hepatic cysts detected by spiral computed tomography. *Clin Radiol.* 2003;58(8):626-629. [CrossRef]
4. Silverman SG, Israel GM, Herts BR, Richie JP. Management of the incidental renal mass. Radiology. 2008;249(1):16-31. [CrossRef]
5. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* 2016;278(2):563-577. [CrossRef]
6. Lubner MG. Radiomics and artificial intelligence for renal mass characterization. *Radiol Clin North Am.* 2020;58(5):995-1008. [CrossRef]
7. Mühlbauer J, Egen L, Kowalewski KF, et al. Radiomics in renal cell carcinoma-a systematic review and meta-analysis. *Cancers (Basel).* 2021;13(6):1348. [CrossRef]
8. Ursprung S, Beer L, Bruining A, et al. Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma-a systematic review and meta-analysis. *Eur Radiol.* 2020;30(6):3558-3566. [CrossRef]
9. Li S, Liu J, Xiong Y, et al. Application values of 2D and 3D radiomics models based on CT plain scan in differentiating benign from malignant ovarian tumors. *Biomed Res Int.* 2022;2022:5952296. [CrossRef]
10. Yang X, He J, Wang J, et al. CT-based radiomics signature for differentiating solitary granulomatous nodules from solid lung adenocarcinoma. *Lung Cancer.* 2018;125:109-114. [CrossRef]
11. Miskin N, Qin L, Silverman SG, Shinagare AB. Differentiating benign from malignant cystic renal masses: a feasibility study of computed tomography texture-based machine learning algorithms. *J Comput Assist Tomogr.* 2023;47:376-381. [CrossRef]
12. Dana J, Lefebvre TL, Savadjiev P, et al. Malignancy risk stratification of cystic renal lesions based on a contrast-enhanced CT-based machine learning model and a clinical decision algorithm. *Eur Radiol.* 2022;32(6):4116-4127. [CrossRef]
13. He QH, Feng JJ, Lv FJ, Jiang Q, Xiao MZ. Deep learning and radiomic feature-based blending ensemble classifier for malignancy risk prediction in cystic renal lesions. *Insights Imaging.* 2023;14(1):6. [CrossRef]
14. Chen M, Cao J, Hu J, et al. Clinical-radiomic analysis for pretreatment prediction of objective response to first transarterial chemoembolization in hepatocellular carcinoma. *Liver Cancer.* 2021;10(1):38-51. [CrossRef]
15. Hu MJ, Yu YX, Fan YF, Hu CH. CT-based radiomics model to distinguish necrotic hepatocellular carcinoma from pyogenic liver abscess. *Clin Radiol.* 2021;76(2):161. [CrossRef]
16. Kocak B, Baessler B, Bakas S, et al. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. *Insights Imaging.* 2023;14(1):75. [CrossRef]
17. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749-762. [CrossRef]
18. Corica FA, Iczkowski KA, Cheng L, et al. Cystic renal cell carcinoma is cured by resection: a study of 24 cases with long-term followup. *J Urol.* 1999;161(2):408-411. [CrossRef]
19. Webster WS, Thompson RH, Cheville JC, Lohse CM, Blute ML, Leibovich BC. Surgical resection provides excellent outcomes for patients with cystic clear cell renal cell carcinoma. *Urology.* 2007;70(5):900-904. [CrossRef]
20. Jhaveri K, Gupta P, Elmi A, et al. Cystic renal cell carcinomas: do they grow, metastasize, or recur? *AJR Am J Roentgenol.* 2013;201(2):292-296. [CrossRef]
21. Cooperberg MR, Mallin K, Kane CJ, Carroll PR. Treatment trends for stage I renal cell carcinoma. *J Urol.* 2011;186(2):394-399. [CrossRef]
22. Daskivich TJ, Tan HJ, Litwin MS, Hu JC. Life expectancy and variation in treatment for early stage kidney cancer. *J Urol.* 2016;196(3):672-677. [CrossRef]
23. Kane CJ, Mallin K, Ritchey J, Cooperberg MR, Carroll PR. Renal cell cancer stage migration: analysis of the National Cancer Data Base. *Cancer.* 2008;113(1):78-83. [CrossRef]
24. Siegel CL, McFarland EG, Brink JA, Fisher AJ, Humphrey P, Heiken JP. CT of cystic renal masses: analysis of diagnostic performance and interobserver variation. *AJR Am J Roentgenol.* 1997;169(3):813-818. [CrossRef]
25. Schoots IG, Zaccai K, Hunink MG, Verhagen PCMS. Bosniak classification for complex renal cysts reevaluated: a systematic review. *J Urol.* 2017;198(1):12-21. [CrossRef]
26. Silverman SG, Israel GM, Trinh QD. Incompletely characterized incidental renal masses: emerging data support conservative management. *Radiology.* 2015;275(1):28-42. [CrossRef]
27. Demircioğlu A. The effect of preprocessing filters on predictive performance in radiomics. *Eur Radiol Exp.* 2022;6(1):40. [CrossRef]



Supplemantary Figure S1. **(a)** SHapley Additive exPlanations (SHAP) value of RF. **(b)** SHAP value of the decision tree.

Supplemantary Table S1. Comparison of the diagnostic efficiencies of 3 CT radiomic feature–based machine learning algorithms with that of a radiologist's diagnosis

Machine learning algorithm	Sensitivity	Specificity	Accuracy	PPV	NPV
		P value vs. radiologist			
RF	0.35	0.75	0.50	0.87	0.35
DT	0.86	0.75	0.73	0.90	0.60
KNN	0.84	0.61	0.90	0.84	0.78

PPV, positive predictive value; NPV, negative predictive value; RF, random forest; DT, decision tree; KNN, k-nearest neighbor; CT, computed tomography.

Supplemantary Table S2. Radiomics quality score of this research study

Criteria		Points
1	Image protocol quality - well-documented image protocols (for example, contrast, slice thickness, energy, etc.) and/or usage of public image protocols allow reproducibility/replicability	+ 2
2	Multiple segmentations - possible actions are: segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, segmentation at different breathing cycles. Analyze feature robustness to segmentation variabilities	+ 1
3	Phantom study on all scanners - detect inter-scanner differences and vendor-dependent features. Analyze feature robustness to these sources of variability	**+ 0**
4	Imaging at multiple time points - collect images of individuals at additional time points. Analyze feature robustness to temporal variabilities (for example, organ movement, organ expansion/shrinkage)	+ 0
5	Feature reduction or adjustment for multiple testing - decreases the risk of overfitting. Overfitting is inevitable if the number of features exceeds the number of samples. Consider feature robustness when selecting features	+ 3
6	Multivariable analysis with non-radiomics features (for example, EGFR mutation) - is expected to provide a more holistic model. Permits correlating/inferencing between radiomics and non-radiomics features	+ 0
7	Detect and discuss biological correlates - demonstration of phenotypic differences (possibly associated with underlying gene–protein expression patterns) deepens understanding of radiomics and biology	+ 0
8	Cut-off analyses - determine risk groups by either the median, a previously published cut-off or report a continuous risk variable. Reduces the risk of reporting overly optimistic results	+ 0
9	Discrimination statistics - report discrimination statistics (for example, C-statistic, ROC curve, AUC) and their statistical significance (for example, *P* values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)	+ 1
10	Calibration statistics - report calibration statistics (for example, calibration-in-the-large/slope, calibration plots) and their statistical significance (for example, *P* values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)	+ 1
11	Prospective study registered in a trial database - provides the highest level of evidence supporting the clinical validity and usefulness of the radiomics biomarker	+ 0
12	Validation - the validation is performed without retraining and without adaptation of the cut-off value, provides crucial information with regard to credible clinical performance	**+ 3**
13	Comparison to 'gold standard' - assess the extent to which the model agrees with/is superior to the current 'gold standard' method (for example, TNM-staging for survival prediction). This comparison shows the added value of radiomics	+ 2
14	Potential clinical utility - report on the current and potential application of the model in a clinical setting (for example, decision curve analysis)	+ 2
15	Cost-effectiveness analysis - report on the cost-effectiveness of the clinical application (for example, QALYs generated)	+ 0
16	Open science and data - make code and data publicly available. Open science facilitates knowledge transfer and reproducibility of the study	+ 2
	Total	**17**

Total points (36 = 100%). ROC, receiver operating characteristic; AUC, area under the curve; TNM, tumor, node and metastasis.

Supplemantary Table S3.

CLEAR Checklist v1.0

Note: Use the checklist in conjunction with the main text for clarification of all items.
Yes, details provided; No, details not provided; n/e, not essential; n/a, not applicable; Page, page number

Section	No.	Item	Yes	No	n/a	Page
Title						
	1	Relevant title, specifying the radiomic methodology	■	☐	☐	1
Abstract						
	2	Structured summary with relevant information	■	☐	☐	2
Keywords						
	3	Relevant keywords for radiomics	■	☐	☐	2
Introduction						
	4	Scientific or clinical background	■	☐	☐	3
	5	Rationale for using a radiomic approach	■	☐	☐	3
	6	Study objective(s)	■	☐	☐	3-4
Method						
Study Design	7	Adherence to guidelines or checklists (e.g., CLEAR checklist)	■	☐	☐	Attachment
	8	Ethical details (e.g., approval, consent, data protection)	■	☐	☐	4
	9	Sample size calculation	☐	■	☐	
	10	Study nature (e.g., retrospective, prospective)	■	☐	☐	4
	11	Eligibility criteria	■	☐	☐	4
	12	Flowchart for technical pipeline	■	☐	☐	Fig 1
Data	13	Data source (e.g., private, public)	■	☐	☐	4
	14	Data overlap	☐	■	☐	
	15	Data split methodology	■	☐	☐	4
	16	Imaging protocol (i.e., image acquisition and processing)	■	☐	☐	4
	17	Definition of non-radiomic predictor variables	☐	☐	■	4-5
	18	Definition of the reference standard (i.e., outcome variable)	■	☐	☐	4
Segmentation	19	Segmentation strategy	■	☐	☐	5
	20	Details of operators performing segmentation	■	☐	☐	5
Pre-processing	21	Image pre-processing details	■	☐	☐	5
	22	Resampling method and its parameters	☐	■	☐	
	23	Discretization method and its parameters	☐	■	☐	

Supplemantary Table S3. continued

Section	No.	Item	Yes	No	n/a	Page
	24	Image types (e.g., original, filtered, transformed)	■	☐	☐	5
Feature extraction	25	Feature extraction method	■	☐	☐	5
	26	Feature classes	■	☐	☐	5
	27	Number of features	■	☐	☐	5
	28	Default configuration statement for remaining parameters	■	☐	☐	5
Data preparation	29	Handling of missing data	☐	☐	■	
	30	Details of class imbalance	☐	☐	■	
	31	Details of segmentation reliability analysis	■	☐	☐	5-6
	32	Feature scaling details (e.g., normalization, standardization)	■	☐	☐	6
	33	Dimension reduction details	☐	■	☐	
Modeling	34	Algorithm details	■	☐	☐	6
	35	Training and tuning details	■	☐	☐	6
	36	Handling of confounders	☐	■	☐	
	37	Model selection strategy	■	☐	☐	6
Evaluation	38	Testing technique (e.g., internal, external)	■	☐	☐	6
	39	Performance metrics and rationale for choosing	■	☐	☐	7
	40	Uncertainty evaluation and measures (e.g., confidence intervals)	■	☐	☐	
	41	Statistical performance comparison (e.g., DeLong's test)	■	☐	☐	7
	42	Comparison with non-radiomic and combined methods	■	☐	☐	7
	43	Interpretability and explainability methods	☐	■	☐	
Results						
	44	Baseline demographic and clinical characteristics	■	☐	☐	7-8
	45	Flowchart for eligibility criteria	■	☐	☐	Fig 2
	46	Feature statistics (e.g., reproducibility, feature selection)	■	☐	☐	8
	47	Model performance evaluation	■	☐	☐	8-9
	48	Comparison with non-radiomic and combined approaches	■	☐	☐	Table 2
Discussion						
	49	Overview of important findings	■	☐	☐	9
	50	Previous works with differences from the current study	■	☐	☐	10
	51	Practical implications	■	☐	☐	11
	52	Strengths and limitations (e.g., bias and generalizability issues)	■	☐	☐	11-12

Supplemantary Table S3. continued

Section	No.	Item	Yes	No	n/a	Page
Open Science						
Data availability	53	Sharing images along with segmentation data [n/e]	☐	■	☐	
	54	Sharing radiomic feature data	■	☐	☐	14
Code availability	55	Sharing pre-processing scripts or settings	☐	☐	■	
	56	Sharing source code for modeling	■	☐	☐	14
Model availability	57	Sharing final model files	☐	■	☐	
	58	Sharing a ready-to-use system [n/e]	☐	☐	■	

Kocak B, Baessler B, Bakas S, Cuocolo R, Fedorov A, Maier-Hein L, Mercaldo N, Müller H, Orlhac F, Pinto Dos Santos D, Stanzione A, Ugga L, Zwanenburg A. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights Imaging. 2023 May 4;14(1):75. doi: 10.1186/s13244-023-01415-8

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.232472

BREAST IMAGING

ORIGINAL ARTICLE

Diagnostic value of qualitative and quantitative enhancement parameters on contrast-enhanced mammography

Musa Kul[1]
Selçuk Akkaya[2]
Sibel Kul[2]

[1]University of Health Sciences, Trabzon Faculty of Medicine, Kanuni Health Research Center, Department of Radiology, Trabzon, Türkiye

[2]Karadeniz Technical University, Faculty of Medicine, Department of Radiology, Trabzon, Türkiye

PURPOSE

To determine whether qualitative and quantitative enhancement parameters obtained from contrast-enhanced mammography (CEM) can be used in predicting malignancy.

METHODS

After review board approval, consecutive 136 suspicious lesions with definite diagnosis were retrospectively analyzed on CEM. Acquisition was routinely started with craniocaudal view and ended with mediolateral oblique view of the affected breast. Lesion conspicuity (low, moderate, high), internal enhancement pattern (homogeneous, heterogeneous, rim), contrast-to-noise ratio (CNR), percentage of signal difference (PSD) and relative enhancement from early to late view were analyzed. PSD and relative enhancements were used to determine patterns of descending, steady or ascending enhancements. Receiver operating characteristic analysis, Cohen's kappa statistics and Spearman correlation tests were used.

RESULTS

There were 29 benign and 107 malignant lesions. 64% of the malignant lesions exhibited high conspicuity compared to 14% of the benign lesions ($P < 0.001$). CNR values were higher in malignant lesions compared to benign ones ($P \leq 0.004$). CNR from early view yielded 82% sensitivity, 72% specificity and PSD yielded 79% sensitivity, 65% specificity. Descending pattern and rim enhancement observed in 44% and 21% of breast cancers, respectively, and both provided 96% positive predictive value for malignancy.

CONCLUSION

Diagnostic accuracy of quantitative parameters was higher than that of qualitative parameters. High CNR, rim enhancement, and descending pattern were features commonly seen in malignant lesions, while low CNR, homogeneous enhancement, and ascending pattern were commonly seen in benign lesions.

KEYWORDS

BI-RADS, breast cancer, contrast-enhanced mammography, enhancement parameters, pharmacokinetics

Corresponding author: Sibel Kul

E-mail: sibel_ozy@yahoo.com

Received 27 September 2023; revision requested 13 November 2023; last revision received 05 February 2024; accepted 18 March 2024.

Epub: 15.04.2024

Publication date: 08.07.2024

DOI: 10.4274/dir.2024.232472

Contrast-enhanced mammography (CEM) is a recently developed advanced digital mammography (DM) technique that uses low- and high-energy acquisitions following the administration of an intravenous iodine contrast agent. Low-energy and recombined images are finally obtained for each of the craniocaudal (CC) and mediolateral oblique (MLO) projections involved.[1,2] The low-energy images are similar and comparable to those of conventional DM.[3,4] Recombined images demonstrate the iodine uptake of breast lesions secondary to angiogenesis on a suppressed background of normal fibroglandular tissue. The physiological and morphological information obtained from CEM is thus similar to that yielded by dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). CEM combines the relative simplicity and low cost of mammography with the high sensitivity of contrast-enhanced imaging. The technique dramatically improves the ability of DM in the detection and characterization of breast lesions.[5,6]

You may cite this article as: Kul M, Akkaya S, Kul S. Diagnostic value of qualitative and quantitative enhancement parameters on contrast-enhanced mammography. *Diagn Interv Radiol*. 2024;30(4):248-255.

In CEM, low-energy and recombined images are reviewed, and morphological parameters, together with the presence of enhancement, are used for lesion characterization.[7] Since both benign and malignant lesions may exhibit enhancement, it may not be possible to differentiate them solely through enhancement.[1,2,6] Therefore, routinely evaluated parameters of enhancement intensity, pattern, and kinetics used in the diagnosis, molecular subtyping, and prognostication of breast cancer in DCE-MRI[8-11] have recently become the subject of CEM research.[12-18] This retrospective study aims to determine whether qualitative and quantitative enhancement parameters obtained from CEM could be used for predicting malignancy.

Main points

- The conspicuity and contrast-to-noise ratio (CNR) of malignant breast lesions are usually higher than those of benign lesions.
- The diagnostic values of quantitative enhancement parameters are higher than those of the qualitative parameters, and CNR at early-phase images is the most successful among them.
- Rim enhancement is the least common internal pattern but is highly predictive of malignancy.
- Quantitatively evaluated descending patterns and negative percentage of signal difference values are highly predictive of malignancy.
- On contrast-enhanced mammography, while high enhancement intensity, rim enhancement, and the descending pattern of relative enhancement in a lesion indicate a malignancy, low enhancement intensity, homogeneous enhancement, and ascending pattern indicate a benign lesion.

Methods

Patient population

This retrospective study was approved by the KTU Medical Faculty SCI Research Ethics Committee Ethics Committee (date: 12.04.2019, decision no: 24237859-295), with informed consent being waived. Consecutive cases that had undergone CEM between June 2014 and February 2022 in our hospital were retrospectively reviewed. In our practice, we offer CEM instead of DM to diagnose patients with complaints of a palpable mass and spontaneous nipple discharge or nipple retraction if our targeted fast ultrasound (US) related to the area of interest reveals suspicious findings. CEM is also performed in some patients with suspicious findings following screening mammography if there is no contraindication for contrast administration.

The inclusion criteria of the present study were as follows:

1. CEM was performed due to suspicious breast lesions determined by clinical examination, mammography, or an US.

2. The CEM exam obtained both CC and MLO views for the breast with the suspicious lesion.

3. Definite diagnoses were provided through either a histopathological examination of the surgically excised or needle-biopsied specimens, or by follow-up for lesion stability of at least 2 years.

The exclusion criteria were as follows:

1. Patients receiving neoadjuvant chemotherapy,

2. Patients with any contraindications to contrast material administration,

3. Patients with a known or suspected pregnancy,

4. For multifocal cases, lesions that were superimposed on one another in the projections,

5. Suspicious breast lesions displayed on a single projection.

Contrast-enhanced mammography technique and analysis

CEM examinations were performed using Senographe Essential full-field DM equipment (GE Healthcare, Buc, France) in our breast imaging unit. Iopromide (Ultravist 300) (300 mg/mL at 1.5 mL/kg and not exceeding a maximum dose of 120 mL) was administered intravenously from the antecubital fossa at a rate of 3 mL/s using a power injector, followed by a 20-mL saline flush. Two minutes after the injection, the acquisition was started with a CC image of the affected breast and continued with a CC image of the normal breast and an MLO image of the normal breast, and ended with an MLO image of the affected breast. The four recombined images were generated automatically by processing low- and high-kV images.

All images were evaluated on a mammography workstation in consensus by two radiologists who were experienced in breast imaging. During these analyses, the radiologists were blinded to clinical information and the final diagnosis.

The qualitative assessment steps involved the following:

1. Lesion type [mass or non-mass enhancement (NME)],

2. Conspicuity according to the enhancement intensity of the lesion (low, moderate, or high),

3. Internal enhancement pattern (homogeneous, heterogeneous, or rim enhancement for masses, and homogeneous, heterogeneous, or clumped for NME),[19]

4. Assessment of the relative enhancement pattern (ascending, steady, or descending) by visual evaluation of the change in conspicuity of the lesion from the CC (early phase) to MLO (late phase) views. Increase conspicuity from the CC to MLO views was defined as an ascending pattern; if no visual alteration was present in the lesion conspicuity it was defined as steady; a decrease in the conspicuity from the CC to MLO views was defined as a descending pattern.

Examples of the enhancement parameters that were used in the qualitative assessment are shown in Figures 1 and 2.

The quantitative assessments involved the following:

1. Maximum tumor size,

2. Contrast-to-noise ratio (CNR),

3. The percentage of signal difference (PSD) from the CC to MLO views and classification of the relative enhancement patterns as ascending, steady, or descending.

To calculate the CNR, relative gray values were used. A region of interest (ROI) covering the entire lesion was manually drawn on the recombined images. A separate circular ROI was placed over the background tissue showing the most homogeneous signal, adjacent to the tumor. The following formula was applied:

$$CNR = T_{mean} - BG_{mean} / BG_{sd}$$

T_{mean} = mean pixel value in the ROI of the tumor

BG_{mean} = mean pixel value in the ROI of the background

BG_{sd} = standard deviation in the ROI of the background

The CNRs obtained from the recombined images of the CC and MLO views were used as a quantitative measure of early (CNR_1) and late (CNR_2) phase tumor enhancement, respectively (Figure 3).

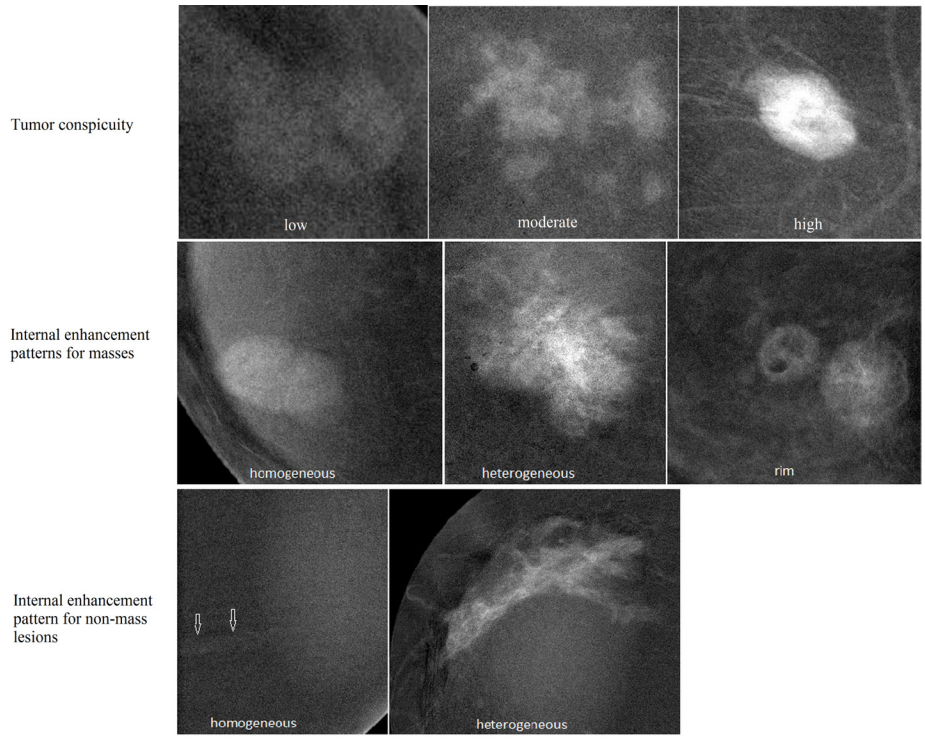


Figure 1. Examples of tumor conspicuity groups (row 1) and internal enhancement patterns for both masses (row 2) and non-mass enhancements (row 3) are demonstrated on contrast-enhanced mammography images.

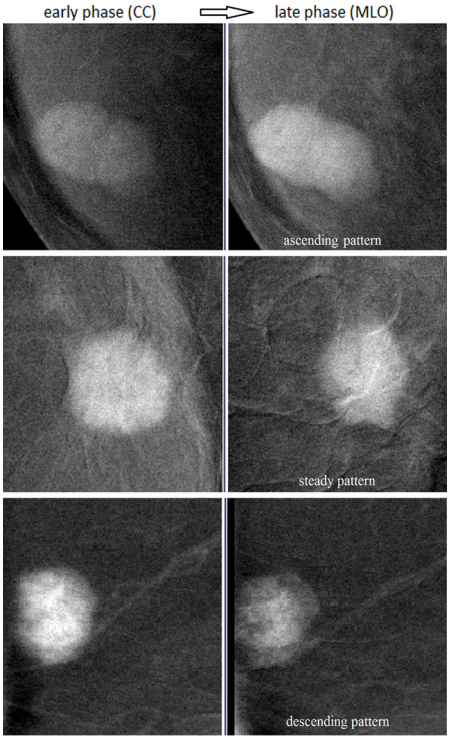


Figure 2. Examples of three qualitatively evaluated relative enhancement patterns on contrast-enhanced mammography. The images in the first column are craniocaudal views (early phase) and those in the second column are mediolateral oblique views (late phase).

The PSD was calculated using the following formula;

$$PSD = (CNR_2 - CNR_1 / CNR_1) \times 100$$

An increase in CNR from the early to late phase exceeding 10% was recorded as ascending and a decrease greater than 10% as descending, while values in between were considered to reflect a steady pattern.

Statistical analysis

Data analysis was performed using the SPSS (v.23.0, IBM, Armonk, NY, USA) software. Descriptive statistics were expressed as mean ± standard deviation, median and minimum–maximum values for continuous variables, and as a number (n) and percentage (%) for categorical variables. The normal distribution of variables was assessed using the Kolmogorov–Smirnov test. The chi-square and Fisher's exact tests were used to test for a significant difference between benign and malignant groups for the frequency distribution of categorical variables, and a Student's t-test or the Mann–Whitney U test for interval variables. Interval variables were compared between the independent groups (CNR_1 with CNR_2) using the Wilcoxon test. Receiver operating characteristic (ROC) analysis was performed to determine the diagnostic accuracy of the CNR and PSD parameters by calculating the area under the curve (AUC) values. Sensitivity and specificity were calculated after determining the optimal cut-off values for those parameters. Cohen's kappa statistics and Spearman correlation tests were conducted to document the agreement and association between the qualitative and quantitative CEM parameters. A *P* value of <0.05 was considered statistically significant. The Power and Precision (v3.2) software was used for power calculations.

Results

A total of 136 clinically or radiologically suspicious breast lesions among 105 women (3 lesions in 1, 2 lesions in 26, and single lesions in the remaining 81 women) with a median age of 46 years (range: 26–71 years) were evaluated. Of the 136 lesions, 29 (21%) were benign and 107 (79%) were malignant. The mean size of the lesions was 26 ± 17 mm. There was no statistical difference between the sizes of benign and malignant lesions (*P* = 0.390). Final diagnoses were obtained through the histopathological examination of the surgically excised or needle-biopsied specimens in 130 lesions. A total of 71 invasive ductal carcinoma (IDC), 10 invasive lobular carcinoma, 10 IDC with ductal carcinoma in situ (DCIS), 7 pure DCIS, 4 mixed invasive ductal-lobular carcinoma, and 5 other malignancies were identified. For the benign lesions, there were 6 cases of fibrocystic changes/epithelial hyperplasia, 4 cases of mastitis, 4 fibroadenomas, 3 papillo-

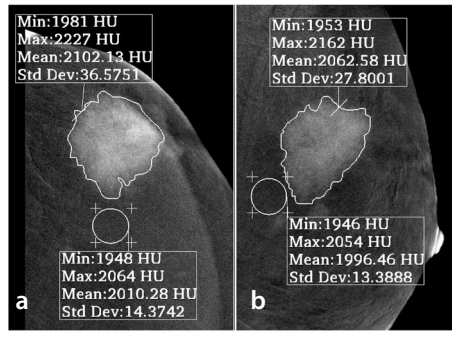


Figure 3. To calculate the enhancement intensity of the lesion, a region of interest (ROI) covering the entire lesion was manually drawn on the recombined images, and a separate circular ROI was placed over the most homogenous background tissue adjacent to the tumor. The mean pixel values of the tumor (T_{mean}) and background tissue (BG_{mean}), as well as the standard deviation of the signal from the background tissue (BG_{sd}), were used to calculate the contrast-to-noise ratio ($CNR = T_{mean} - BG_{mean} / BG_{sd}$); CNR_1 of the mass characterized as an invasive ductal carcinoma (measured on the craniocaudal view) was 6.40 **(a);** CNR_2 measured on the mediolateral oblique view was 5.01 **(b).** The percentage of signal difference was –22, representing a descending pattern. CNR, contrast-to-noise ratio.

mas, and 6 other benign lesions. Another 6 lesions were characterized as stable during a follow-up of at least 2 years.

For the affected breasts, the mean time interval between the start of the contrast injection and CC view was 148 ± 34 s; between the start of the contrast injection and the MLO view, this interval was 270 ± 55 s. Descriptive and CEM imaging features of the lesions are documented in Table 1. Among the lesions, 121 (89%) were masses and 15 (11%) were classified as NME.

Tumor conspicuity was significantly higher in the malignant lesions compared with benign lesions ($P < 0.001$); 86% of the benign lesions had low-moderate conspicuity, whereas 64% of the malignant lesions had high conspicuity. Of the 4 benign lesions with high conspicuity, 3 were mastitis and 1 was epithelial hyperplasia. Eight malignant tumors had low conspicuity, 2 of which were DCIS; the remainder were invasive cancer tumors smaller than 15 mm.

As quantitative parameters of tumor enhancement, both CNR_1 and CNR_2 were higher for the malignant lesions compared with the benign legions ($P < 0.001$). In the benign lesions, CNR_2 was significantly higher than CNR_1 ($P < 0.001$), whereas CNR_1 was significantly higher than CNR_2 in the malignant lesions ($P = 0.045$) (Figure 4). The PSD was significantly higher in benign lesions compared with malignant lesions (36.3 vs. 0.6, $P < 0.001$) (Figure 5). 57% (61/107) of the malignant lesions had negative PSD values, and 86% (25/29) of benign lesions had positive PSD values. While malignant lesions frequently tended to lose their enhancement at late acquisition, benign lesions commonly exhibited ascending enhancement. When ROC curves were plotted using the final diagnoses as a reference, AUC values of 0.816 [95% confidence interval (CI), 0.738–0.893], 0.717 (95% CI, 0.624–0.818), and 0.726 (95% CI, 0.627–0.825) were obtained for CNR_1, CNR_2, and PSD, respectively (Figure 6). A cut-off value of 2.50 for CNR_1 yielded 82% sensitivity and 72% specificity, whereas a cut-off value of 10% for PSD yielded 79% sensitivity and 65% specificity. The power of the study for CNR_1 and PSD was 99.4% and 94.9%, respectively.

Of the 11 malignant NMEs, 6 exhibited clumped, 4 heterogeneous, and 1 homo-

Table 1. Descriptive and contrast-enhanced mammography imaging features of the 136 breast lesions in this study

	Benign	Malignant	*P* value
Number of cases	29	107	
Lesion size (mm)			
Mean ± standard deviation	23.8 ± 21.4	27.1 ± 16.2	
Minimum–maximum	5–96	6–100	0.390
Median	18	25	
Lesion type**			
Mass	25 (86%)	96 (90%)	0.019
NME	4 (14%)	11 (10%)	
CEM qualitative enhancement parameters			
Lesion conspicuity**			
Low	12 (41%)	8 (8%)	
Moderate	13 (45%)	30 (28%)	<0.001
High	4 (14%)	69 (64%)	
Internal enhancement pattern**			
Homogeneous	15 (52%)	22 (21%)	
Heterogeneous	13 (45%)	58 (54%)	0.001
Rim	1 (3%)	27 (25%)	
Relative enhancement pattern**			
Ascending	16 (55%)	34 (32%)	
Steady	12 (35%)	38 (35%)	0.024
Descending	3 (10%)	35 (33%)	
CEM quantitative enhancement parameters			
CNR_1*	2.26 ± 0.93	4.34 ± 2.09	<0.001
CNR_2*	2.83 ± 1.07	4.05 ± 1.81	<0.001
PSD*	36.3 ± 48.8	0.6 ± 32.6	<0.001
Relative enhancement pattern**			
Ascending	20 (69%)	36 (34%)	
Steady	7 (24%)	24 (22%)	<0.001
Descending	2 (7%)	47 (44%)	

*Data represent mean values ± standard deviation, **Data represent numbers and percentages of cases. NME, non-mass enhancement; CEM, contrast-enhanced mammography; CNR, contrast-to-noise ratio; PSD, percentage of signal difference.

geneous enhancement. Of the 4 benign NMEs, 2 exhibited homogeneous, 1 heterogeneous, and 1 clumped enhancement. Clumped enhancement was grouped in the heterogeneous pattern for statistical analyses. The most common internal enhancement pattern was homogeneous in benign lesions (52%) and heterogeneous (54%) in malignant lesions ($P = 0.001$). Rim enhancement was seen in 28% (27/96) of malignant and 4% (1/25) of benign masses ($P = 0.020$). The benign lesion with rim enhancement was a 25-mm papilloma. Rim enhancement revealed a positive predictive value (PPV) of 96% for a breast cancer diagnosis.

Ninety percent of benign lesions exhibited an ascending-steady pattern on the visual evaluation of relative enhancements. Descending enhancement was detected in 33% of malignant and 10% of benign lesions ($P = 0.032$). In the quantitative evaluation, the most common pattern was ascending in benign lesions (69%) and descending in malignant lesions (44%). The ascending pattern was significantly more common in benign than malignant lesions (69% vs. 35%) ($P = 0.002$). A descending pattern was observed in 44% (47/107) of malignant and 7% (2/29) of benign lesions ($P = 0.001$). All benign lesions, except for 2 papillomas, exhibited ascending or steady-type enhancement. Visually and quantitatively evaluated descending patterns had a PPV of 92% and 96% for breast cancer diagnosis, respectively.

A strong positive correlation was present between tumor conspicuity and CNR_1 values (correlation coefficient: 0.831, $P < 0.001$) (Figure 7). Additionally, visually and quantitatively analyzed relative enhancement patterns were compatible in 70% of cases (kappa: 0.548, $P < 0.001$).

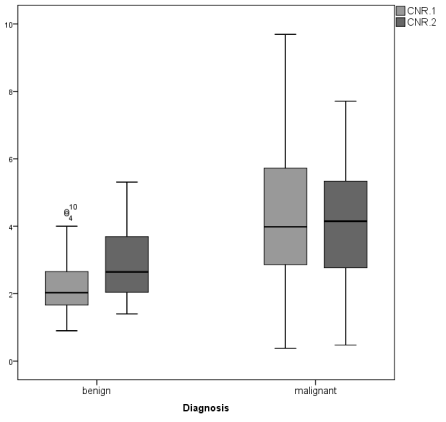


Figure 4. Comparison of early- (CNR_1) and late-phase enhancement intensity (CNR_2) values of benign and malignant lesions. Both CNR_1 and CNR_2 were higher in malignant lesions compared with benign lesions. CNR, contrast-to-noise ratio.

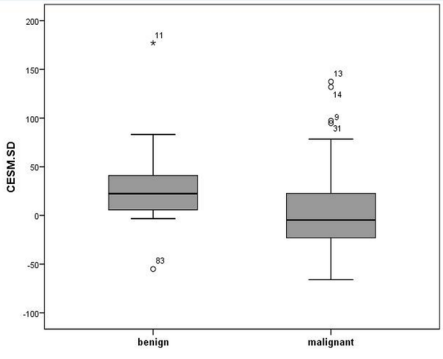


Figure 5. A comparison of the percentage signal difference (PSD) values of benign and malignant lesions; the PSD was significantly higher in benign lesions compared with malignant lesions (28.3 vs. 1.3).

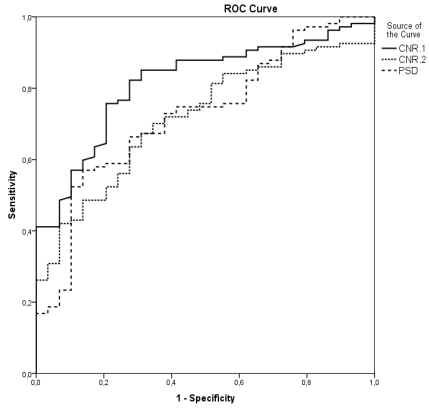


Figure 6. Receiver operating characteristic curves for CNR_1, CNR_2, and PSD in the diagnosis of malignant breast lesions. Area under the curve values of 0.816, 0.717, and 0.726 were obtained for CNR_1, CNR_2, and PSD, respectively. CNR, contrast-to-noise ratio; PSD, percentage signal difference.

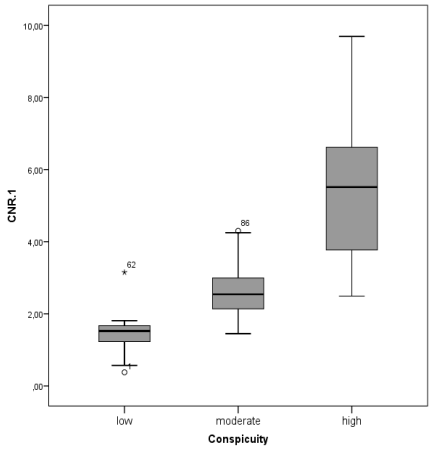


Figure 7. The correlation between lesion conspicuity and quantitative enhancement intensity (CNR_1). CNR, contrast-to-noise ratio.

For both benign and malignant lesions, tumor conspicuity and CNR_1 were found to be significantly higher in tumors equal to or larger than 20 mm in diameter compared with smaller tumors ($P < 0.010$). A statistically significant positive correlation was detected between lesion size and CNR_1 value (0.693 for benign, and 0.313 for malignant lesions, $P < 0.001$). There was no difference between the enhancement intensities of mass and non-mass-like (NML) lesions. In the benign group, the mean CNR_1 was 2.19 for masses and 2.67 for NMLs ($P = 0.647$). In the malignant group, the mean CNR_1 was 4.39 for masses and 3.93 for NMLs ($P = 0.821$). Additionally, no significant difference was observed for CNR_1 between invasive and non-invasive cancers (4.4 and 3.6, respectively, $P = 0.457$), or between benign lesions and non-invasive cancer (2.5 and 3.6, respectively, $P = 0.231$).

Discussion

In the present study, we evaluated the three enhancement parameters (degree of enhancement, internal enhancement pattern, and relative change in enhancement intensity from early to late projection) of 136 clinically or radiologically suspicious breast lesions using qualitative and quantitative analyzes on CEM. The distribution of all the enhancement parameters differed significantly between benign and malignant lesions.

In 2022, the American College of Radiology released a supplement that included the first version of the breast imaging reporting and data system lexicon for CEM to standardize the interpretation and reporting of imaging findings.[19] One of the parameters investigated in the present study was an enhancement descriptor of this lexicon called "lesion conspicuity," which is defined as the degree of enhancement relative to the background. Low conspicuity, which refers to enhancement equal to or slightly greater compared with the background, was present in 41% of the benign and 8% of the malignant cases. Conversely, 64% of malignant and 14% of benign tumors had high conspicuity. Tumors with high conspicuity were more likely to be malignant and reflected 64% sensitivity and 86% specificity. Previously, Nicosia et al.[20] evaluated lesion conspicuity in recombined CEM images and reported 80% sensitivity and 72% specificity when moderate and high conspicuity were accepted as predictive of malignancy. Several studies were also conducted on tumor enhancement before publication of the CEM lexicon.[2,9,15,16,21] These studies also found a significant correlation between the

degree of enhancement and the probability of a lesion being malignant. Minimal or no enhancement has been reported in 22%–73% of benign and up to 8% of malignant lesions. More than 40% of malignant lesions with minimal or no enhancement in those studies were DCISs. In the present study, only 2 of 8 lesions with low conspicuity were DCISs, and the remaining lesions were invasive cancer with a size smaller than 15 mm. Low conspicuity and the absence of enhancement, although highly predictive of benign lesions, do not exclude in situ or invasive cancer on CEM images.

We detected a strong positive correlation between tumor conspicuity and the CNR_1 value (correlation coefficient: 0.831). According to quantitatively evaluated parameters, CNR_1 (a measure of early-phase enhancement intensity) (AUC: 0.816) was more effective than CNR_2 (a measure of late-phase enhancement intensity) (AUC: 0.717) and PSD (a measure of percentage changes in the enhancement intensity from the early to late phase) (AUC: 0.726) for differentiating malignant from benign lesions. As in the case of lesion conspicuity, tumors with a higher CNR_1 were more likely to be malignant, reflecting 82% sensitivity and 72% specificity. Liu et al.[14] and Rudnicki et al.[15] also used quantitative methods to evaluate the enhancement intensity of lesions in two projections. Liu et al.[14] reported a significant difference between the enhancement intensities in two projections and reported that earlier projections of contrast-enhanced images (AUC: 0.843) played a more important role in the differential diagnosis of breast lesions compared with later projections (AUC: 0.755), which was similar to the results obtained in our study. However, Rudnicki et al.[15] could not detect a significant enhancement difference between early- and late-phase images (AUCs of 0.725 and 0.713), which was likely related to the short time interval between the images.

The optimal timing for imaging the affected breast to better differentiate benign from malignant lesions is not known. Additionally, there are many variations in the methods used to evaluate the degree of enhancement in CEM studies.[2,14-16] These are not just at the level of formulas used to calculate enhancement level but also at the area chosen for background signal measurement and the ROI size, extent and placement. A study conducted by Lv et al.[16] compared the effectiveness of relative gray values when different locations for background signal (the area around the lesion, away from the lesion, close to the chest wall, and the chest wall) were chosen. The authors found that relative gray values were more effective when the background area around the lesion was used. In our study, we preferred using the background area around the lesion, with reference to the research of Lv et al.[16] However, future studies are needed to compare the effectiveness of different techniques and methods.

Both in the present study and the study conducted by Liu et al.[14], no significant difference was observed between the enhancement intensities of non-invasive and invasive cancers. However, Rudnicki et al.[15] reported significantly higher enhancement levels for infiltrating compared with non-infiltrating cancers. On the other hand, tumor size emerged as an important determinant of enhancement intensity in our study. Larger tumors demonstrated higher conspicuity and CNR_1 values, which was likely related to the effect of tumor volume on the projection images. While this positive correlation between size and CNR_1 was prominent for benign tumors (correlation coefficient: 0.693), it was fair (correlation coefficient: 0.313) for breast cancers; this was likely the result of a reduced enhancement caused by tumoral necrosis, which is commonly observed in large malignant lesions.

The level of tumor enhancement is not the only enhancement descriptor that should be considered; another descriptor is the internal enhancement pattern, which is classified as homogeneous, heterogeneous, or rim types for masses and homogeneous, heterogeneous, or clumped types for NMLs.[19] Different from the MRI lexicon, the internal patterns of clustered ring and non-enhancing septations are not present in the CEM lexicon. The lower resolution of CEM compared with MRI hinders the discernibility between these two patterns. In our study population, the most common internal enhancement pattern was homogeneous for benign and heterogeneous for malignant lesions. Heterogeneous and rim enhancement as indicators of malignancy exhibited 79% sensitivity. However, specificity was extremely low (52%). Although not a common finding, rim enhancement was documented in 28% of malignant and 4% of benign masses and provided 96% PPV. Previously, Chi et al.[13] reported rim enhancement in 11% (33/312) of lesions, among which 67% were malignant and 33% were benign. Kamal et al.[22] detected rim enhancement in 14% (24/168) of masses, 54% of which were benign and 46% were malignant. Contrary to these studies[9,22] that reported rim enhancement as an unreliable sign for predicting malignancy, we found rim enhancement to be a highly predictive feature for breast cancer.

Enhancement kinetics are routinely used for the characterization of breast tumors in DCE-MRI. In general, benign lesions exhibit a persistent pattern, whereas malignant lesions reflect a wash-out pattern. A plateau can be observed in both benign and malignant lesions. Kuhl et al.[11] previously reported a washout pattern in 57% of malignant and 6% of benign lesions, and a persistent pattern in 83% of benign and 9% of malignant lesions as the worst curve-type on DCE-MRI. However, a low percentage of malignant tumor volume (reported as 7%–40% for invasive cancers) shows washout pattern.[23-25] Kim et al.[23] reported the worst curve type as wash-out in 84% of breast cancers, whereas the predominant curve type was persistent in 96% of cases.

Only two mammographic projections were used in this study to evaluate changes in the enhancement intensity of lesions from the early to late phase, which was not sufficient for conducting an actual kinetic evaluation; instead, we identified it as a relative enhancement pattern in the present study. As previously described, enhancement values that were evaluated on recombined views represented the entire tumor volume, and the mean gray values were used in the CNR calculation. The relative enhancement patterns that we obtained were more like the predominant curves of DCE-MRI studies. Therefore, the descending pattern rate in the present and previous CEM studies[2,5,14,18,26] was not as high as the wash-out pattern reported in Kuhl's et al.[11] study. In the present study, ascending enhancement was significantly more common in benign lesions (69% vs. 35%), while descending enhancement was observed in 43% of malignant and 7% of benign lesions; these results reflect those of previous studies.[2,14] The PSD was significantly lower in malignant lesions compared with benign ones (1.3 vs. 28.3, $P < 0.001$), and 94% of lesions with negative PSD values were malignant. Quantitatively evaluated descending patterns and negative PSD values were highly predictive for malignancy (PPV of 96% and 94%, respectively). The quantitative assessment of relative enhancement patterns was more effective than the qualitative assessment concerning the characterization of breast tumors. However, CNR_1 was the most valuable parameter in our study in terms of the characterization of breast lesions, in contrast with the recent study conducted by Rong et al.[18], which reported a kinetic pattern as being more effective than enhancement intensity.

There are several limitations in the present study. First, it included patients from a single institution, and, specifically, the number of benign cases, non-mass lesions, and non-invasive cancers among the patients was limited. Since patients for whom there was a high suspicion of the presence of a malignancy had undergone CEM, there was an inherent bias in favor of malignancy. Nevertheless, the study had adequate statistical power. Second, the enhancement parameters in different histopathologies and lesion types were not discussed due to the small number of cases included in each subtype. To remove the bias and increase the impact of the findings, further large-scale multicenter studies including screening cohorts may be helpful. Third, two different views were applied for the analysis of relative enhancement patterns through the evaluation of changes in enhancement intensity from the early to late phases. This does not reflect a true kinetic evaluation, and we are unsure how comparable this method is with kinetics obtained from DCE-MRI. Further studies with MRI correlations are thus needed. The more accurate evaluation of enhancement kinetics could likely have been achieved if two or more acquisitions in the same projection had been used in the analysis. Additionally, the optimal time intervals for demonstrating enhancement kinetics in CEM are unclear. Although we adjusted the time intervals, in line with previous DCE-MRI studies, it was unclear whether iodine in CEM acted similarly to gadolinium in DCE-MRI. Fourth, CEM is a two-dimensional method, and the enhancement values are relative values affected by the size of the lesion, as well as the size and composition of superimposed normal breast tissue. Therefore, it might offer only limited insight into the temporal changes in tumoral enhancement. Furthermore, we used the entire tumor area and mean values for the calculation of quantitative enhancement parameters and did not compare these with other potential measurements, such as using the tumor area exhibiting the highest enhancement or using maximum values instead of means. Additionally, inter-observer variability was not evaluated. Finally, we evaluated the probable value of CEM enhancement parameters in the differentiation of benign and malignant lesions. However, these parameters may also correlate with the prognostic factors and molecular subtypes of breast cancer, which were not analyzed in the present study.

In conclusion, enhancement intensity and the relative enhancement patterns of breast tumors can be evaluated both qualitatively and quantitatively on CEM images. Combined with internal enhancement patterns, they can be used in the differential diagnosis of breast lesions. Quantitative parameters appear to be more diagnostic than qualitative parameters, and the relative enhancement intensity on early-phase images (CNR_1) is the most successful among them. However, it should be noted that the enhancement intensity on CEM depends on the lesion size, and although low enhancement is highly predictive of a lesion being benign, it does not exclude in situ or even invasive cancers. While high enhancement intensity at the early phase, rim enhancement, and descending patterns are features that are considered highly predictive for malignancy, low enhancement intensity, homogeneous enhancement, and ascending patterns are more predictive for benign lesions. These enhancement parameters are capable of contributing to CEM in lesion characterization and may also have prognostic value for breast cancer patients. This is a subject that requires further investigation.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Polat DS, Evans WP, Dogan BE. Contrast-enhanced digital mammography: technique, clinical applications, and pitfalls. *AJR Am J Roentgenol.* 2020;215(5):1267-1278. [CrossRef]
2. Deng CY, Juan YH, Cheung YC, et al. Quantitative analysis of enhanced malignant and benign lesions on contrast-enhanced spectral mammography. *Br J Radiol.* 2018;91(1086):20170605. [CrossRef]
3. Francescone MA, Jochelson MS, Dershaw DD, et al. Low energy mammogram obtained in contrast-enhanced digital mammography (CEDM) is comparable to routine full-field digital mammography (FFDM). *Eur J Radiol.* 2014;83(8):1350-1355. [CrossRef]
4. Lalji UC, Jeukens CR, Houben I, et al. Evaluation of low-energy contrast-enhanced spectral mammography images by comparing them to full-field digital mammography using EUREF image quality criteria. *Eur Radiol.* 2015;25(10):2813-2820. [CrossRef]
5. Dromain C, Balleyguier C, Muller S, et al. Evaluation of tumor angiogenesis of breast carcinoma using contrast-enhanced digital mammography. *AJR Am J Roentgenol.* 2006;187(5):528-537. [CrossRef]
6. James JJ, Tennant SL. Contrast-enhanced mammography (CESM). *Clin Radiol.* 2018;73(8):715-723. [CrossRef]
7. Perry H, Phillips J, Dialani V, et al. Contrast-enhanced mammography: a systematic guide to interpretation and reporting. *AJR Am J Roentgenol.* 2019;212(1):222-231. [CrossRef]
8. Kim JY, Kim SH, Kim YJ, et al. Enhancement parameters on dynamic contrast enhanced breast MRI: do they correlate with prognostic factors and subtypes of breast cancers? *Magn Reson Imaging.* 2015;33(1):72-80. [CrossRef]
9. Wang C, Wei W, Santiago L, Whitman G, Dogan B. Can imaging kinetic parameters of dynamic contrast-enhanced magnetic resonance imaging be valuable in predicting clinicopathological prognostic factors of invasive breast cancer? *Acta Radiol.* 2018;59(7):813-821. [CrossRef]
10. Macura KJ, Ouwerkerk R, Jacobs MA, Bluemke DA. Patterns of enhancement on breast MR images: interpretation and imaging pitfalls. *Radiographics.* 2006;26(6):1719-1734. [CrossRef]
11. Kuhl CK, Mielcareck P, Klaschik S, et al. Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology.* 1999;211(1):101-110. [CrossRef]
12. Mohamed Kamal R, Hussien Helal M, Wessam R, Mahmoud Mansour S, Godda I, Alieldin N. Contrast-enhanced spectral mammography: Impact of the qualitative morphology descriptors on the diagnosis of breast lesions. *Eur J Radiol.* 2015;84(6):1049-1055. [CrossRef]
13. Chi X, Zhang L, Xing D, Gong P, Chen Q, Lv Y. Diagnostic value of the enhancement intensity and enhancement pattern of CEM to benign and malignant breast lesions. *Medicine (Baltimore).* 2020;99(37):e22097. [CrossRef]
14. Liu Y, Zhao S, Huang J, et al Quantitative analysis of enhancement intensity and patterns on contrast-enhanced spectral mammography. *Sci Rep.* 2020;10(1):9807. [CrossRef]
15. Rudnicki W, Heinze S, Niemiec J, et al. Correlation between quantitative assessment of contrast enhancement in contrast-enhanced spectral mammography (CESM) and histopathology-preliminary results. *Eur Radiol.* 2019;29(11):6220-62266. [CrossRef]
16. Lv Y, Chi X, Sun B, Lin S, Xing D. Diagnostic value of quantitative gray-scale analysis of contrast-enhanced spectral mammography for benign and malignant breast lesions. *J Comput Assist Tomogr.* 2020;44(3):405-412. [CrossRef]
17. Rudnicki W, Heinze S, Popiela T, Kojs Z, Luczynska E. Quantitative assessment of contrast enhancement on contrast enhancement spectral mammography (CESM) and comparison with qualitative assessment. *Anticancer Res.* 2020;40(5):2925-2932. [CrossRef]
18. Rong X, Kang Y, Xue J, Han P, Guang Y, Li Z. Contrast-enhanced spectral mammography:

are kinetic patterns useful for differential diagnoses of enhanced lesions? *Diagn Interv Radiol*. 2023;29(2):244-250. [CrossRef]

19. American College of Radiology. ACR BI-RADS Atlas 5th edition. Published 2022. Accessed Feb 22, 2023. [CrossRef]
20. Nicosia L, Bozzini AC, Palma S, et al. Breast imaging reporting and data system and contrast enhancement mammography: lesion conspicuity likelihood of malignancy and relationship with breast tumor receptor status. *Acad Radiol*. 2023;30(10):2243-2251. [CrossRef]
21. Łuczyńska E, Niemiec J, Hendrick E, et al. Degree of enhancement on contrast enhanced spectral mammography (CESM) and lesion type on mammography (MG): comparison based on histological results. *Med Sci Monit*. 2016;22:3886-3893. [CrossRef]
22. Kamal RM, Helal MH, Mansour SM, et al. Can we apply the MRI BI-RADS lexicon morphology descriptors on contrast-enhanced spectral mammography? *Br J Radiol*. 2016;89(1064):20160157. [CrossRef]
23. Kim JY, Kim JJ, Hwangbo L, et al. Kinetic heterogeneity of breast cancer determined using computer-aided diagnosis of preoperative MRI scans: relationship to distant metastasis-free survival. *Radiology*. 2020;295(3):517-526. [CrossRef]
24. Nam SY, Ko ES, Lim Y, et al. Preoperative dynamic breast magnetic resonance imaging kinetic features using computer-aided diagnosis: association with survival outcome and tumor aggressiveness in patients with invasive breast cancer. *PLoS One*. 2018;13(4):e0195756. [CrossRef]
25. Song SE, Cho KR, Seo BK, Woo OH, Jung SP, Sung DJ. Kinetic features of invasive breast cancers on computer-aided diagnosis using 3T MRI data: correlation with clinical and pathologic prognostic factors. *Korean J Radiol*. 2019;20(3):411-421. [CrossRef]
26. Jong RA, Yaffe MJ, Skarpathiotakis M, et al. Contrast-enhanced digital mammography: initial clinical experience. *Radiology*. 2003;228(3):842-850. [CrossRef]

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2023.232129

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE

Role of interventional radiology in the management of iatrogenic urinary tract injury: the factors affecting the outcome

Selin Ardalı Düzgün[1]
Emre Ünal[1]
Türkmen Turan Çiftçi[1]
Ebru Öztürk[2]
Okan Akhan[1]
Devrim Akıncı[1]

[1]Hacettepe University Faculty of Medicine, Department of Radiology, Ankara, Türkiye

[2]Hacettepe University Faculty of Medicine, Department of Bioistatistics, Ankara, Türkiye

PURPOSE

To evaluate the efficacy of interventional radiological (IR) procedures in iatrogenic urinary tract injury and investigate the factors affecting the outcome.

METHODS

Fifty-eight patients (21 male) with a mean age of 50.3 ± 15.8 years referred for iatrogenic urinary tract injury were enrolled in this study. Technical success was defined as (i) successful placement of a nephrostomy catheter within the renal pelvis and/or (ii) successful antegrade ureteral stent placement (double J stent) between the renal pelvis and bladder lumen. Complete resolution was defined as maintained ureteral patency without an external drain and ureteral stent. The factors that may affect complete resolution [ureteral avulsion, ureterovaginal fistula (UVF), history of malignancy/radiotherapy, and time to IR management] were also investigated. The receiver operating characteristic analysis was performed to estimate the cut-off time point for the IR management timing affecting complete resolution.

RESULTS

The technical success rate for nephrostomy and ureteral stent placement was 100% (n = 58/58) and 78% (n = 28/36), respectively. In 14 patients, non-dilated pelvicalyceal systems were evident. In 18 patients, no further intervention after percutaneous nephrostomy was performed due to (i) poor performance status (n = 6) and (ii) reconstruction surgery upon clinicians' and/or patients' request (n = 12). Reconstruction surgery was required in 11 of the remaining 40 patients due to failure of percutaneous treatment (n = 11/40, 27.5%). In six of the patients, ureteral stents could not be removed due to the development of benign ureteral strictures (n = 6/40, 15%). Our complete resolution rate was 57.5% (n = 23/40). Age, gender, type of surgery (endoscopic or open), side and location of the injury did not statistically affect the complete resolution rate. The presence of ureteral avulsion, history of malignancy and radiotherapy individually or in combination significantly affected the complete resolution rate negatively. The presence of UVF also had a negative effect on the complete resolution rate; however, it did not reach statistical significance. Delayed intervention was also a significant factor related to lower complete resolution. The optimal cut-off point of the time interval for favorable clinical outcome was found to be 0–19th day following the surgery.

CONCLUSION

IR procedures are safe and effective in the management of iatrogenic urinary tract injuries. Antegrade ureteral stenting should be performed as soon as possible to establish ureteral integrity without the development of stricture.

KEYWORDS

Iatrogenic injury, urinary tract, urinary leak, percutaneous nephrostomy, ureteral stent

Corresponding author: Selin Ardalı Düzgün

E-mail: selin.ardali@gmail.com

Received 01 February 2023; revision requested 22 February 2023; last revision received 05 April 2023; accepted 08 May 2023.

Epub: 05.06.2023

Publication date: 08.07.2024

DOI: 10.4274/dir.2023.232129

You may cite this article as: Ardalı Düzgün S, Ünal E, Çiftçi TT, Öztürk E, Akhan O, Akıncı D. Role of interventional radiology in the management of iatrogenic urinary tract injury: the factors affecting the outcome. *Diagn Interv Radiol.* 2024;30(4):256-261.

Iatrogenic urinary tract injuries can be encountered following various abdominopelvic surgeries. Patients may present with fever, abdominal pain, and sepsis. Delayed diagnosis, particularly in asymptomatic patients, can lead to stricture, ureterovaginal fistula (UVF), or kidney failure.[1,2] Intraoperative detection is relatively rare, but it allows for immediate repair. The majority of cases are identified in the post-operative period, and delayed diagnosis is related to lower treatment success.[2-4]

The management of urinary tract injuries may vary depending on the location, severity, and recognition time of the injury.[5] Minimally invasive procedures are the commonly preferred methods of treatment due to the associated lower morbidity/mortality rates and shorter hospital stays.[3,6] Lask et al.[7] reported shorter hospital stay following interventional radiological (IR) procedures (3–5 days) compared with reconstructive surgery (14–35 days).

The European Association of Urology guideline on iatrogenic urinary trauma recommends initial urinary diversion via percutaneous nephrostomy.[8] Urinary diversion by percutaneous nephrostomy may serve as a bridging therapy prior to surgery or can be the definitive treatment. Although Lask et al.[7] reported a complete recovery rate of 80% with percutaneous nephrostomy, Borkowski et al.[9] reported a recovery rate of 28.6% in patients treated with percutaneous nephrostomy alone. Therefore, ureteral stent placement should be performed following nephrostomy to preserve ureteral integrity.[8,10]

This study aims to (i) investigate the efficacy of IR management in iatrogenic urinary tract injury and (ii) find out the factors affecting the outcome.

Methods

This retrospective study was approved by Hacettepe University, Faculty of Medicine Institutional Review Board (GO15/533-27). Informed consent for each procedure was provided by all patients.

Main points

- Ureteral avulsion, ureterovaginal fistula, history of malignancy and radiotherapy individually or in combination negatively affected interventional radiological treatment success.
- Delayed intervention was a significant factor related to a lower complete resolution rate.
- The optimal cut-off point of the time interval for favorable clinical outcome was found to be 0-19 day following the surgery.

Study population

Fifty-eight patients referred to our unit due to iatrogenic urinary tract injury over an 11-year period were enrolled in this study. The diagnosis of iatrogenic urinary tract injury was made by (i) contrast-enhanced abdominal computed tomography with a urography phase and (ii) laboratory analysis of samples obtained from intraabdominal collections. The patients' clinical data, laboratory results, and imaging findings were recorded individually. The factors that may affect the complete resolution rate (ureteral avulsion, UVF, malignancy, radiotherapy, and time to IR management) were also evaluated. Complications were classified according to the Society of Interventional Radiology classification system.[11]

Inclusion and exclusion criteria

The inclusion criteria were as follows: the presence of (i) urinary extravasation on cross-sectional imaging, (ii) urine leak via surgically or percutaneously placed drainage tubes (proven by laboratory analysis), or (iii) UVF. The exclusion criteria were as follows: (i) <18 years of age and (ii) urinary leak due to non-iatrogenic incidents.

Definitions

The results of the treatment were evaluated by reviewing the patients' electronic records. Technical success was defined as (i) the successful placement of a nephrostomy catheter within the renal pelvis and/or (ii) successful antegrade ureteral stent placement (double J stent) between the renal pelvis and bladder lumen. The data of the patients who underwent reconstruction surgery, upon clinicians' and/or patients' request, and in whom further management was not considered due to poor performance status were excluded from further analysis.

Complete resolution was defined as maintained ureteral patency without an external drain and ureteral stent. The location of injury was classified as (i) pelvicalyceal system, (ii) ureter, and (iii) bladder. Ureteral avulsion was recognized as complete discontinuity of the ureter.[12] Time to IR management was defined as the time interval between the surgery and percutaneous nephrostomy (n = 40, patients managed with IR procedures alone).

Technique

Routine hemogram, blood biochemistry, and the coagulation profile (international normalized ratio <1.5 and platelet >50,000/mL) were checked before each procedure. All patients received prophylactic broad-spectrum antibiotics (ceftriaxone or ciprofloxacin) prior to the procedure. All procedures were performed in an IR unit under conscious sedation.

Percutaneous nephrostomy

All procedures were performed under ultrasonographic and fluoroscopy guidance while patients were in the prone position. Lower or middle calyceal puncture was performed via an 18G needle in patients with severe hydronephrosis. Following contrast material administration under fluoroscopy, a 0.035-inch guidewire (Amplatz Super Stiff, Boston Scientific, Natick, MA, USA) insertion and tract dilatation were performed. Over the guidewire, a nephrostomy catheter was placed into the renal pelvis. In patients with a non-dilated pelvicalyceal system or mild hydronephrosis, a 21G needle was used for calyceal puncture. Then, the pelvicalyceal system was opacified under fluoroscopy, and a 0.018-inch guide wire was introduced through the renal pelvis, followed by the introducer set (AccuStick, Boston Scientific, USA). Finally, tract dilatation and catheter placement were performed over the 0.035-inch guidewire.

Antegrade ureteral stent placement

Ureteral stent placement was scheduled as a further intervention in a different session following nephrostomy. First, the nephrostomy catheter was removed with the support of a stiff guide wire (Amplatz Super Stiff, Boston Scientific, Natick, MA, USA). Then, a 0.035-inch hydrophilic wire (Terumo, Tokyo, Japan) was delivered through the ureter with the manipulation of a 5F guiding catheter (Imager II Angiographic Catheter, Bern, Boston Scientific, USA). Contrast material was given to reveal the bladder lumen, and the hydrophilic guide wire was exchanged for a stiff guide wire. A 9F vascular sheath was introduced, and a double J stent (8Fr, 20–26 cm, Flexima Ureteral Stent, Boston Scientific, USA) was placed with the support of pushers through the sheath. After obtaining the desired position of the ureteral stents, the nylon threads were removed under fluoroscopic guidance. All patients were evaluated at regular intervals and underwent stent exchange every four months.

Statistical analysis

The data were tested for normal distribution using the Kolmogorov–Smirnov and

Shapiro–Wilk tests. Descriptive statistics were presented as n (%) for categorical variables. If the continuous variables satisfied the normal distribution assumption, they were expressed as mean and standard deviation; otherwise, they were presented as median, first and third quartiles (Q1–Q3) or interquartile range (IQR). Pearson's chi-square test or Fisher's exact test was used to compare the difference association of two groups for categorical variables. For continuous variables, differences between the two groups were compared using the Mann–Whitney U test and t-test based on the normality assumption. The receiver operating characteristic (ROC) analysis was performed to estimate the cut-off time point for the IR management timing affecting complete resolution. The area under the curve (AUC) and 95% confidence interval (CI) were calculated. The optimal cut-off value for the time interval was specified with the maximizing metric in bootstrapped samples using the cutpointr package in R.[13] The maximizing metric is the sum of sensitivity and specificity. Moreover, in order to find the best discrimination point of the time interval for favorable clinical outcomes, the bootstrapped samples were preferred.

A *P* value of less than 0.05 was considered statistically significant. Statistical analyses were performed using IBM SPSS Statistics (version 23, IBM SPSS, Chicago, IL, USA).

Results

A total of 78 patients were referred for urinary diversion due to a urinary leak. In eight patients, the urinary leak occurred as UVF following radiotherapy, and in 12 of the patients, ureteral integrity was disrupted due to malignant ureteral invasion. These patients were excluded from the study. The final study group consisted of 58 patients (21 male, 36.2%) with a mean age of 50.3 ± 15.8 years (Figure 1).

Urinary tract injury was more frequently encountered following abdominal hysterectomy and bilateral salpingo-oophorectomy surgery (n = 23/58, 40%) and ureterorenoscopy (n=16/58, 28%). In men, the most common surgical indication was urinary stone disease (n = 12/21, 57%), and in women, the most common surgical indications were cervical carcinoma (n = 7/37, 19%) (Figure 2) and myoma uteri (n = 7/37, 19%). Patient and injury characteristics are given in detail in Table 1.

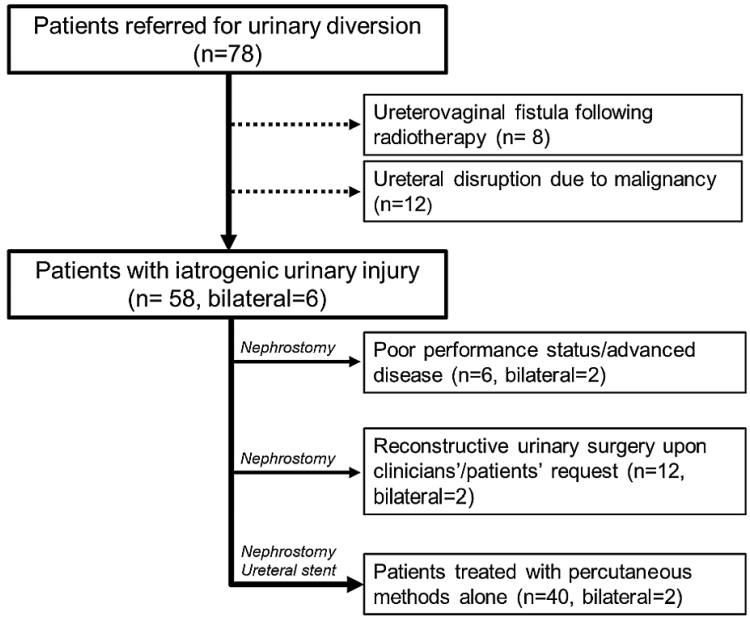


Figure 1. Flowchart of patient selection.

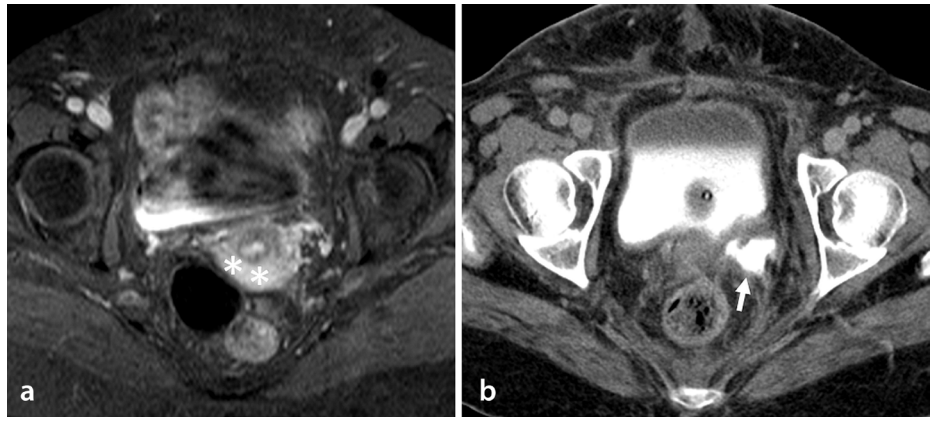


Figure 2. A 55-year-old woman with cervical carcinoma. Axial post-contrast fat-saturated T1 weighted magnetic resonance image shows carcinoma of the posterior cervical wall (a, asterisk). Post-surgical abdominal CT urography demonstrated distal ureteral contrast extravasation (b, arrow). CT, computed tomography.

Table 1. Data of patients with iatrogenic urinary tract injury (n = 58)

		n (%)
Gender (F)		37 (63.8)
Injury site	Ureter	43 (74) (n = 6 bilateral)
	Renal pelvicalyceal	11 (19)
	Bladder	4 (7)
Renal pelvicalyceal injury side	Right	6 (55)
	Left	5 (45)
Ureteric injury side	Right	16 (37)
	Left	21 (49)
	Bilateral	6 (14)
Ureteric injury localization	Proximal	7 (14)
	Middle	8 (16)
	Distal	34 (69)
Hydronephrosis		44 (76)
Pyonephrosis		13 (22)

In all cases, the initially performed procedure was percutaneous nephrostomy. A total of 64 percutaneous nephrostomy procedures were performed in 58 patients (bilateral: 6) with a technical success rate of 100%. In 12 patients (n = 12/58, 21%), percutaneous urinoma drainage was also necessary. Thirty-six patients had indwelling surgically placed drainage tubes. Further management was not considered in 6 patients (n = 6/58, 10%) (bilateral: 2) due to poor performance status and associated comorbid diseases, and these patients opted for permanent nephrostomy. In 21% of the patients (n = 12/58), surgeons performed reconstruction surgery following nephrostomy upon clinicians' and/or patients' request (mean 49.5 ± 33.3 days after nephrostomy). In these 12 patients, no further IR management after percutaneous nephrostomy was performed.

Finally, a total of 40 patients (bilateral: 2) were managed with IR procedures alone. Four out of these 40 patients (10%) were treated with nephrostomy, and no further intervention was required. In 36 patients (n = 36/40, 90%) (bilateral: 2), antegrade ureteral stent placement was attempted after a median of 12 days following nephrostomy (IQR: 18, range: 4–66 days). In 8 patients (n = 8/40, 20%) (bilateral: 1), antegrade stent placement could not be achieved due to the lack of ureteral continuity. The technical success rate for ureteral stent placement was 78% (n = 28/36). Percutaneous balloon dilatation was necessary in 5 patients due to associated benign ureteral stricture (n = 5/28, 18%). No major complications occurred during any of the procedures.[11]

In 68% (n = 19/28) of the patients with ureteral stents, the stents were removed after a median of 110.5 days (IQR: 149, range: 40–701) (Figure 3). The complete resolution rate was 57.5% [n = 23/40, (nephrostomy alone n = 4, ureteral stent n = 19)]. Six patients (n = 6/40, 15%) (bilateral: 1) are still under treatment with routine ureteral stent exchanges due to associated benign ureteral strictures. Eleven patients (n = 11/40, 27.5%) opted for nephrostomy due to (i) ureteral avulsion (n = 8, bilateral: 1) or (ii) refractory urinary leak despite functioning ureteral stent (n = 3). These patients underwent surgery due to failure of IR treatment (mean 47.6 ± 48 days). Moreover, the median follow-up period was 765 days (IQR: 1021).

A further analysis was carried out for patients treated with IR methods alone (n = 40). Age, gender, and the side and location of the injury did not statistically affect the complete resolution rate. There was no statistically significant difference regarding the complete resolution rate between endoscopic and open surgery (P = 0.117) (Table 2). However, the presence of ureteral avulsion, history of malignancy and radiotherapy individually or in combination significantly affected the complete resolution rate negatively. The presence of UVF also had a negative effect on the complete resolution rate, but it did not reach statistical significance (Table 3). Complete resolution was achieved in 25% (n = 2/8) of the patients with UVF (Figures 3, 4). In the complete resolution group, 74% (n = 17/23) of the patients had no malignancy. In addition, all patients with a history of radiotherapy (n = 4/40, 10%) opted for nephrostomy or ureteral stent.

The median time between surgery and diagnosis of iatrogenic injury by cross-sectional imaging was 10 days (IQR: 13.5, range: 0–75). After the diagnosis, percutaneous nephrostomy was performed after a median of 2 days (IQR: 2.5, range: 0–6). The median time between surgery and IR management

Table 2. Data of the patients managed with interventional radiological procedures alone (n = 40*)

		Complete resolution n = 23 (57.5%)	Nephrostomy/ ureteral stent n = 17 (42.5%)	*P* value
Gender (F)		12 (52.2)	12 (70.6)	0.240
Age		50 (33-60)	45 (40.5-64)	0.522
Injury site	Ureter	16 (69.5)	16 (94)	0.107
	Renal pelvis	7 (30.5)	1 (6)	
Injury side	Right	8 (35)	10 (59)	0.054
	Left	15 (65)	5 (29.5)	
	Bilateral	0 (0)	2 (12)	
Ureteric injury localization	Proximal	3 (19)	2 (12.5)	0.765
	Middle	3 (19)	2 (12.5)	
	Distal	10 (62)	12 (75)	
Type of surgery (endoscopic)		11 (48)	4 (24)	0.117

*In 18 of the patients, no further intervention after percutaneous nephrostomy was performed due to (i) poor performance status (n = 6) and (ii) reconstruction surgery upon clinicians' and/or patients' request (n = 12).

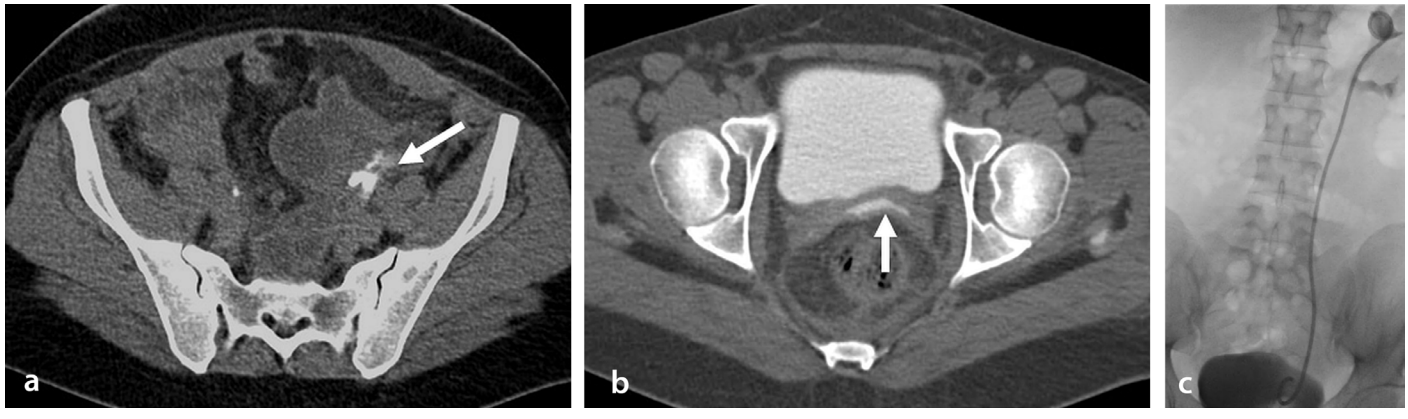


Figure 3. A 33-year-old woman underwent an emergency hysterectomy for postpartum hemorrhage. On follow-up, she developed left flank pain and fever. CT urography demonstrated contrast extravasation (**a**, arrow) and ureterovaginal fistula (**b**, arrow). Percutaneous nephrostomy was initially performed. Four days after nephrostomy, ureteral stent placement was performed **(c)**. During the procedure, ureteral stricture was evident at the level of the pelvic brim (not shown). The ureteral stent was removed on the second exchange period due to the absence of stricture or leak. CT, computed tomography.

was 13 days (IQR: 15, range: 0–78). Time to IR management had also a negative effect on the complete resolution rate. It was shorter in patients with complete recovery than in the remaining patients (a median of 10 days vs. 20 days, $P = 0.018$) (Table 3). According to the ROC analysis, the time to IR management was a significant predictor of clinical outcome (AUC: 0.729, 95% CI: 0.557–0.901, $P = 0.018$) (Figure 5). The optimal cut-off point of the time interval for favorable clinical outcome was found to be 0-19 day following the surgery with respect to the maximizing metric in bootstrapped samples (sensitivity: 0.714, specificity: 0.563).

Discussion

This study evaluated the effectiveness of IR procedures in the management of iatrogenic urinary tract injury. Percutaneous nephrostomy, ureteral stent placement, and collection drainage were the main percutaneous treatment options. The technical success rate for nephrostomy and ureteral stent placement was 100% and 78%, respectively. The lack of ureteral continuity was the major reason for failure of antegrade ureteral stent placement, and the complete resolution rate was 57.5%. The presence of ureteral avulsion, UVF, and history of malignancy and radiotherapy individually or in combination negatively affected the complete resolution rate. In addition, there was a statistically significant negative relation between delayed IR management and the complete resolution rate. The optimal cut-off point of the time interval for favorable clinical outcome was found to be 0-19 day following the surgery. The time prior to IR management was significantly longer in patients who opted for nephrostomy or ureteral stent.

In the management of iatrogenic urinary tract injury, clinical success is based on several conditions: (i) recovery from urosepsis, (ii) preserving renal function, (iii) cessation of urinary leak, and (iv) complete resolution without indwelling nephrostomy and/or ureteral stent. Percutaneous nephrostomy prior to any further management, including surgery, is recommended for urinary decompression and diversion.[6,14] Lask et al.[7] treated 20 patients with percutaneous nephrostomy alone and reported a complete recovery rate of 80%. However, percutaneous nephrostomy per se remains insufficient in most of the cases. Therefore, further management, primarily ureteral stent placement, is mandatory for complete resolution. Borkowski et al.[9] reported a complete recovery of 28.6% (6/21) with percutaneous nephrostomy alone, while this rate was 83% (5/6) for the ureteral stent group. Similarly, our complete resolution rate with percutaneous nephrostomy alone was relatively low (n = 4/12, 30%), while it was 68% (n = 19/28) for ureteral stent. In addition, nephroureteral stents can be used for both urinary diversion and maintaining ureteral patency in patients with urinary tract injury. Zilberman et al.[15] reported a complete resolution rate of 78.5% with nephroureteral stents in a patient population with iatrogenic urinary injury.

Ku et al.[3] reported a complete resolution rate of 65% in 17 patients with urinary leak treated with both antegrade and retrograde ureteral stent placement. Fontana et al.[16] performed ureteral stent placement in 15 patients with urinary leak and reported a complete resolution rate of 53.5%. However, Ustunsoz et al.[17] reported a complete resolution rate of 75% in 22 patients with 24 ureteral injuries. In this study, our complete resolution rate was 57.5%. This may be due to the heterogeneity and complexity of our study population. Ustunsoz et al.[17] reported a higher complete resolution rate in a study consisting of relatively young patients (postpartum urinary injury) without a history of malignancy or radiotherapy. The history of malignancy and/or radiotherapy were significant factors affecting complete resolution in our study. Furthermore, we found complete resolution rates of 71% and 37.5% in patients with benign and malignant diseases, respectively. In a different study, complete resolu-

Table 3. Factors affecting the outcome (n = 40)

	Complete resolution n = 23 (57.5%)	Nephrostomy/ureteral stent n = 17 (42.5%)	*P* value
Benign	17 (74)	7 (41)	0.037
Malignant	6 (26)	10 (59)	
Ureteral avulsion	0 (0)	8 (47)	<0.001
Ureterovaginal fistula	2 (9)	6 (35)	0.053
Radiotherapy	0 (0)	4 (23.5)	0.026
Combination of risk factors	0 (0)	7 (41)	0.001
Time to interventional radiological management (days)	10 (4.50–19.50)	20 (9.50–30.25)	0.018

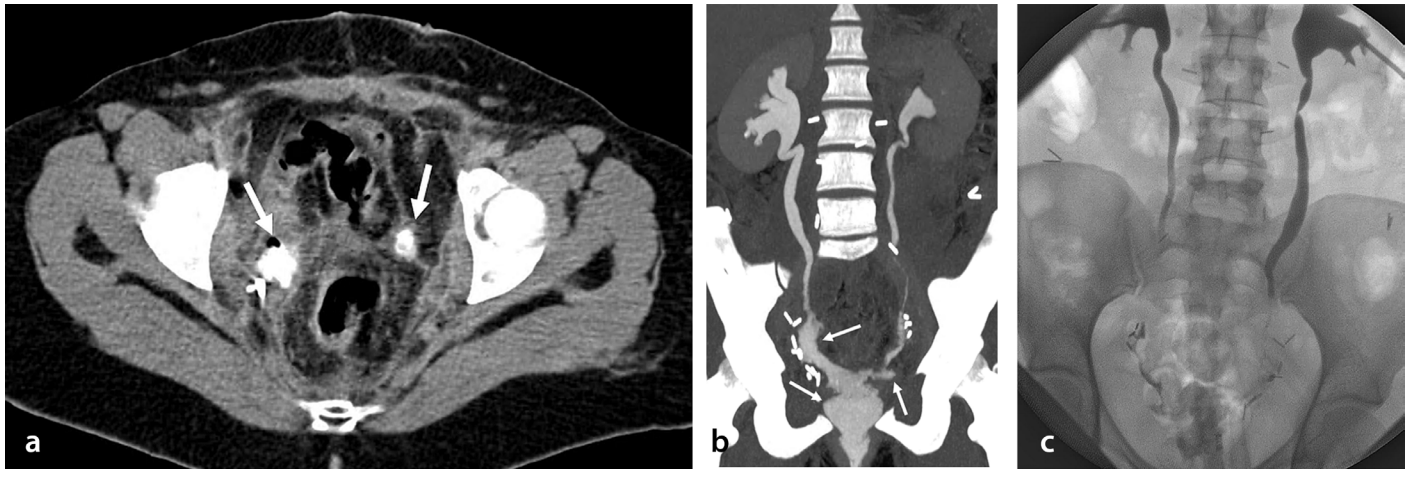


Figure 4. A 40-year-old woman presented with a vaginal urine leak following hysterectomy. CT urography demonstrated distal ureteral contrast extravasation and ureterovaginal fistula (arrows, **a** and **b**). Bilateral nephrostomy was performed **(c)**. The patient underwent bilateral ureteroneocystostomy due to failure of ureteral stent placement. CT, computed tomography.

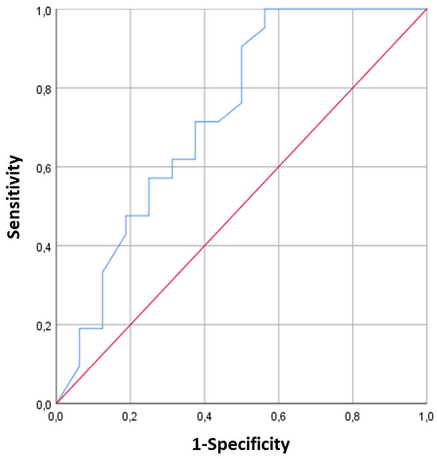


Figure 5. The receiver operating characteristics curve for the time to interventional radiological management (area under the curve: 0.729, 95% confidence interval: 0.557–0.901, $P = 0.018$).

tion was achieved in only 4 out of 19 cancer patients with postoperative ureteral injury.[18]

The presence of ureteral avulsion (loss of integrity) negatively affected the complete resolution rate. In this study, eight patients with ureteral avulsion opted for nephrostomy and ended up with reconstructive surgery. Ustunsoz et al.[17] also reported that 50% of the patients with failure of treatment had ureteral avulsion.

Delayed diagnosis of urinary tract injury is a major factor related to low treatment success.[3,4,9,17] Morrow et al.[19] reported that a longer median time to ureteral stent placement was associated with failure. We also found that time prior to IR management was a significant factor in determining complete resolution. The optimal cut-off point of the time interval for favorable clinical outcome was found to be 0–19th day following the surgery (sensitivity: 0.714, specificity: 0.563).

The presence of UVF also had a negative effect on the complete resolution rate; however, it did not reach statistical significance. This may be because of our small sample size. Chen et al.[20] reported a complete resolution rate of 83% in a series of 12 patients with UVF managed with ureteral stenting. In addition, Rajamaheswari et al.[21] reported successful ureteral stenting in 77% of patients with UVF. Our relatively low success rate in patients with UVF may be due to delayed intervention. Follow-up with nephrostomy alone is not recommended in the treatment of UVF due to an increased rate of failure.[20,22] Ureteral integrity should be established as soon as possible to avoid a mature fistula tract between the ureter and vagina.[20]

This study has several limitations. First, it is a retrospective study. Second, the study population was heterogeneous. Third, the time to IR management was relatively long; therefore, the complete resolution rate of this study might have been negatively affected. Finally, the sample size was small, and in several patients, antegrade ureteral stent placement could not be attempted due to clinicians' decisions.

In conclusion, IR procedures are safe and effective methods of treatment alternative to reconstruction surgery in the management of postoperatively detected iatrogenic urinary tract injury. Ureteral avulsion, UVF, history of malignancy and radiotherapy, and delayed intervention negatively affect treatment success. Antegrade ureteral stent placement should be performed as soon as possible to establish ureteral integrity without the development of stricture.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Burks FN, Santucci RA. Management of iatrogenic ureteral injury. *Ther Adv Urol*. 2014;6(3):115-124. [CrossRef]
2. Selzman AA, Spirnak JP. Iatrogenic ureteral injuries: a 20-year experience in treating 165 injuries. *J Urol*. 1996;155(3):878-881. [CrossRef]
3. Ku JH, Kim ME, Jeon YS, Lee NK, Park YH. Minimally invasive management of ureteral injuries recognized late after obstetric and gynaecologic surgery. *Injury*. 2003;34(7):480-483. [CrossRef]
4. Ghali AM, El Malik EM, Ibrahim AI, Ismail G, Rashid M. Ureteric injuries: diagnosis, management, and outcome. *J Trauma*. 1999;46(1):150-158. [CrossRef]
5. Kim JS, Lee DH, Suh HJ. Double-J stenting: initial management of injured ureters recognized late after gynecological surgery. *Int Urogynecol J*. 2010;21(6):699-703. [CrossRef]
6. Koukouras D, Petsas T, Liatsikos E, et al. Percutaneous minimally invasive management of iatrogenic ureteral injuries. *J Endourol*. 2010;24(12):1921-1927. [CrossRef]
7. Lask D, Abarbanel J, Luttwak Z, Manes A, Mukamel E. Changing trends in the management of iatrogenic ureteral injuries. *J Urol*. 1995;154(5):1693-1695. [CrossRef]
8. Summerton DJ, Kitrey ND, Lumen N, Serafetinidis E, Djakovic N; European Association of Urology. EAU guidelines on iatrogenic trauma. *Eur Urol*. 2012;62(4):628-639. [CrossRef]
9. Borkowski T, Judycki J, Borkowski A, Czaplicki M, Radziszewski P. Evolution in the approach to overlooked ureteral injuries after gynecological surgery. *Ginekol Pol*. 2016;87(10):690-696. [CrossRef]
10. Brandes S, Coburn M, Armenakas N, McAninch J. Diagnosis and management of ureteric injury: an evidence-based analysis. *BJU Int*. 2004;94(3):277-289. [CrossRef]
11. Khalilzadeh O, Baerlocher MO, Shyn PB, et al. Proposal of a new adverse event classification by the Society of Interventional Radiology Standards of Practice Committee. *J Vasc Interv Radiol*. 2017;28(10):1432-1437. Erratum in: *J Vasc Interv Radiol*. 2018;29(1):146. [CrossRef]
12. Tang K, Sun F, Tian Y, Zhao Y. Management of full-length complete ureteral avulsion. *Int Braz J Urol*. 2016;42(1):160-164. [CrossRef]
13. Thiele C, Hirschfeld G. Cutpointr: improved estimation and validation of optimal cutpoints in R. *J Stat Softw*. 2021;98:1-27. [CrossRef]
14. Narasimham DL, Jacobsson B, Nyman U, Vijayan P. Primary double pigtail stenting as a treatment of upper urinary tract leaks. *J Urol*. 1990;143(2):234-236. [CrossRef]
15. Zilberman DE, Rimon U, Morag R, Winkler HZ, Ramon J, Mor Y. Non-surgical treatment of latrogenic postoperatively diagnosed ureteral injuries. *Isr Med Assoc J*. 2015;17(4):227-230. [CrossRef]
16. Fontana F, Piacentino F, Ossola C, et al. Diagnostic and ınterventional radiology management of ureteral ıatrogenic leakage after gynecologic surgery. *Diagnostics* (Basel). 2021;11(5):750. [CrossRef]
17. Ustunsoz B, Ugurel S, Duru NK, Ozgok Y, Ustunsoz A. Percutaneous management of ureteral injuries that are diagnosed late after cesarean section. *Korean J Radiol*. 2008;9(4):348-353. [CrossRef]
18. Kawada H, Sato Y, Inaba Y, et al. Stenting using the rendezvous technique for postoperative ureteral complications in cancer patients. *Cardiovasc Intervent Radiol*. 2020;43(10):1486-1491. [CrossRef]
19. Morrow J, Curry D, Dooher M, Woolsey S. Minimally invasive management of delayed recognition iatrogenic ureteric injury. *Ulster Med J*. 2017;86(3):181-184. [CrossRef]
20. Chen YB, Wolff BJ, Kenton KS, Mueller ER. Approach to ureterovaginal fistula: examining 13 years of experience. *Female Pelvic Med Reconstr Surg*. 2019;25(2):e7-11. [CrossRef]
21. Rajamaheswari N, Chhikara AB, Seethalakshmi K. Management of ureterovaginal fistulae: an audit. *Int Urogynecol J*. 2013;24(6):959-962. [CrossRef]
22. Dowling RA, Corriere JN Jr, Sandler CM. Iatrogenic ureteral injury. *J Urol*. 1986;135:912-915. [CrossRef]

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2023.221424

INTERVENTIONAL RADIOLOGY

ORIGINAL ARTICLE

The temporal and spatial relationship between percutaneous vertebral augmentation and new symptomatic fractures

Jing Tang[1*]
Jin Liu[2*]
Zuchao Gu[3]
Yu Zhang[3]
Haosen Yang[2]
Zhenlin Li[1]

[1]Sichuan University West China Hospital, Department of Radiology, Chengdu, China

[2]Chengdu Seventh People's Hospital, Department of Orthopaedics, Chengdu, China

[3]Chengdu First People's Hospital, Department of Orthopaedics, Chengdu, China

PURPOSE

This study aimed to explore the relationship between the time from percutaneous vertebral augmentation (PVA) until subsequent fracture and the risk of new symptomatic fractures (NSFs) in untreated vertebrae at different distances from "augmented vertebrae".

METHODS

Patients who underwent PVA for the treatment of osteoporotic vertebral compression fractures at the West China Hospital of Sichuan University from May 2014 to April 2019 were retrospectively recruited. Vertebrae not treated during PVA were stratified based on their distance from the nearest augmented vertebra and the time elapsed since PVA. Survival curves were plotted to compare the risk of NSFs in untreated vertebrae at different distances from augmented vertebrae. The Cox proportional hazards model was used to identify risk factors of NSFs in untreated vertebrae.

RESULTS

In total, 162 patients with 228 NSFs (2.760 vertebrae) were analyzed. More than half of the NSFs (56.6%) occurred within the first year after PVA. Rates and hazard ratios (HRs) of NSFs were higher in vertebrae located one segment away from the augmented vertebrae (21.0%, HR: 3.99, $P < 0.001$), two segments away (10.6%, HR: 1.97, $P = 0.003$), or three segments away (10.5%, HR: 2.26, $P < 0.001$) than in vertebrae located five or more segments away (3.81%, HR: 1.00). Similar results were observed regardless of whether the untreated vertebrae were located in the thoracolumbar junction. In addition to distance, other risk factors of NSFs were the thoracolumbar location of untreated vertebrae, the number of augmented vertebrae, and percutaneous vertebroplasty.

CONCLUSION

The risk of NSFs is greater for untreated vertebrae located closer to augmented vertebrae than for untreated vertebrae further away. This distance dependence occurs mainly within the three segments closest to the augmented vertebra. The risk of NSFs decreases with time after augmentation, and it is also related to the number of augmented vertebrae, the type of augmentation, and whether the untreated vertebrae are thoracolumbar or not.

KEYWORDS

Vertebral augmentation, new symptomatic fracture, osteoporotic vertebral compression fracture, percutaneous kyphoplasty, percutaneous vertebroplasty

***Joint first authors**

Corresponding author: Zhenlin Li, Haosen Yang

E-mail: hx_lizhenlin@126.com, 914374012@qq.com

Received 26 June 2022; revision requested 30 August 2022; last revision received 26 May 2023; accepted 17 June 2023.

Epub: 09.08.2023

Publication date: 08.07.2024

DOI: 10.4274/dir.2023.221424

Since the first use of bone cement for the treatment of invasive cervical hemangioma in 1987,[1] percutaneous vertebral augmentation (PVA) has been considered an effective treatment for osteoporotic vertebral compression fractures (OVCFs). This technique, which can involve percutaneous kyphoplasty (PKP) or percutaneous vertebroplasty (PVP), can provide immediate, effective analgesia, as well as quick recovery of daily activities.[2-9] The use of PVA can also reduce mortality associated with OVCFs.[10] Despite these benefits, new symptomatic fractures (NSFs) after augmentation remain a vexing problem.

You may cite this article as: Tang J, Liu J, Gu Z, Zhang Y, Yang H, Li Z. The temporal and spatial relationship between percutaneous vertebral augmentation and new symptomatic fractures. *Diagn Interv Radiol*. 2024;30(4):262-269.

The American Society for Bone and Mineral Research has developed guidelines on how to prevent secondary fractures for patients with osteoporotic fractures.[11] However, these guidelines should be adapted for patients undergoing PVA, given the potential effects of bone cement on the biomechanics of the spine and subsequent NSFs after PVA.[12-14] Therefore, it is necessary to study the occurrence and risk of NSFs after PVA. At the patient level, several studies have shown that the occurrence of NSFs after augmentation depends on bone mineral density (BMD), age, and several other factors.[15-21] Vertebra-level studies of patients with OVCF who underwent PVA have suggested that adjacent vertebrae, especially the sandwich vertebrae, are more prone to NSFs.[17,18] However, the risk of NSFs in different untreated vertebrae within the same patient, and the factors that affect that risk, are still unknown.

An analysis of the time course and location of NSFs after PVA can provide further information about this risk. Therefore, the present study compared the risk of post-PVA fracture in untreated vertebrae at different distances from augmented vertebrae using vertebra-level survival analysis. It also analyzed different potential risk factors using the Cox proportional hazards model.

Methods

Eligibility criteria

This retrospective study was approved by the Ethics Committee of West China Hospital of Sichuan University (approval number: 2019-992). The requirement for informed consent was waived because, at the time of surgery, patients gave written consent for their anonymized medical data to be analyzed and published for research purposes.

Main points

- The risk of new fractures is greater for untreated vertebrae nearest to the augmented vertebra.
- This distance dependence occurs mainly within the three segments closest to the treated vertebra, and the risk of new fractures decreases with time since augmentation.
- Distance from the treated vertebrae, thoracolumbar location, percutaneous vertebroplasty, and higher number of treated vertebrae were identified as risk factors for new symptomatic fractures after percutaneous vertebral augmentation.

This study retrospectively examined clinical and imaging data from patients with OVCFs who underwent PVA at our institution between May 2014 and April 2019. The OVCFs were defined as vertebral compression fractures without obvious cause or those caused by low-energy injury. Patients with OVCFs were included if they were (a) ≥70 years old or had a dual-energy X-ray absorptiometry (DXA) BMD T-score ≤-2.5 (when BMD was measured) and (b) complained of recurrent pain associated with NSFs. These NSFs were confirmed using magnetic resonance imaging (MRI). Patients were followed up for at least 12 months after PVA.

Patients who (a) received PVA because of pathological fractures caused by spinal neoplasms, (b) had a history of PVA at other hospitals for whom relevant data was inaccessible, (c) experienced new fractures caused by high-energy trauma during follow-up, or (d) had a history of thoracic or lumbar internal fixation were excluded.

Surgical procedures

All PVA procedures were performed after MRI had confirmed acute OVCF. The purpose of the PVA and the surgical procedures involved were explained to the patients in detail. All PKP and PVP procedures were performed based on standard guidelines.[18,21] The same bone cement (Osteopal V, Germany) was used in all procedures.

Postoperative treatment and follow-up care

After surgery, patients rested in the supine position for three hours and gradually resumed activities out of bed. Routine anteroposterior and lateral X-ray examinations were performed to assess the distribution and leakage of bone cement after procedures. After surgery, all patients were routinely given calcium (800 mg daily) and active vitamin D (0.5 μg daily), which they were told to continue indefinitely. A total of 28 patients also opted for zoledronic acid therapy at the time of the initial fracture.

After discharge, patients were followed up with via telephone every three months to enquire about pain levels and daily activities. Patients complaining of back or lower back pain that lasted longer than three days or those who did not experience significant relief after taking non-steroidal anti-inflammatory drugs were requested to come to the hospital for an X-ray examination. In the case of a suspected NSF, MRI was performed.

Assessment indices

Baseline data on sex, age, BMD, body mass index (BMI), augmentation method (PKP or PVP), puncture method (unilateral or bilateral), and cause of fractures, were collected, as well as imaging data from X-ray and MRI examinations. All radiographic results were independently evaluated by a spine surgeon with 11 years of experience and a radiologist with seven years of experience in musculoskeletal system imaging. If there was a dispute, a radiology professor with more than 30 years of experience in musculoskeletal system imaging was consulted for the final evaluation. The inter-observer correlation coefficient (ICC) was excellent (ICC: 0.84, $P < 0.001$).

Fracture data were also collected, including the number of fractures; location of fractures [thoracolumbar (T11–L2) or non-thoracolumbar], which was defined as the region with the higher number of OVCFs in patients with multiple OVCFs; degree of compression of fractures, which was defined as the worst degree in patients with multiple OVCFs; kyphosis angle between the upper and lower endplates of the fractured vertebra, which was defined as the greatest angle in patients with multiple OVCFs; cleft signs in OVCF vertebrae; distribution of bone cement; intradiscal cement leakage; and number of vertebrae treated.

The distribution of bone cement was evaluated using the 12-score method (Figure 1).[21] The distance between untreated vertebrae in the T4–L5 segment and the nearest treated vertebra was measured. The date of surgery at the nearest treated vertebra served as a start time for each untreated vertebra. When the distances between the untreated vertebrae and two separate treated vertebrae were equal, the date of the most recent surgery was considered as the start time. The time of diagnosis served as the end point for calculating survival time of newly fractured vertebrae, while the end of follow-up was the end point for calculating survival time of unfractured vertebrae.

Based on the interval between the occurrence of NSF and the last augmentation procedure, patients were stratified into those who suffered early (within 3 months), mid-term (3–12 months), or late NSFs (>12 months). The interval between the occurrence of NSF and the final augmentation procedure was defined as the time from the most recent PVA until the definitive diagnosis of NSF.

Statistical analysis

Statistical analyses were conducted using SPSS version 23.0 (IBM, Armonk, New York, USA). Normally distributed continuous data were presented as mean ± standard deviation. Enumeration data were reported as median (minimum–maximum). Categorical data were expressed as frequencies with percentages. Where appropriate, results were reported as hazard ratio (HR), along with 95% confidence intervals.

Differences in normally distributed continuous data were assessed for significance using One-Way analysis of variance and pairwise comparisons using the least significant difference (LSD) test. Skewed data were assessed using the Kruskal–Wallis test for significance assessment, while the Wilcoxon rank–sum test was used for pairwise comparisons. Differences in categorical data were assessed using chi-squared tests, and pairwise comparisons were conducted using chi-squared tests with Bonferroni correction.[22] The Bonferroni correction compensated by raising the test standard

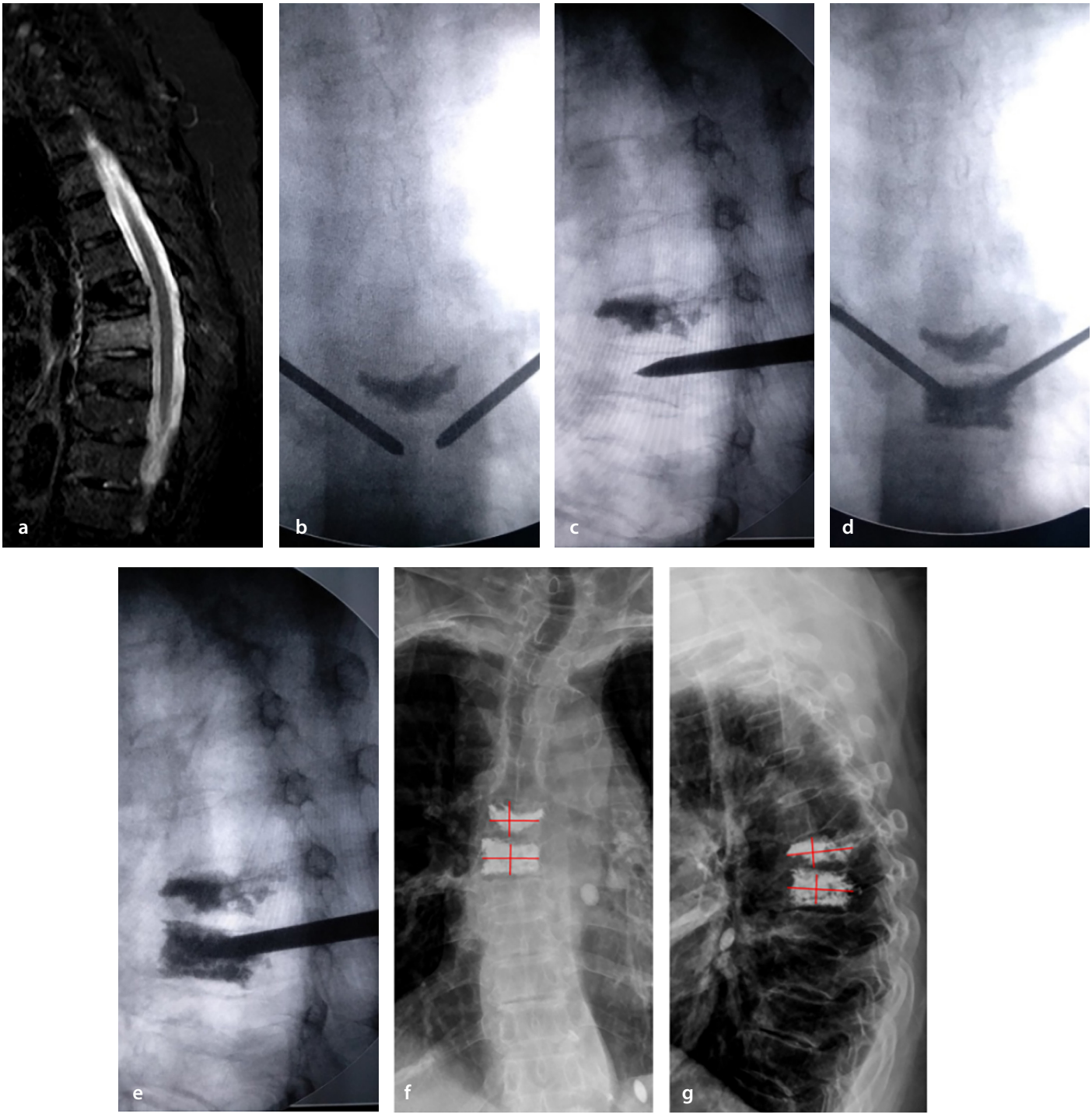


Figure 1. An 81-year-old female was treated with percutaneous vertebroplasty (PVP) at T6 in our hospital. She was admitted to the hospital because of back pain for the preceding 24 hours. Magnetic resonance imaging showed fresh compression fracture of T7, which was treated by repeat PVP. **(a)** New compression fracture at vertebra T7. **(b, c)** Intraoperative fluoroscopic X-ray images of puncture. **(d, e)** Intraoperative fluoroscopic X-ray images of bone cement injection. **(f, g)** Anteroposterior and lateral X-ray images after repeat PVP. The red line shows quadrants for evaluating cement distribution. Cement was distributed across nine quadrants at vertebra T6 and across 12 quadrants at vertebra T7.

for each individual hypothesis at the level of significance. Survival curves were drawn using the Kaplan–Meier method and compared using the log-rank test. The Cox proportional hazards model was used to identify risk factors of NSFs. Level of significance was taken as α: 0.050.

Results

Study population

Among the 1.280 patients with OVCFs who were treated with PVA in our hospital, 190 (14.8%) suffered NSFs. In the end, 228 NSFs in 162 patients met the eligibility criteria; in these patients, 432 vertebrae had been augmented. A total of 2.760 non-treated T4–L5 vertebrae before NSFs were analyzed, of which 273 were affected by the 228 NSFs. The median number of NSFs was similar between patients who took bisphosphate at the time of initial fracture [1.19 (1–3)] and those who did not [1.35 (1–4); $P = 0.131$].

Characteristics of new symptomatic fractures after percutaneous vertebral augmentation

The median follow-up time for all patients enrolled in the study was 39.8 (12.7–71.6) months, and the median time until occurrence of NSFs was 11.4 (0.2–66.0) months. Most new fractures (56.6%, 129/228) occurred within the first year after PVA, while 21.0% (48/228) occurred in the second year, 13.2% (30/228) in the third year, 5.26% (12/228) in the fourth year, and 3.95% (9/228) after the fourth year.

The NSFs were also stratified into those occurring early (<3 months after PVA; n = 79), in the mid-term (3–12 months after PVA; n = 50), or late (>12 months after PVA; n = 99) (Table 1). In the patients who underwent DXA testing, there was no overall significant difference between the three groups ($P = 0.823$). No significant differences were found in the average BMD T-score between the three groups ($P = 0.099$), but the pairwise comparison found that BMD was lower for patients who suffered early NSFs than for those who suffered late ones ($P = 0.033$). No significant differences were found between the patients who suffered mid-term NSFs and those who suffered early or late ones ($P = 0.280$, $P = 0.475$).

The average age of the three groups was different ($P = 0.044$). Pairwise comparisons using the LSD method revealed no significant difference in age between the early and mid-term groups ($P = 0.725$), but there was a significantly younger age in the late group than in the early group ($P = 0.042$) and mid-term group ($P = 0.033$; Appendix Table 1). Significant differences were also observed between the three groups in kyphosis angle, thoracolumbar location, number of treated vertebrae, and augmentation method (PKP or PVP) ($P \leq 0.037$) (Table 1). However, there were no significant differences between the three groups in BMI, sex, cause of fractures, cleft sign, puncture method (unilateral or bilateral), bone cement distribution, intradiscal cement leakage, and degree of compression ($P > 0.05$).

New symptomatic fractures at different distances from the nearest treated vertebra

Significant differences in cumulative NSF rates among vertebrae that were one, two, three, four, or five or more segments away from the nearest treated vertebra were observed ($P < 0.001$; Table 2). Similar results were observed regardless of whether the untreated vertebrae were located in the thoracolumbar junction ($P < 0.001$; Table 2) or not ($P < 0.001$; Table 2). Pairwise comparisons

Table 1. Baseline demographic and clinical characteristics of patients with new symptomatic fractures, stratified based on time since the last percutaneous vertebral augmentation

Characteristic	Timing of fracture since last surgical procedure			
	Early	Mid-term	Late	*P*
New symptomatic fractures	79	50	99	-
Age (years)	76.64 ± 8.33	77.10 ± 7.38	74.22 ± 7.44	**0.044**
Female	67 (84.8)	40 (80.0)	82 (82.8)	0.779
Body mass index (kg/m²)	21.98 ± 3.39	21.89 ± 3.53	22.02 ± 3.14	0.977
Distribution of bone cement (no of quadrants)	10.69 ± 1.33	10.46 ± 1.53	10.27 ± 1.57	0.176
T-score of bone mineral density*	-3.47 ± 0 .46	-3.39 ± 0.59	-3.26 ± 0.42	0.099
No of treated vertebrae	2 (1-6)	2 (1-6)	1 (1-5)	**<0.001**
DXA results available	42 (53.2)	24 (48.0)	49 (49.5)	0.823
Presence of clefts	17 (21.5)	5 (10.0)	13 (13.1)	0.150
Cause of injury known	18 (22.8)	15 (30.0)	39 (39.4)	0.058
Degree of compression of previous fractures ≥50%	21 (26.6)	9 (18.0)	17 (17.2)	0.266
Kyphosis angle of previous fractures ≥10°	49 (62.0)	20 (40.0)	58 (58.6)	**0.037**
Thoracolumbar	43 (54.4)	23 (46.0)	69 (69.7)	**0.012**
PKP	14 (17.7)	16 (32.0)	53 (53.5)	**<0.001**
Unilateral puncture	20 (25.3)	10 (20.0)	20 (20.2)	0.667
Intervertebral leakage	11 (13.9)	6 (12.0)	10 (10.1)	0.735

*For patients with new symptomatic fractures for whom data on bone mineral density were available. Values indicated as n, n (%), median (minimum–maximum) or mean ± standard deviation, unless otherwise noted. DXA, dual-energy X-ray absorptiometry; PKP, percutaneous kyphoplasty.

Table 2. Incidence of new symptomatic fractures inside or outside the thoracolumbar region, stratified by distance from the nearest treated vertebra

Location of fracture#	No of segments away from treated vertebra*					*P*
	One	Two	Three	Four	Five or more	
Within the thoracolumbar region	87 (211)	23 (151)	15 (85)	11 (47)	7 (53)	**<0.001**
Outside the thoracolumbar region	38 (232)	23 (236)	25 (254)	11 (260)	33 (958)	**<0.001**
Total	**125 (443)**	**46 (387)**	**40 (339)**	**22 (307)**	**40 (1011)**	**<0.001**

*The data in the table represent the number of vertebrae, including the number of fractured vertebrae outside the brackets and the number of unfractured vertebrae inside the brackets, sample size (n) = fractured + unfractured vertebrae. #All the vertebrae located in T4-L5 that were studied were divided into two types: those located in the thoracolumbar (T11-L2) region and those not located in the thoracolumbar region (T4-T10, L3-L5).

showed that rates of NSFs affecting vertebrae that were one (21.0%), two (10.6%), or three (10.5%) segments away were significantly higher than rates of NSFs affecting vertebrae that were five or more segments away (3.81%; $P < 0.001$; Table 3). There was no difference in rates of NSFs affecting vertebrae that were four or five or more segments away ($P = 0.028$, Bonferroni correction method).

Risk curves of NSFs were plotted for varying distances from the nearest treated vertebrae using the Kaplan–Meier method (Figure 2), and a log-rank test confirmed significant differences between them ($P < 0.001$). The Cox proportional hazards model and forward stepwise method based on the conditional likelihood ratio and several categorical variables (Appendix Table 2) were used to investigate risk factors associated with NSFs occurring in 2.760 untreated vertebrae in segments T4–L5. An omnibus test of the Cox model coefficient was statistically significant ($P < 0.001$). The following variables were identified as significantly related to the occurrence of NSFs: distance from the nearest treated vertebrae ≤3, thoracolumbar junction of untreated vertebrae, higher number of treated vertebrae, and method of vertebral augmentation (PVP) ($P < 0.001$).

Table 3. Pairwise comparison of rates of new symptomatic fractures in segments at different distances from the treated vertebra

	One segment away	Two segments away	Three segments away	Four segments away	Five or more segments away
One segment away	-	-	-	-	-
Two segments away	<0.001	-	-	-	-
Three segments away	<0.001	0.974		-	-
Four segments away	<0.001	0.059	0.069	-	-
Five or more segments away	<0.001	<0.001	<0.001	0.028	-

The adjusted *P* value according to Bonferroni correction: α: 0.05/10 = 0.005.

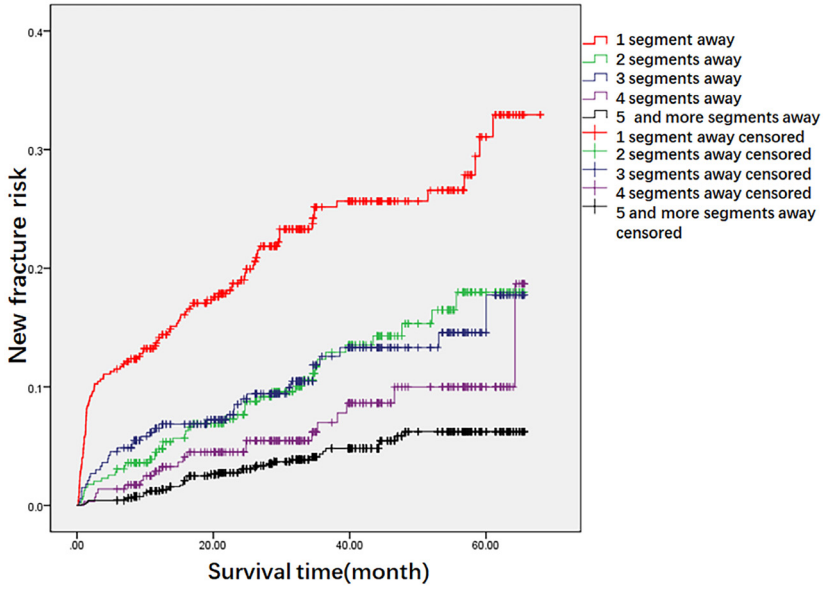


Figure 2. Curves showing the risk of new symptomatic fractures (NSFs) at one, two, three, four or five or more segments away from the nearest treated vertebra. Data were obtained from 228 NSFs affecting 2.760 untreated vertebrae located in segments T4–L5.

Table 4. Cox proportional hazards modeling to identify risk factors for the occurrence of new symptomatic fractures

Factor*	B (regression coefficient)	SE	*P* value	HR	95% CI	
					Lower	Upper
Distance from nearest treated vertebra	-	-	**<0.001**	-	-	-
One segment away	1.384	0.201	**<0.001**	3.989	2.689	5.916
Two segments away	0.676	0.227	**0.003**	1.967	1.260	3.071
Three segments away	0.814	0.228	**<0.001**	2.258	1.444	3.532
Four segments away	0.438	0.267	0.100	1.550	0.919	2.614
Whether thoracolumbar or not	0.842	0.136	**<0.001**	2.322	1.780	3.028
No of augmented vertebrae	0.109	0.054	**0.043**	1.115	1.004	1.239
Augmentation method (PVP vs PKP)	0.405	0.129	**0.002**	1.499	1.164	1.930

*Definitions of categorical variables: distance from the nearest augmented vertebra, 1: one segment away, 2: two segments away, 3: three segments away, 4: four segments away; location of the untreated vertebra, 0: thoracolumbar segment, 1: non-thoracolumbar segment; augmentation method, 1: PVP, 2: PKP. CI, confidence interval; HR, hazard ratio; PVP, percutaneous vertebroplasty; PKP, percutaneous kyphoplasty; SE, standard error.

The HR of NSFs was higher at one (HR: 3.99, $P < 0.001$), two (HR: 1.97, $P = 0.003$), and three (HR: 2.26, $P < 0.001$) segments away from the nearest treated vertebrae than at five or more segments away (Table 4). The HRs were similar for NSFs four or five or more segments away ($P = 0.100$). The HR for NSFs was 1.499 times higher among patients who underwent PVP than among those who underwent PKP ($P = 0.002$). The HR of NSFs in the untreated vertebrae of the thoracolumbar junction was 2.322 times higher than in those of the non-thoracolumbar junction ($P < 0.001$), and it was 1.115 times higher for each increase in the number of treated vertebrae ($P = 0.043$).

Discussion

A PVA is a minimally invasive procedure that has been shown to significantly benefit patients with acute pain caused by OVCF,[2-9] so it has become the main surgical method for treating OVCF. However, new fractures after augmentation remain a substantial concern. In this study, 228 NSFs at the vertebra level in 162 patients were retrospectively analyzed, and the risks of fractures in untreated vertebrae at different distances from augmented vertebrae were compared. The results of the study suggest that the risk of NSFs is greater for untreated vertebrae located closer to augmented vertebrae. This distance dependence occurred mainly within the three segments adjacent to the treated vertebra, while the risk of NSFs decreased with time since augmentation. Closer proximity to the treated vertebrae, thoracolumbar location, PVP as the method of vertebral augmentation, and a higher number of treated vertebrae were identified as risk factors for NSFs after PVA.

Several studies have identified risk factors for new vertebral fractures,[18,23-30] but most of them assessed risk at the level of the patient, not individual vertebrae. This approach can help identify patients at high risk of new fractures, but only comparisons of fracture risk at different untreated vertebrae in the same patient can help predict which vertebrae are at increased risk. The patterns identified in the present study may provide clues to clinicians and radiologists about which vertebrae require greater attention after augmentation. Comparisons within the same patient should also be free of confusion caused by factors other than the distance between untreated and treated vertebrae and their location. The present study suggests that the rate of NSFs decreases with increasing distance from the nearest treated vertebra. This result confirms and extends previous vertebra-level analyses.[17,18,24] Adjacent vertebrae are more prone to NSFs, while the "sandwich" vertebrae, defined as the untreated vertebrae either side of the augmented vertebrae, are understandably at highest risk of NSF. While the thoracolumbar junction may be inherently prone to fractures,[18,24,26] the present study found that the distance dependence of the risk of NSFs applies inside and outside this region. In fact, the Cox proportional hazards analysis suggested that proximity to augmented vertebrae may influence risk of NSFs in untreated vertebrae to a greater extent than thoracolumbar location.

The present study also found that augmentation method (PKP or PVP) influenced the risk of NSFs.[9] Although similar numbers of patients underwent each procedure, PVP was associated with a larger proportion of early (82.3%) and mid-term NSFs (68.0%) and a smaller proportion of late NSFs (46.5%). These results are consistent with the idea that PKP is superior to PVP for improving local kyphosis of the fracture: PKP can restore vertebral height, reduce the negative effects of load transfer, and decrease the risk of NSFs;[29,30] PVP, in contrast, may be less effective at counteracting biomechanical changes at the fracture site,[13,14] increasing the risk of NSFs.

Augmentation exerts biomechanical effects on the spine, although it is unclear whether these effects increase the risk of NSFs.[12-14] The risk of NSFs has been previously shown to increase with the number of cement-filled vertebrae,[18] and the present study found similar results, reflecting greater deleterious biomechanical effects with a higher number of augmented vertebrae. A higher number of augmented vertebrae inevitably expands the reach of their influence on untreated vertebrae, and more fractures may also mean lower BMD.

Low BMD is another important risk factor for NSFs,[23-27] and the present study found that patients with lower BMD experienced NSFs earlier than those with higher BMD. At the vertebra level, it was also found that the closer an untreated vertebra was to an augmented vertebra, the more likely it was to suffer an NSF. This difference in the risk of NSFs in the same patient's untreated vertebrae likely reflects the effect of augmentation on NSFs.[22-24] Differences in the risk of NSF when patients, rather than individual vertebrae, are the unit of analysis may be more closely related to BMD.[23]

Previous studies have suggested that the risk of new fractures can depend on puncture method (unilateral or bilateral puncture),[18,31] excess bone cement distribution,[21,26] and intradiscal cement leakage.[24,26] The puncture method can affect the distribution of bone cement, and one study[21] concluded that cement distribution slightly alters the risk of NSFs. This conclusion may not be generalizable, however, given that the average distribution of bone cement in the study was 10.5 points on a 12-point scale, which was already close to the optimal distribution.[21] Just under 12% of patients in the present study showed leakage into the intervertebral space, but this may not be unusually high given that all patients in the study had already suffered NSFs.

The present study had several limitations. Its retrospective nature increased the risk of various types of bias. In addition, BMD data was unavailable for some patients, so it was not possible to include BMD in the Cox proportional hazards analysis. Prospective studies are therefore needed to examine the occurrence of NSFs in different vertebrae.

In conclusion, although a causal relationship between augmentation and subsequent vertebral fractures cannot be conclusively demonstrated, the data in the present study add valuable information to the continuing question of how PVA affects risk of NSFs in vertebrae at different distances from the augmented site. The results of the present study suggest that fracture risk depends on proximity to treated vertebrae and on time since augmentation. This distance dependence appears to hold mainly within the three segments closest to the treated vertebra. The risk of NSFs is also related to the location of the untreated vertebrae, the number of augmented vertebrae, and the type of augmentation.

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This work was supported by the Science and Technology Support Program of Sichuan Province (grant no: 2021YFG0137).

References

1. Galibert P, Deramond H, Rosat P, Le Gars D. Note préliminaire sur le traitement des angiomes vertébraux par vertébroplastie acrylique percutanée [Preliminary note on the treatment of vertebral angioma by percutaneous acrylic vertebroplasty]. *Neurochirurgie*. 1987;33(2):166-168. [CrossRef]

2. Clark W, Bird P, Gonski P, et al. Safety and efficacy of vertebroplasty for acute painful osteoporotic fractures (VAPOUR): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet*. 2016;388(10052):1408-1416. Erratum in: *Lancet*. 2017;389(10069):602. [CrossRef]
3. Klazen CA, Lohle PN, de Vries J, et al. Vertebroplasty versus conservative treatment in acute osteoporotic vertebral compression fractures (Vertos II): an open-label randomised trial. *Lancet*. 2010;376(9746):1085-1092. [CrossRef]
4. De Leacy R, Chandra RV, Barr JD, et al. The evidentiary basis of vertebral augmentation: a 2019 update. *J Neurointerv Surg*. 2020;12(5):442-447. [CrossRef]
5. Nas ÖF, İnecikli MF, Hacıkurt K, et al. Effectiveness of percutaneous vertebroplasty in patients with multiple myeloma having vertebral pain. *Diagn Interv Radiol*. 2016;22(3):263-268. [CrossRef]
6. Papanastassiou ID, Phillips FM, Van Meirhaeghe J, et al. Comparing effects of kyphoplasty, vertebroplasty, and non-surgical management in a systematic review of randomized and non-randomized controlled studies. *Eur Spine J*. 2012;21(9):1826-1843. [CrossRef]
7. Anderson PA, Froyshteter AB, Tontz WL Jr. Meta-analysis of vertebral augmentation compared with conservative treatment for osteoporotic spinal fractures. *J Bone Miner Res*. 2013;28(2):372-382. [CrossRef]
8. Lou S, Shi X, Zhang X, Lyu H, Li Z, Wang Y. Percutaneous vertebroplasty versus non-operative treatment for osteoporotic vertebral compression fractures: a meta-analysis of randomized controlled trials. *Osteoporos Int*. 2019;30(12):2369-2380. [CrossRef]
9. Zhu RS, Kan SL, Ning GZ, et al. Which is the best treatment of osteoporotic vertebral compression fractures: balloon kyphoplasty, percutaneous vertebroplasty, or non-surgical treatment? A Bayesian network meta-analysis. *Osteoporos Int*. 2019;30(2):287-298. [CrossRef]
10. Hinde K, Maingard J, Hirsch JA, Phan K, Asadi H, Chandra RV. Mortality outcomes of vertebral augmentation (vertebroplasty and/or balloon kyphoplasty) for osteoporotic vertebral compression fractures: a systematic review and meta-analysis. *Radiology*. 2020;295(1):96-103. [CrossRef]
11. Anderson PA, Conley RB. Secondary fracture prevention: review of recent American Society for Bone and Mineral Research multidisciplinary stakeholder consensus recommendations. *Spine J*. 2020;20(7):1044-1047. [CrossRef]
12. Chen XS, Jiang JM, Sun PD, Zhang ZF, Ren HL. How the clinical dosage of bone cement biomechanically affects adjacent vertebrae. *J Orthop Surg Res*. 2020;15(1):370. [CrossRef]
13. Luo J, Annesley-Williams DJ, Adams MA, Dolan P. How are adjacent spinal levels affected by vertebral fracture and by vertebroplasty? A biomechanical study on cadaveric spines. *Spine J*. 2017;17(6):863-874. [CrossRef]
14. Polikeit A, Nolte LP, Ferguson SJ. The effect of cement augmentation on the load transfer in an osteoporotic functional spinal unit: finite-element analysis. *Spine (Phila Pa 1976)*. 2003;28(10):991-996. [CrossRef]
15. Lee BG, Choi JH, Kim DY, Choi WR, Lee SG, Kang CN. Risk factors for newly developed osteoporotic vertebral compression fractures following treatment for osteoporotic vertebral compression fractures. *Spine J*. 2019;19(2):301-305. [CrossRef]
16. Chiu PY, Kao FC, Hsieh MK, et al. A retrospective analysis in 1347 patients undergoing cement augmentation for osteoporotic vertebral compression fracture: is the sandwich vertebra at a higher risk of further fracture? *Neurosurgery*. 2021;88(2):342-348. [CrossRef]
17. Nieuwenhuijse MJ, Putter H, van Erkel AR, Dijkstra PD. New vertebral fractures after percutaneous vertebroplasty for painful osteoporotic vertebral compression fractures: a clustered analysis and the relevance of intradiskal cement leakage. *Radiology*. 2013;266(3):862-870. [CrossRef]
18. Liu J, Tang J, Gu Z, Zhang Y, Yu S, Liu H. Fracture-free probability and predictors of new symptomatic fractures in sandwich, ordinary-adjacent, and non-adjacent vertebrae: a vertebra-specific survival analysis. *J Neurointerv Surg*. 2021;13(11):1058-1062. [CrossRef]
19. Wang YT, Wu XT, Chen H, Wang C, Mao ZB. Adjacent-level symptomatic fracture after percutaneous vertebral augmentation of osteoporotic vertebral compression fracture: a retrospective analysis. *J Orthop Sci*. 2014;19(6):868-876. [CrossRef]
20. Baek SW, Kim C, Chang H. The relationship between the spinopelvic balance and the incidence of adjacent vertebral fractures following percutaneous vertebroplasty. *Osteoporos Int*. 2015;26(5):1507-1513. [CrossRef]
21. Liu J, Tang J, Liu H, Gu Z, Zhang Y, Yu S. A novel and convenient method to evaluate bone cement distribution following percutaneous vertebral augmentation. *Sci Rep*. 2020;10(1):16320. [CrossRef]
22. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ*. 1995;310(6973):170. [CrossRef]
23. Kim SH, Kang HS, Choi JA, Ahn JM. Risk factors of new compression fractures in adjacent vertebrae after percutaneous vertebroplasty. *Acta Radiol*. 2004;45(4):440-445. [CrossRef]
24. Liu J, Tang J. Letter: a retrospective analysis in 1347 patients undergoing cement augmentation for osteoporotic vertebral compression fracture: is the sandwich vertebra at a higher risk of further fracture? *Neurosurgery*. 2021;88(6):E566-E567. [CrossRef]
25. Tseng YY, Yang TC, Tu PH, Lo YL, Yang ST. Repeated and multiple new vertebral compression fractures after percutaneous transpedicular vertebroplasty. *Spine (Phila Pa 1976)*. 2009;34(18):1917-1922. [CrossRef]
26. Sun G, Tang H, Li M, Liu X, Jin P, Li L. Analysis of risk factors of subsequent fractures after vertebroplasty. *Eur Spine J*. 2014;23(6):1339-1345. [CrossRef]
27. Kong M, Zhou C, Zhu K, et al. 12-month teriparatide treatment reduces new vertebral compression fractures incidence and back pain and improves quality of life after percutaneous kyphoplasty in osteoporotic women. *Clin Interv Aging*. 2019;14:1693-1703. [CrossRef]
28. Lindsay R, Silverman SL, Cooper C, et al. Risk of new vertebral fracture in the year following a fracture. *JAMA*. 2001;285(3):320-323. [CrossRef]
29. Dohm M, Black CM, Dacre A, Tillman JB, Fueredi G; KAVIAR investigators. A randomized trial comparing balloon kyphoplasty and vertebroplasty for vertebral compression fractures due to osteoporosis. *AJNR Am J Neuroradiol*. 2014;35(12):2227-2236. [CrossRef]
30. Wang H, Sribastav SS, Ye F, et al. Comparison of percutaneous vertebroplasty and balloon kyphoplasty for the treatment of single level vertebral compression fractures: a meta-analysis of the Literature. *Pain Physician*. 2015;18(3):209-222. [CrossRef]
31. Cheng X, Long HQ, Xu JH, Huang YL, Li FB. Comparison of unilateral versus bilateral percutaneous kyphoplasty for the treatment of patients with osteoporosis vertebral compression fracture (OVCF): a systematic review and meta-analysis. *Eur Spine J*. 2016;25(11):3439-3449. [CrossRef]

Appendix Table 1. Pairwise comparison of demographic and clinical characteristics of patients with early, middle, and late new symptomatic fractures*

Characteristic	Early vs. mid-term	Early vs. late	Mid-term vs. late
Age (years)	0.725	0.042	0.033
Female	0.479	0.722	0.672
Body mass index (kg/m²)	0.884	0.942	0.830
Distribution of bone cement	0.393	0.062	0.462
T-score of bone mineral density	0.280	0.033	0.475
No of treated vertebrae	<0.001	<0.001	0.952
DXA results available	0.568	0.627	0.863
Presence of clefts	0.090	0.137	0.580
Cause of injury known	0.360	0.018	0.260
Degree of compression of previous fractures ≥50%	0.261	0.128	0.900
Kyphosis angle of previous fractures ≥10°	0.015	0.642	0.032
Thoracolumbar	0.351	0.036	0.005
PKP	0.061	<0.001	0.013
Unnilateral puncture	0.486	0.417	0.977
Intervertebral leakage	0.753	0.432	0.724

*The adjusted *P* value according to Bonferroni correction: α: 0.05/3 = 0.017. Numbers are *P* values. The intervals for defining "early," "mid-term," and "late" are explained in the methods section of the manuscript.
Differences in normally distributed continuous data were assessed for significance using One-Way analysis of variance and pairwise comparisons using the least significant difference test, including data of age, body mass index, distribution of bone cement, and T-score of bone mineral density. Skewed data were assessed using the Kruskal-Wallis test for significance assessment or Wilcoxon tests for pairwise comparisons, including data of number of treated vertebrae. Differences in categorical data were assessed using chi-squared tests, including data of sex, presence of clefts, cause of injury known, degree of compression of previous fractures ≥50%, kyphosis angle of previous fractures ≥10°, thoracolumbar location, PKP or PVP, unipedicular puncture, intervertebral leakage, and DXA results available. DXA, dual energy X-ray absorptiometry; PKP, percutaneous kyphoplasty; PVP, percutaneous vertebroplasty.

Appendix Table 2. Definitions of categorical variables used to identify potential risk factors for new symptomatic fractures

Variable	Categories	No of vertebrae with/without NSF
Sex	0: female	222/2.069
	1: male	51/418
Distance from the nearest augmented vertebra	1: one segment away	125/443
	2: two segments away	46/387
	3: three segments away	40/339
	4: four segments away	22/307
	5: five or more segments away	40/1.011
Location of the untreated vertebra	0: thoracolumbar segment	130/1.940
	1: non-thoracolumbar segment	143/547
Location of the nearest augmented vertebra	0: thoracolumbar segment	162/1.498
	1: non-thoracolumbar segment	111/989
Intradiscal cement leakage	0: no	266/2.459
	1: yes	7/28
Degree of compression ≥50%	0: no	220/2.033
	1: yes	53/454
Cleft sign	0: no	240/2.124
	1: yes	33/363
Augmentation method	1: PVP	160/1.467
	2: PKP	113/1.020
Puncture method	0: bilateral	57/536
	1: unilateral	216/1.951
Cause of injury	0: no	182/1.689
	1: yes	91/798

NSF, new symptomatic fracture; PKP, percutaneous kyphoplasty; PVP, percutaneous vertebroplasty.

Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.232550

NEURORADIOLOGY

REVIEW

Survival prediction using apparent diffusion coefficient values in recurrent glioblastoma under bevacizumab treatment: an updated systematic review and meta-analysis

Dong Liu
Zhangyu Li

Huzhou Central Hospital, The Affiliated Central Hospital of Huzhou Teachers College, Department of Radiology, Zhejiang, China

ABSTRACT

Bevacizumab is a common strategy for the treatment of recurrent glioblastoma. Survival status is a crucial issue for patients with recurrent glioblastoma, and the apparent diffusion coefficient (ADC) values of the lower Gaussian curve have been reported to have the potential to predict prognosis in recurrent glioblastoma. In the present study, we aimed to clarify the survival prediction of ADC values in patients with recurrent glioblastoma receiving bevacizumab treatment through a systematic review and meta-analysis of randomized clinical trials, comparing ADC values higher than the cut-off values with those lower than the cut-off values to determine which type of ADC values can be associated with significant survival benefits. Different survival indicators were analyzed, including overall survival (OS) and progression-free survival (PFS). Ten studies with a total of 782 patients with recurrent glioblastoma were included. The focused outcomes were OS and PFS. Our results showed that ADC values lower than the cut-off values were associated with significant benefits for OS status compared with ADC values higher than the cut-off values. Similar significant benefits were observed for PFS. The meta-analysis results suggest that ADC values lower than the cut-off values might be associated with significant benefits for OS and PFS when compared with ADC values higher than the cut-off values. However, bias in relation to the different stages of recurrent glioblastoma and different types, doses, and regimens of bevacizumab should not be ignored.

KEYWORDS

Glioblastoma, bevacizumab treatment, apparent diffusion coefficient, overall survival, progression-free survival

Glioblastoma is an aggressive and malignant brain tumor[1] with a median survival duration of 8–14 months.[2,3] Concurrent chemotherapy and radiotherapy with surgery is still unable to achieve a favorable prognosis, and most glioblastomas are recurrent.[1,4] Glioblastoma is a tumor characterized by cell anaplasia, necrosis, prominent angiogenesis, and hyperoxygenation,[5] which activate vascular endothelial growth factor A, a target molecule in the treatment of the disease.[6]

To inhibit this target molecule, bevacizumab, a humanized monoclonal antibody, is a reasonable option to treat glioblastoma. Its clinical efficacy has been established in many types of cancers, such as renal cell carcinoma,[7] colorectal cancer,[8] cervical cancer,[9] and lung cancer.[10] For glioma, the clinical effects of bevacizumab on overall survival (OS) and progression-free survival (PFS) might be controversial,[11-13] with one trial finding no evidence of improved OS with bevacizumab treatment.[11] In addition, bevacizumab has not been approved for chemotherapy in patients with recurrent glioblastoma in the European Union, probably due to the lack of evidence for its anti-tumor effects. However, bevacizumab was approved by the US Food and Drug Administration to treat recurrent glioblastoma in 2009. European guidelines also include bevacizumab as a treatment option for recurrent glioblastoma because of its demonstrated improvement in quality of life, safety,[13,14] and the prolongation of OS and PFS in patients.[15]

Corresponding author: Zhangyu Li

E-mail: lizhyuzj@sina.com

Received 13 October 2023; revision requested 13 November 2023; last revision received 11 December 2023; accepted 01 January 2024.

Epub: 31.01.2024

Publication date: 08.07.2024

DOI: 10.4274/dir.2024.232550

You may cite this article as: Liu D, Li Z. Survival prediction using apparent diffusion coefficient values in recurrent glioblastoma under bevacizumab treatment: an updated systematic review and meta-analysis. *Diagn Interv Radiol*. 2024;30(4):270-274.

Magnetic resonance imaging (MRI) is usually used to diagnose and evaluate therapeutic effects in patients with recurrent glioblastoma. A recent meta-analysis showed that perfusion MRI might be beneficial for predicting prognosis in patients with recurrent glioblastoma receiving bevacizumab treatment.[16] One type of MRI method, the diffusion-weighted MRI, uses the diffusion process of water molecules in the brain to generate contrast and obtain the diffusion values to detect the structural characteristics of brain white matter.[17-19] In addition, diffusion-weighted imaging might be useful for predicting prognosis in recurrent glioblastoma, especially by obtaining the mean apparent diffusion coefficient (ADC) value of the lower Gaussian curve, which is calculated from the histogram analysis.[20,21] In the current systematic review and meta-analysis, we aimed to clarify the role of high and low ADC values in prognosis prediction for patients with recurrent glioblastoma receiving bevacizumab treatment, especially regarding OS and PFS. We included up-to-date eligible studies to confirm the role of ADC values in a prediction biomarker.

Methods

Literature search criteria

A set of keywords was used to search for and collect relevant studies using the Web of Science, PubMed, Embase, Cochrane Central Register of Controlled Trials, and ScienceDirect databases. The keywords were as follows: "bevacizumab," "chemotherapy," "glioblastoma," "recurrent," "magnetic," "MRI," "apparent diffusion coefficient," "ADC," "cohort," "prognosis," "prediction," "treatment," "therapy," "survival," "outcome," "comparison," "prognostic," and "observational." We only considered articles published (including online) before September 2023.

The inclusion criteria for the articles were as follows: (1) cohort or observational studies, (2) comparisons between ADC values higher and lower than the cut-off values for the survival status of patients with recurrent glioblastoma receiving bevacizumab treatment, (3) outcome profiles at baseline and endpoint for survival (including OS and PFS), (4) inclusion of detailed survival data such as the *P* value, 95% confidence interval (CI), or hazard ratio (HR), and (5) publication in journals in the science citation index database and in the English language.

Main points

- The white matter diffusion value, the apparent diffusion coefficient (ADC), may help predict prognosis in recurrent glioblastoma, a type of brain cancer.
- For patients with recurrent glioblastoma receiving bevacizumab treatment, ADC may inform the survival prognosis.
- An ADC value lower than the cut-off values may predict improved status in overall survival and progression-free survival.

Reporting bias assessment

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was used to evaluate the risk of bias for the eligible studies, which included the following dimensions: patient selection, index test, reference standard, and flow and timing. We chose QUADAS because it is a useful and validated tool to evaluate the risk of bias of diagnostic accuracy studies in a systematic review.[22] The risk-of-bias assessment was reported and visualized according to the above four dimensions. In addition, a funnel plot was used to assess the publication bias of the included studies.

Data quality evaluation and collection

We performed the current systematic review and meta-analysis study according to the Cochrane Handbook for Systematic Reviews and Interventions and reported the results according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[23] The following data were collected from the included articles: first, the HR and either the *P* value or 95% CI for OS as well as the patient number of patients with recurrent glioblastoma with ADC values higher than the cut-off values under bevacizumab treatment; second, the HR and either the *P* value or 95% CI for OS as well as the patient number for patients with recurrent glioblastoma with ADC values lower than the cut-off values receiving bevacizumab treatment; third, the HR and either the *P* value or 95% CI for PFS as well as the patient number for patients with recurrent glioblastoma with ADC values higher than the cut-off values receiving bevacizumab treatment; fourth, the HR and either the *P* value or 95% CI for PFS as well as the patient number for patients with recurrent glioblastoma with ADC values higher than the cut-off values receiving bevacizumab treatment.

Critical appraisal of data

Two researchers (D.L. and Z.L.) assessed the abstracts to screen out articles. Each reviewer independently evaluated the full text version of the included articles. An independent extraction of clinical outcome data from the text, tables, and figures of the selected citations was also performed. The included articles all had data on OS or PFS in the full text. A strong agreement was achieved through a collaborative review by all reviewers (kappa: 0.8). All researchers reviewed the final results.

Meta-analysis and statistical analysis

For OS or PFS, pooled HR estimates were generated with the associated 95% CI or *P* value or individual HR. The summary statistics for each eligible study were assessed, and we extracted the reported HRs and *P* value or 95% CIs if patient-level data were lacking. We used the Cochrane Collaboration Review Manager Software Package (Rev Man Version 5.4, Cochrane library, 11-13 Cavendish Square, London, UK). to perform the meta-analyses. The log HRs were calculated by transforming the HR and *P* value or beginning and end of the 95% CIs in the Rev Man calculation function. The risk estimates of eligible studies were also evaluated by the inverse variance weighted averages of log HRs in the random-effects model.

ADC values higher than the cut-off values were compared with those lower than the cut-off values to determine which type of ADC values could be associated with an improved OS and PFS profile. Chi-square tests were used to assess the heterogeneity between the eligible citations, and the derived I^2 statistic was applied to assess the statistical heterogeneity of the eligible citations in the meta-analysis. The cut-off value for the Higgins I^2 index was based on the suggestions of the Cochrane Handbook for Systematic Reviews of Interventions (2nd edition),[24] and two-sided *P* values were also calculated.

Results

Description of studies

The PRISMA flowchart for our article selection process is presented in Figure 1. Finally, 10 studies were included.[20,21,25-32] The QUADAS risk-of-bias assessment is presented in Figure 2. A symmetric distribution is shown in the funnel plot of eligible studies.

Log hazard ratio of apparent diffusion coefficient values higher than the cut-off values against those lower than the cut-off values for overall survival

The I^2 was 0%, which indicated low heterogeneity. The test for overall effect was $Z = 6.64$ ($P < 0.00001$), and the meta-analysis

results revealed a significant difference in the log HR of OS events between ADC values higher than the cut-off values and those lower than the cut-off values, suggesting a significant benefit for OS for ADC values lower than the cut-off values (Figure 3).

Log hazard ratio of apparent diffusion coefficient values higher than the cut-off values against those lower than the cut-off values for progression-free survival

The I^2 was 11%, which indicated low heterogeneity. The test for overall effect was $Z = 5.58$ ($P < 0.00001$), and the meta-analysis results revealed a significant difference in the log HR of PFS events between ADC values higher than the cut-off values and those lower than the cut-off values, suggesting a significant benefit for PFS for ADC values lower than the cut-off values (Figure 4).

Discussion

We found that ADC values lower than the cut-off values were superior to ADC values higher than the cut-off values for OS and PFS in patients with recurrent glioblastoma receiving bevacizumab treatment. In addition, the low heterogeneity within the eligible studies in the current meta-analysis was noted. Despite the characteristics of diffusion-weighted imaging for detecting the microstructure of the brain and tumor, the prognostic potential of ADC values in patients with recurrent glioblastoma receiving bevacizumab treatment still needs to be clarified. The low heterogeneity might decrease the potential impact from the statistical and clinical heterogeneity in the current meta-analysis. However, the possible biases in the eligible studies should not be ignored. Our meta-analysis was different from a previous meta-analysis[33] in terms of the following: (1) our met-analysis included the most up-to-date studies on ADC values for OS and PFS for recurrent glioblastoma with bevacizumab treatment; (2) our meta-analysis identified more significant differences with greater Z values; (3) the heterogeneity of our meta-analysis was lower than that of the previous meta-analysis; (4) our QUADAS assessment results showed a more stringent evaluation for the included studies. Therefore, our meta-analysis provides up-to-date and valuable information on the topic and can help confirm the prognostic role of ADC values in patients with recurrent glioblastoma receiving bevacizumab treatment.

ADC values measure the water diffusivity and viscosity of the brain, indicating that ADC values might represent the bio-physi-

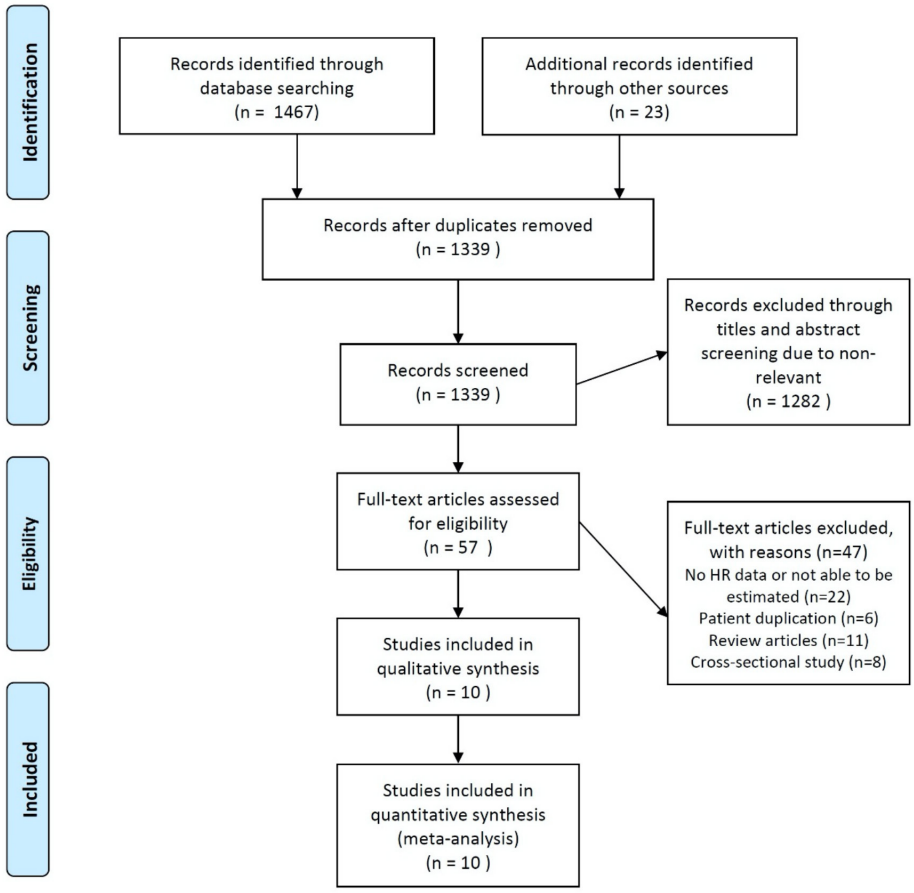


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart for the selection of eligible studies. Identification of the potentially relevant literature and screening of the identified literature using abstract and title selection adhered to PRISMA guidelines. The assessment of the full text of the screened literature aimed to find eligible studies. Suitable studies were then included in the final meta-analysis. HR, hazard ratio.

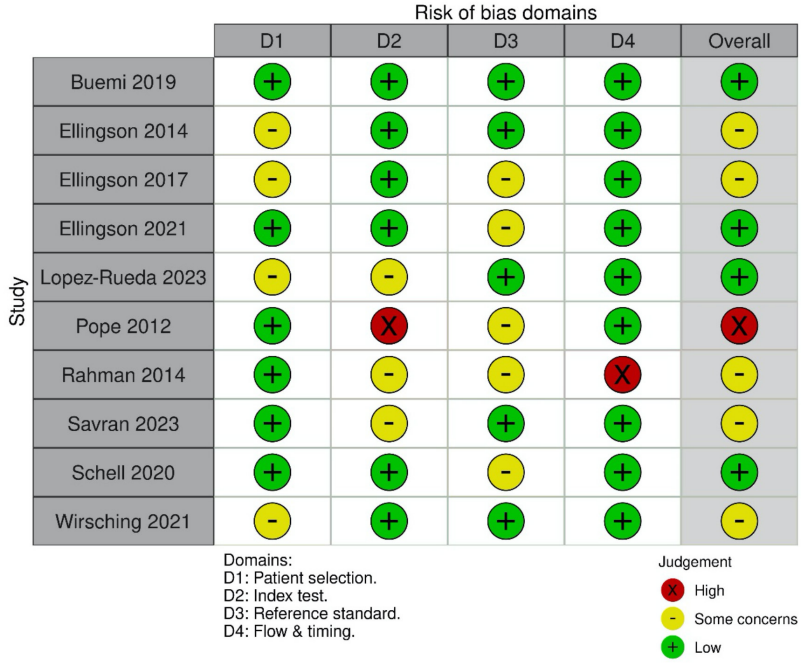


Figure 2. Risk-of-bias assessment. The Quality Assessment of Diagnostic Accuracy Studies tool was used to assess the risk of bias in the included articles.

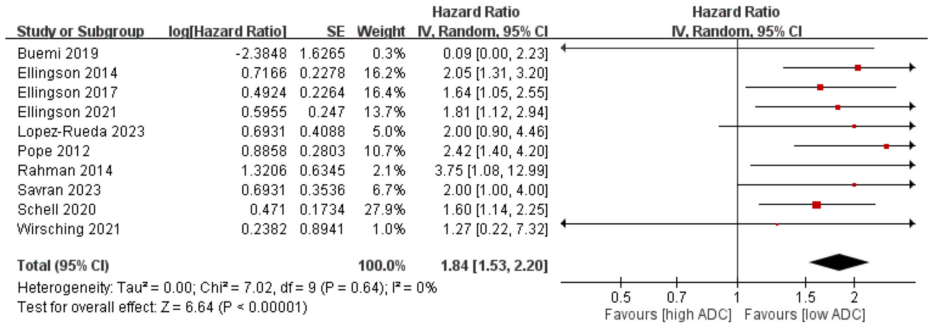


Figure 3. Forest plot of the log hazards ratio for the meta-analysis results for overall survival (OS): apparent diffusion coefficient (ADC) values higher than the cut-off values against those lower than the cut-off values. ADC values lower than the cut-off values showed a significant benefit in terms of improved OS compared with ADC values higher than the cut-off values (statistically significant). CI, confidence interval; SE, standard error.

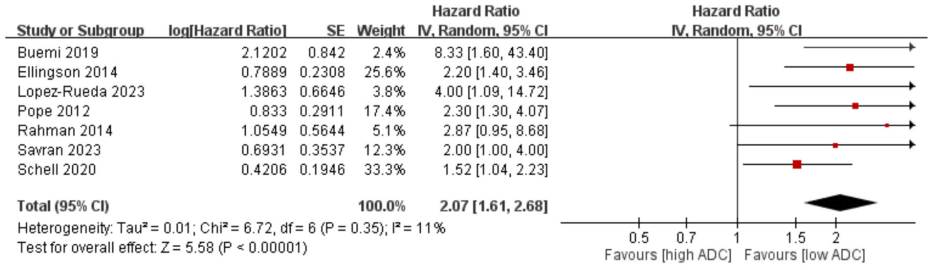


Figure 4. Forest plot of log hazards ratio for the meta-analysis results for progression-free survival (PFS): apparent diffusion coefficient (ADC) values higher than the cut-off values against those lower than the cut-off values. ADC values lower than the cut-off values showed a significant benefit in terms of improved PFS compared with ADC values higher than the cut-off values (statistically significant). CI, confidence interval; SE, standard error.

cal characteristics of tissue diffusivity within the glioblastoma area. Although ADC values might predict prognosis in recurrent glioblastoma, consistent pathological evidence and reliable biological models have not been established. One study suggested that ADC values might be associated with the oxygenated or cellular status of glioblastoma, which might influence and interfere with the effectiveness of bevacizumab in the case of an aggressive glioblastoma.[26] A recent biological study also suggested that ADC values might be associated with the increased expression of decorin, a small proteoglycan that modulates angiogenesis and viscosity;[20,34,35] it also binds to various macromolecules and activates metalloproteinases in the extracellular matrix.[20,36,37] This might explain the possible underlying mechanisms of the characteristics of independent imaging biomarkers related to ADC values in the prognosis of recurrent glioblastoma with bevacizumab treatment. These results suggest that patients with recurrent glioblastoma with ADC values lower than the cut-off values might be appropriate candidates for bevacizumab chemotherapy. By contrast, patients with ADC values higher than the cut-off values might not be suitable for bevacizumab chemotherapy. These findings might provide an initial model in terms of precision medicine for chemotherapy for patients with recurrent glioblastoma.

Our meta-analysis has several limitations. First, histopathological evidence to support the role of ADC values in the prediction of prognosis in recurrent glioblastoma is lacking. Determining consistent histopathological evidence in a future study is warranted to clarify the underlying biological mechanisms. Currently, most theoretical explanations relating to the role of ADC values in prediction are speculative and not based on solid evidence. In addition, the cut-off point or threshold of ADC values was diverse in the studies included in the meta-analysis. A further meta-analysis with a homogenous threshold in relation to this aspect is warranted in the future. Second, the age and gender variances in the included studies might influence the interpretation of our study results. More consistent age and gender distribution patterns might be needed in future randomized clinical trials to decrease the bias caused by different age and gender distributions. Third, variations in bevacizumab regimens might also bias our meta-analysis results. More consistent bevacizumab regimens might be helpful for improving the accuracy of the meta-analytic results. Fourth, the different techniques, doses, regimens, and durations of combined radiotherapy in the included studies might also influence the interpretations of our results. Fifth, the lack of a meta-analysis on treatment adverse events, toxicities, and compliance might prevent detailed conclusions. Sixth, most of the included studies were from Europe and the USA. The ethnicity bias might influence the interpretations of our current meta-analysis. Seventh, it is impossible to control the bias from the different brands, magnetic strengths, pulse sequences, and default settings of MRI machines at different sites in the different included studies. This type of bias should not be ignored. Eighth, one included study[20] involved patients with isocitrate dehydrogenase (IDH) mutation; the role of IDH mutation in ADC values might need to be clarified in our meta-analysis. Finally, the variable cut-off values of different included studies provide another limitation to our meta-analysis.

In conclusion, the meta-analysis results suggest that ADC values lower than the cut-off values might be associated with significant benefits for OS and PFS when compared with ADC values higher than the cut-off values. However, the bias caused by the different stages of recurrent glioblastoma and different types, doses, and regimens of bevacizumab should not be ignored.

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This research was funded by a grant from the Huzhou Public Welfare Research General Project (2021GBY46).

References

1. Stupp R, Taillibert S, Kanner A, et al. Effect of tumor-treating fields plus maintenance temozolomide vs maintenance temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial. *JAMA*. 2017;318(23):2306-2316. [CrossRef]
2. Ostrom QT, Patil N, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2013-2017. *Neuro Oncol*. 2020;22(12 Suppl 2):iv1-iv96. [CrossRef]
3. Van Meir EG, Hadjipanayis CG, Norden AD, Shu HK, Wen PY, Olson JJ. Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma. *CA Cancer J Clin*. 2010;60(3):166-193. [CrossRef]

4. Kumar A A, Abraham Koshy A. Regression of recurrent high-grade glioma with temozolomide, dexamethasone, and levetiracetam: case report and review of the literature. *World Neurosurg*. 2017;108:990.e11-990.e16. [CrossRef]
5. Lu-Emerson C, Duda DG, Emblem KE, et al. Lessons from anti-vascular endothelial growth factor and anti-vascular endothelial growth factor receptor trials in patients with glioblastoma. *J Clin Oncol*. 2015;33(10):1197-1213. [CrossRef]
6. Seyedmirzaei H, Shobeiri P, Turgut M, Hanaei S, Rezaei N. VEGF levels in patients with glioma: a systematic review and meta-analysis. *Rev Neurosci*. 2020;32(2):191-202. [CrossRef]
7. Hainsworth JD, Sosman JA, Spigel DR, Edwards DL, Baughman C, Greco A. Treatment of metastatic renal cell carcinoma with a combination of bevacizumab and erlotinib. *J Clin Oncol*. 2005;23(31):7889-7896. [CrossRef]
8. Qu C-Y, Zheng Y, Zhou M, et al. Value of bevacizumab in treatment of colorectal cancer: a meta-analysis. *World J Gastroenterol*. 2015;21(16):5072-5080. [CrossRef]
9. Kehoe S. Bevacizumab and treatment of cervical cancer. *Maturitas*. 2014;79(4):355-356. [CrossRef]
10. Besse B, Le Moulec S, Mazières J, et al. Bevacizumab in patients with nonsquamous non-small cell lung cancer and asymptomatic, untreated brain metastases (BRAIN): a nonrandomized, phase II study. *Clin Cancer Res*. 2015;21(8):1896-1903. [CrossRef]
11. van den Bent MJ, Klein M, Smits M, et al. Bevacizumab and temozolomide in patients with first recurrence of WHO grade II and III glioma, without 1p/19q co-deletion (TAVAREC): a randomised controlled phase 2 EORTC trial. *Lancet Oncol*. 2018;19(9):1170-1179. [CrossRef]
12. Weller M, Van Den Bent M, Tonn JC, et al. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. *Lancet Oncol*. 2017;18(6):e315-e329. [CrossRef]
13. Weller M, van den Bent M, Preusser M, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol*. 2021;18(3):170-186. [CrossRef]
14. Fu M, Zhou Z, Huang X, et al. Use of Bevacizumab in recurrent glioblastoma: a scoping review and evidence map. *BMC Cancer*. 2023;23(1):544. [CrossRef]
15. Diaz RJ, Ali S, Qadir MG, De La Fuente MI, Ivan ME, Komotar RJ. The role of bevacizumab in the treatment of glioblastoma. *J Neurooncol*. 2017;133(3):455-467. [CrossRef]
16. Choi SH, Jung SC, Kim KW, et al. Perfusion MRI as the predictive/prognostic and pharmacodynamic biomarkers in recurrent malignant glioma treated with bevacizumab: a systematic review and a time-to-event meta-analysis. *J Neurooncol*. 2016;128(2):185-194. [CrossRef]
17. Vanderweyen DC, Theaud G, Sidhu J, et al. The role of diffusion tractography in refining glial tumor resection. *Brain Struct Funct*. 2020;225(4):1413-1436. [CrossRef]
18. Wang Y, Lv X, Gong H, et al. Acute irradiation injury of canine brain with pathology control is detected by diffusion-weighted imaging of MRI. *Clin Imaging*. 2013;37(3):440-445. [CrossRef]
19. Sawlani V. Diffusion-weighted imaging and apparent diffusion coefficient evaluation of herpes simplex encephalitis and Japanese encephalitis. *J Neurol Sci*. 2009;287(1-2):221-226. [CrossRef]
20. Ellingson BM, Patel K, Wang C, et al. Validation of diffusion MRI as a biomarker for efficacy using randomized phase III trial of bevacizumab with or without VB-111 in recurrent glioblastoma. *Neurooncol Adv*. 2021;3(1):vdab082. [CrossRef]
21. Schell M, Pflüger I, Brugnara G, et al. Validation of diffusion MRI phenotypes for predicting response to bevacizumab in recurrent glioblastoma: post-hoc analysis of the EORTC-26101 trial. *Neuro Oncol*. 2020;22(11):1667-1676. [CrossRef]
22. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536. [CrossRef]
23. Knobloch K, Yoon U, Vogt PM. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *J Craniomaxillofac Surg*. 2011;39(2):91-92. [CrossRef]
24. Cumpston MS, McKenzie JE, Welch VA, Brennan SE. Strengthening systematic reviews in public health: guidance in the Cochrane Handbook for Systematic Reviews of Interventions, 2nd edition. *J Public Health (Oxf)*. 2022;44(4):e588-e592. [CrossRef]
25. Buemi F, Guzzardi G, Del Sette B, et al. Apparent diffusion coefficient and tumor volume measurements help stratify progression-free survival of bevacizumab-treated patients with recurrent glioblastoma multiforme. *Neuroradiol J*. 2019;32(4):241-249. [CrossRef]
26. Ellingson B, Sahebjam S, Kim H, et al. Pretreatment ADC histogram analysis is a predictive imaging biomarker for bevacizumab treatment but not chemotherapy in recurrent glioblastoma. *AJNR Am J Neuroradiol*. 2014;35(4):673-679. [CrossRef]
27. Ellingson BM, Gerstner ER, Smits M, et al. Diffusion MRI phenotypes predict overall survival benefit from anti-VEGF monotherapy in recurrent glioblastoma: converging evidence from phase II trials. *Clin Cancer Res*. 2017;23(19):5745-5756. [CrossRef]
28. Lopez-Rueda A, Puig J, Thió-Henestrosa S, et al. Texture analysis of the apparent diffusion coefficient focused on contrast-enhancing lesions in predicting survival for bevacizumab-treated patients with recurrent glioblastoma. *Cancers (Basel)*. 2023;15(11):3026. [CrossRef]
29. Pope WB, Qiao XJ, Kim HJ, et al. Apparent diffusion coefficient histogram analysis stratifies progression-free and overall survival in patients with recurrent GBM treated with bevacizumab: a multi-center study. *J Neurooncol*. 2012;108(3):491-498. [CrossRef]
30. Rahman R, Hamdan A, Zweifler R, et al. Histogram analysis of apparent diffusion coefficient within enhancing and nonenhancing tumor volumes in recurrent glioblastoma patients treated with bevacizumab. *J Neurooncol*. 2014;119(1):149-158. [CrossRef]
31. Savran B, Göç MF, Öztürk FU, Duran AO, Ünal Ö. Diffusion restriction associated with bevacizumab treatment in recurrent glial tumors, evaluation of survival with ADC measurement analysis. *Eur Rev Med Pharmacol Sci*. 2023;27(9):4153-4161. [CrossRef]
32. Wirsching HG, Roelcke U, Weller J, et al. MRI and ^{18}FET-PET predict survival benefit from bevacizumab plus radiotherapy in patients with isocitrate dehydrogenase wild-type glioblastoma: results from the randomized ARTE trial. *Clin Cancer Res*. 2021;27(1):179-188. [CrossRef]
33. Kurokawa R, Baba A, Kurokawa M, et al. Pretreatment ADC histogram analysis as a prognostic imaging biomarker for patients with recurrent glioblastoma treated with bevacizumab: a systematic review and meta-analysis. *AJNR Am J Neuroradiol*. 2022;43(2):202-206. [CrossRef]
34. Patel KS, Yao J, Raymond C, et al. Decorin expression is associated with predictive diffusion MR phenotypes of anti-VEGF efficacy in glioblastoma. *Sci Rep*. 2020;10(1):14819. [CrossRef]
35. Rosca EV, Koskimaki JE, Rivera CG, Pandey NB, Tamiz AP, Popel AS. Anti-angiogenic peptides for cancer therapeutics. *Curr Pharm Biotechnol*. 2011;12(8):1101-1116. [CrossRef]
36. Järveläinen H, Sainio A, Wight TN. Pivotal role for decorin in angiogenesis. *Matrix Biol*. 2015;43:15-26. [CrossRef]
37. Sofeu Feugaing DD, Götte M, Viola M. More than matrix: the multifaceted role of decorin in cancer. *Eur J Cell Biol*. 2013;92(1):1-11. [CrossRef]