

EISSN 1305-3612

DIR

Diagnostic and Interventional
Radiology

Editor in Chief

Mehmet Ruhi Onur, MD

Department of Radiology, Hacettepe University Faculty of Medicine, Ankara, Türkiye

ORCID ID: 0000-0003-1732-7862

Section Editors and Scientific Editorial Board

Abdominal Imaging

İlkay S. İdilman, MD 

Department of Radiology, Hacettepe University Faculty of Medicine, Ankara, Türkiye

ORCID ID: 0000-0002-1913-2404

Sonay Aydın, MD 

Department of Radiology, Erzincan Binali Yıldırım University Faculty of Medicine, Erzincan, Türkiye

ORCID ID: 0000-0002-3812-6333

Artificial Intelligence and Informatics

Burak Koçak, MD 

Department of Radiology, University of Health Sciences, Başakşehir Çam and Sakura City Hospital, İstanbul, Türkiye

ORCID ID: 0000-0002-7307-396X

Breast Imaging

Füsun Taşkın, MD 

Department of Radiology, Acıbadem University Faculty of Medicine, İstanbul, Türkiye

ORCID ID: 0000-0001-7985-3660

Chest and Cardiovascular Imaging

Furkan Ufuk, MD 

Department of Radiology, Pamukkale University Faculty of Medicine, Denizli, Türkiye

ORCID ID: 0000-0002-8614-5387

Hybrid Imaging and Nuclear Medicine

Evrin Bengi Türkbey, MD 

Radiology and Imaging Sciences, Clinical Center, National Institutes of Health Bethesda, Maryland, United States

ORCID ID: 0000-0002-5216-3528

Interventional Radiology

Barbaros Çil, MD, FCIIRSE 

Department of Radiology, Koç University School of Medicine, İstanbul, Türkiye

ORCID ID: 0000-0003-1079-0088

Bahri Üstünsöz, MD 

Department of Radiology, LSUHSC (Louisiana State University Health Science Center) School of Medicine, New Orleans, United States

ORCID ID: 0000-0003-4308-6708

James Milburn, MD 

Department of Radiology, Ochsner Medical System, New Orleans, Louisiana, USA

ORCID ID: 0000-0003-3403-2628

Musculoskeletal Imaging

Zeynep Maraş Özdemir, MD 

Department of Radiology, İnönü University Faculty of Medicine, Malatya, Türkiye

ORCID ID: 0000-0003-1085-8978

Neuroradiology

Gülgün Yılmaz Ovalı, MD 

Department of Radiology, Celal Bayar University Faculty of Medicine, Manisa, Türkiye

ORCID ID: 0000-0001-8433-5622

Erkan Gökçe, MD 

Department of Radiology, Tokat Gaziosmanpaşa University Faculty of Medicine, Tokat, Türkiye

ORCID ID: 0000-0003-3947-2972

Pediatric Radiology

Meltem Ceyhan Bilgici, MD 

Department of Radiology, 19 Mayıs University Faculty of Medicine, Samsun, Türkiye

ORCID ID: 0000-0002-0133-0234

Evrin Özmen, MD 

Department of Radiology, Koç University Hospital, İstanbul, Türkiye

ORCID ID: 0000-0003-3100-4197

Publication Coordinator

Şükrü Mehmet Ertürk, MD 

Department of Radiology, İstanbul University, İstanbul Faculty of Medicine, İstanbul, Türkiye

ORCID ID: 0000-0003-4086-675X

Biostatistical Consultant

İlker Ercan, PhD 

Department of Biostatistics, Uludağ University School of Medicine, Bursa, Türkiye

ORCID ID: 0000-0002-2382-290X

Publication Services

Galenos Publishing, İstanbul, Türkiye

Past Editors

Editors in Chief

Mustafa Seçil, MD (2016-2023)

Nevzat Karabulut, MD (2011-2016)

Üstün Aydingöz, MD (2010-2011)

Okan Akhan, MD (2001-2010)

Ferhun Balkancı, MD (1999-2001)

Aytekin Besim, MD (1994-1999)*

* Dr. Aytekin Besim actually served as the General Coordinator. His work in this capacity, however, was in effect that of an Editor in Chief.

Editors

Ayşenur Cila, MD (2001-2002)

Suat Kemal Aytaç, MD (1997-2001)

Erhan Ilgıt, MD (1994-2001)

Okan Akhan, MD (1994-2001)

Ferhun Balkancı, MD (1994-2000)

Serdar Akyar, MD (1994-1997)

Section Editors

Section Editorship was established in 2002 at the tenure of Dr Okan Akhan, Editor in Chief.

Abdominal Imaging

Bengi Gürses, MD (2020-2023)

Mehmet Ruhi Onur, MD (2016-2023)

Bariş Türkbey, MD (2014-2020)

Mustafa N. Özmen, MD (2012-2018)

Murat Acar, MD (2015-2016)

Mustafa Seçil, MD (2011-2016)

Ahmet Tuncay Turgut, MD (2011)

Deniz Akata, MD (2007-2011)

Ayşe Erden, MD (2002-2011)

Okan Akhan, MD (2002-2010)

Hakan Özdemir, MD (2002-2010)

Artificial Intelligence and Informatics

Barış Türkbey, MD (2020-2023)

Breast Imaging

Mustafa Erkin Arıbal, MD (2016-2023)

Sibel Kul (2015-2018)

Ayşenur Oktay, MD (2009-2014)

Ayşegül Özdemir, MD (2004-2009)

Cardiovascular Imaging

Uğur Bozlar, MD (2016-2023)

Muşturay Karçaaltıncaba, MD (2007-2010)

Mecit Kantarcı (2010-2016)

Chest Imaging

Nevzat Karabulut, MD (2010-2014)

Çetin Atasoy, MD (2007-2010)

Macit Arıyürek, MD (2002-2007)

Figen Demirkazık, MD, (2014-2018)

General Radiology

Ersin Öztürk, MD (2014-2017)

Utku Şenol, MD (2010-2013)

Oğuz Dicle, MD (2007-2010)

Interventional Radiology

Cüneyt Aytekin, MD (2016-2023)

Bora Peynircioğlu, MD (2012-2015)

Levent Oğuzkurt, MD (2011-2014)

Fatih Boyvat, MD (2007-2010)

İsmail Oran, MD (2015-2019)

Musculoskeletal Imaging

Hatice Tuba Sanal, MD (2016-2023)

Fatih Kantarcı, MD (2014-2016)

Ayşenur Oktay, MD (2011-2013)

Üstün Aydıngöz, MD (2002-2011)

Berna Dirim Mete (2016-2017)

Neuroradiology and Head & Neck Imaging

Kubilay Aydın, MD (2016-2023)

Nafi Aygün, MD (2016-2023)

Kader Karlı Oğuz, MD (2011-2015)

Süleyman Men, MD (2007-2013)

Muhteşem Ağıldere, MD (2002-2011)

Nuclear Medicine

A. Cahid Civelek, MD (2016-2023)

Oktay Sarı, MD (2015)

Akın Yıldız, MD (2011-2014)

Pediatric Radiology

Korgün Koral, MD (2016-2023)

Murat Kocaoğlu, MD (2016-2023)

Ensar Yekeler, MD (2014-2016)

Suat Fitöz, MD (2007-2013)

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 Girişim Radyol), the current title's exact Turkish translation.

Diagnostic and Interventional Radiology is an open access publication, and the journal's publication model is based on Budapest Open Access Initiative (BOAI) declaration. All published content is available online, free of charge at <https://www.dirjournal.org/>. Authors retain the copyright of their published work in Diagnostic and Interventional Radiology. The journal's content is licensed under a Creative Commons Attribution-NonCommercial (CC BY-NC) 4.0 International License which permits third parties to share and adapt the content for non-commercial purposes by giving the appropriate credit to the original work.

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 ULAKBİM 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: March 2024

EISSN 1305-3612

International scientific journal published bimonthly.



Contents

ABDOMINAL IMAGING

74 Original Article. Donor bile duct evaluation with magnetic resonance cholangiography in living-donor liver transplantation: a novel anatomical classification for predicting surgical techniques. Afak Durur Karakaya, Cemal Aydın Gündoğmuş, Turan Kanmaz, Cihan Karataş, Samet Kapakin

ARTIFICIAL INTELLIGENCE AND INFORMATICS

80 Review. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. Tugba Akinci D'Antonoli, Arnaldo Stanzione, Christian Bluethgen, Federica Vernuccio, Lorenzo Ugga, Michail E. Klontzas, Renato Cuocolo, Roberto Cannella, Burak Koçak

BREAST IMAGING

91 Original Article. Diagnostic value of synthetic diffusion-weighted imaging on breast magnetic resonance imaging assessment: comparison with conventional diffusion-weighted imaging. Ebru Yılmaz, Nilgün Gündoğan, Sıla Ulus, Ebru Banu Türk, Mustafa Enes Mısırlı, Aydan Arslan, Mustafa Erkin Arıbal

HYBRID IMAGING AND NUCLEAR MEDICINE

99 Original Article. Head-to-head comparison of ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI for lymph node metastasis staging in non-small cell lung cancer: a meta-analysis. Min Zhang, Zhikang Liu, Yuhang Yuan, Wenwen Yang, Xiong Cao, Minjie Ma, Biao Han

INTERVENTIONAL RADIOLOGY

107 Original Article. Preoperative simulation results and intraoperative image fusion guidance for transjugular intrahepatic portosystemic shunt placement: a feasibility study of nineteen patients. Xin Wei, Hong Hu, Lin Qi, Liming Zhong, Yunguo Liao, Jiaqi Pu

117 Original Article. Effect of lipiodol marking before CT-guided cryoablation on the outcome of sporadic renal cell carcinoma. Yasuhiro Ushijima, Daisuke Okamoto, Nobuhiro Fujita, Keisuke Ishimatsu, Noriaki Wada, Seiichiro Takao, Ryo Murayama, Masahiro Itoyama, Kousei Ishigami

NEURORADIOLOGY

124 Original Article. Reproducibility of rCBV in glioblastomas using T2*-weighted perfusion MRI: an evaluation of sampling, normalization, and experience. Sabahattin Yüzkan, Samet Mutlu, Mehmet Karagülle, Merve Şam Özdemir, Hamit Özgül, Mehmet Ali Arıkan, Burak Koçak

PEDIATRIC RADIOLOGY

135 Original Article. Focal fatty sparing areas of the pediatric steatotic liver: pseudolesions on hepatobiliary phase magnetic resonance images. Gözde Özer, H. Nursun Özcan, Berna Oğuz, Mithat Haliloğlu



Donor bile duct evaluation with magnetic resonance cholangiography in living-donor liver transplantation: a novel anatomical classification for predicting surgical techniques

Afak Durur Karakaya¹
 Cemal Aydın Gündoğmuş¹
 Turan Kanmaz²
 Cihan Karataş²
 Samet Kapakin³

¹Koç University Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

²Koç University Faculty of Medicine, Liver Transplantation Center, İstanbul, Türkiye

³Atatürk University Faculty of Medicine, Department of Anatomy, Erzurum, Türkiye

PURPOSE

To propose a novel, inclusive classification that facilitates the selection of the appropriate donor and surgical technique in living-donor liver transplantation (LDLT).

METHODS

The magnetic resonance cholangiography examinations of 201 healthy liver donors were retrospectively evaluated. The study group was classified according to the proposed classification. The findings were compared with the surgical technique used in 93 patients who underwent transplantation. The Couinaud, Huang, Karakas, Choi, and Ohkubo classifications were also applied to all cases.

RESULTS

There were 118 right-lobe donors (58.7%) and 83 left-lateral-segment donors (41.3%). Fifty-six (28.8%) of the cases were classified as type 1, 136 (67.7%) as type 2, and 7 (3.5%) as type 3 in the proposed classification; all cases could be classified. The number of individuals able to become liver donors was 93. A total of 36 cases were type 1, 56 were type 2, and 1 was type 3. Of the type 1 donors, 83% required single anastomosis during transplantation, whereas six patients classified as type 1 required two anastomoses, all of which were caused by technical challenges during resection. Moreover, 51.8% of the cases classified as type 2 required additional anastomosis during transplantation. The type 3 patient required three anastomoses. The type 1 and type 2 donors required a different number of anastomoses ($P < 0.001$).

CONCLUSION

The proposed classification in this study includes all anatomical variations. This inclusive classification accurately predicts the surgical technique for LDLT.

KEYWORDS

Bile duct variations, intrahepatic bile ducts, liver transplantation, magnetic resonance imaging, magnetic resonance cholangiography

Corresponding author: Afak Durur Karakaya

E-mail: afkarakaya@ku.edu.tr

Received 23 May 2023; revision requested 26 June 2023;
last revision received 20 August 2023; accepted 22 August 2023.



Epub: 19.09.2023

Publication date: 05.03.2024

DOI: 10.4274/dir.2023.232321

Liver transplantation is being increasingly performed worldwide. This is mainly because of the increasing incidence of fatty liver disease and alcohol abuse. Liver transplantation is a treatment option for acute or chronic liver diseases. It offers a second chance to live and enhances the quality of life, especially for patients who do not respond to medical or surgical treatments.¹ Liver transplantation can involve cadavers or living donors. Whereas cadaveric transplantation involves the transplantation of an organ from a deceased patient, living-donor liver transplantation (LDLT) is a surgical operation in which a portion of the liver from a healthy living person is removed and transplanted into the patient in need. In LDLT, the donor should have normal physical and mental health to ensure their protection. Specifically, liver function, vascular structure, and bile ducts should have a suitable structure and be

You may cite this article as: Durur Karakaya A, Gündoğmuş CA, Kanmaz T, Karataş C, Kapakin S. Donor bile duct evaluation with magnetic resonance cholangiography in living-donor liver transplantation: a novel anatomical classification for predicting surgical techniques. *Diagn Interv Radiol.* 2024;30(2):74-79.

of sufficient volume in terms of anatomy and function in both the donor and recipient.^{2,3} Despite advances in liver surgery techniques, complications in vascular and biliary structures are major causes of morbidity and mortality in the postoperative period. Therefore, the accurate evaluation of biliary tract anatomy and vascular structures is essential in the preoperative period. Determining the anatomy and variations of vessels and bile ducts reduces the complications that may occur in the donor or recipient. Identifying situations hindering the operation is vital, especially for donor candidates.⁴ Bile duct variations present one of these situations, with bile leakage, biliary tract stricture, cholangitis, and biliary tract stones constituting the main complications.⁵ Knowledge of intrahepatic biliary tract variations is essential in determining the surgical technique.

Various classifications for the evaluation of the biliary tract exist. These include the Huang, Couinaud, Choi, Ohkubo, and Karakas classifications, all of which describe the anatomy of the intrahepatic biliary tract. However, the current classifications cannot predict the surgical technique and number of anastomoses required. In addition, none of these classifications can be applied to all possible anatomical variations.

This study aims to propose a novel, inclusive classification that facilitates the selection of the appropriate donor and surgical technique.

Methods

This research was approved by the Ethics Committee of Koc University (protocol number: 2019.140.IRB1.014). Informed consent was obtained from all patients. A total of 201 magnetic resonance cholangiography (MRCP) examinations of healthy liver donor candidates obtained at our university hospital between June 2019 and October 2021 were retrospectively reviewed. Of the 201 healthy donors [mean age $\pm$ standard deviation (SD): 34 ± 9 years], 118 (58.7%) were men and 83 (41.3%) were women. The Couinaud,⁶ Huang,⁷ Karakas,⁸ Choi,⁹ and Ohkubo¹⁰ classi-

fications were applied to all cases. Following this, the study group was classified according to the proposed classification. The number of anastomoses in the operated cases was recorded. Three patients with non-optimal imaging were excluded from the study.

Magnetic resonance cholangiography protocol

Patients were required to fast for 4 h to reduce gastric and duodenal fluid secretions, intestinal peristalsis, and increased gallbladder distension. An antiperistaltic agent (butylscopolamine, Buscopan®) was routinely used. The antiperistaltic agent was administered intravenously immediately before the examination. The MRCP images were obtained using a 1.5 and 3 T magnetic resonance imaging system (Siemens, Erlangen, Germany). First, a two-dimensional, breath-hold half Fourier single-shot turbo spin-echo sequence was obtained in the axial plane. The entire biliary tract could be visualized up to the duodenal ampulla with two breath-holds. Subsequently, two three-dimensional respiration-triggered heavily T2-weighted fast spin echo sequences were obtained in the coronal oblique plane. The imaging plane was selected from the first axial T2-weighted images. The first acquisition was aligned to the common hepatic duct (CHD) at the level

of the pancreatic head, and the second acquisition was aligned to the pancreatic duct approximately 90° to the first imaging plane. Breath triggering was performed by monitoring the respiratory movement with navigation. The navigator was placed at the edge of the diaphragm in the coronal and sagittal localizers. The images were acquired when the position of this diaphragm interface with the lung entered the predetermined window, which ensured a consistent position in the imaging slice. An imaging series with 40 consecutive sections (all 1.5-mm thick) was obtained. The pancreaticobiliary tree has high signal intensity, whereas neighboring structures have low signal intensity because the images are predominantly T2-weighted. A maximum intensity projection (MIP) format was generated from this data volume; MIP reformats can be generated in various planes, such as coronal and sagittal obliques.

Standards of the novel classification

From the MRCP images, the separation and length of the right hepatic duct (RHD) and left hepatic duct (LHD), the total number of bile ducts joining to the CHD, and the way the cystic duct joined the CHD (directly or in the form of a cystohepatic duct) were evaluated. Based on these data, three types emerged (Figure 1). In type 1, there must be

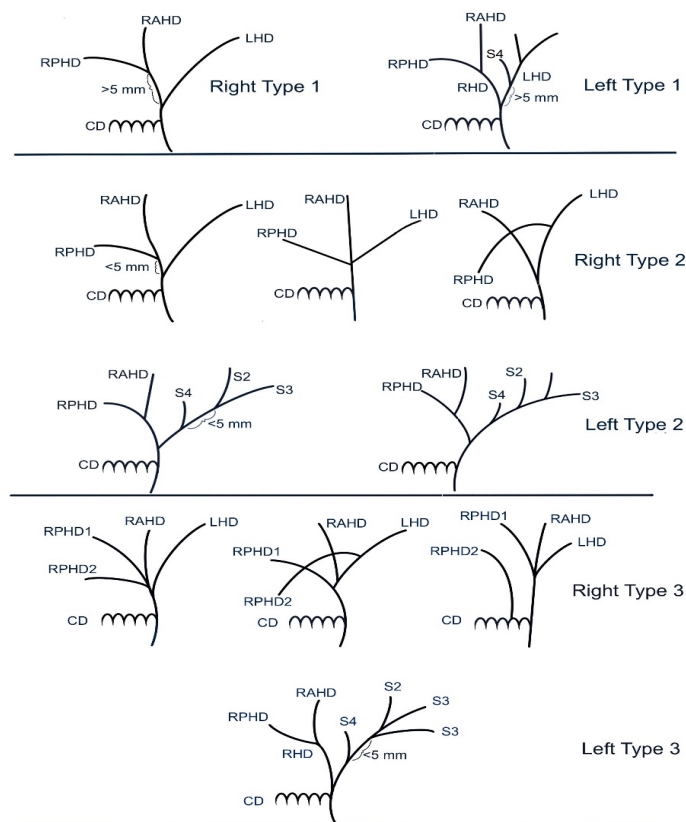


Figure 1. New classification and subtypes for both right-lobe and left-lateral-segment donors. RAHD, right anterior hepatic duct; RPHD, right posterior hepatic duct; RHD, right hepatic duct; LHD, left hepatic duct.

Main points

- The classification used in this study covers all variations.
- It estimates the surgical technique as well as the biliary tract classifications devised thus far.
- The proposed classification can accurately predict the surgical technique.

only one duct from the lobe to be donated that joins to the CHD, and the length of this duct must be at least 5 mm (Figure 2). In type 2, there must be a maximum of two ducts from the lobe to be donated that join to the CHD or one duct with a length of <5 mm (Figure 3). Type 3 involves complex biliary variants, with more than two ducts from the lobe to be donated that join to the CHD (Figure 4). All the participants could be grouped according to this new classification. The predictions of our classification in relation to surgical technique were as follows: type 1 can be considered a safe donor, type 2 can be a donor but may require additional anastomosis, and type 3 is not suited to being a donor because this type requires more than two anastomoses (Figure 5).

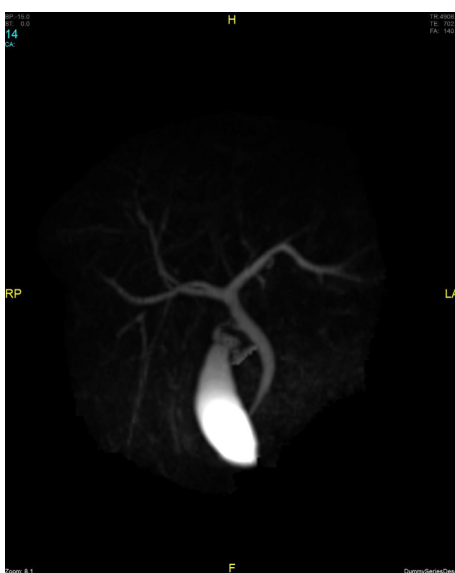


Figure 2. Type 1 configuration in right-lobe donors according to the new classification.

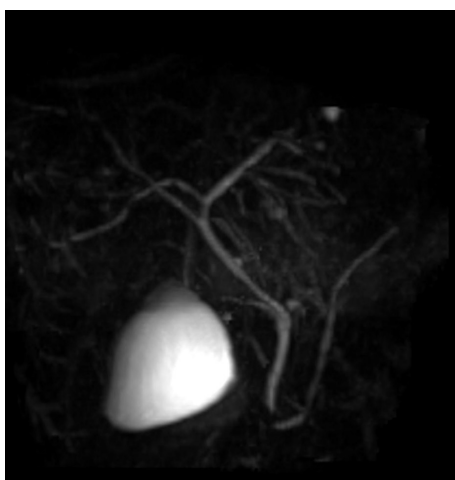


Figure 3. Type 2 configuration in right-lobe donors according to the new classification. The right posterior hepatic duct drains into the left hepatic duct.

Statistical analysis

The mean and SD were used to express all continuous data, and frequencies were used to express all categorical data. The relationship between the donor type of the proposed classification and the surgical technique was evaluated using the chi-square test. Statistical significance was defined as $P < 0.05$. The analysis was performed using IBM SPSS v.21.0 software.

Results

In this study, in which a total of 201 donors were evaluated, 118 participants were right-lobe (58.7%) and 83 were left-lateral-lobe (41.3%) donors (Table 1). In 196 of the donors, the cystic duct joined the CHD independently. Cystohepatic duct variation was present in five donors. In four of these cases, the right posterior duct joined the cystic duct, and in one case, the RHD joined the cystic duct. Regarding the total number of ducts joining the CHD, two were identified in 152 donors, three in 41 donors, four in seven donors, and five in one donor.

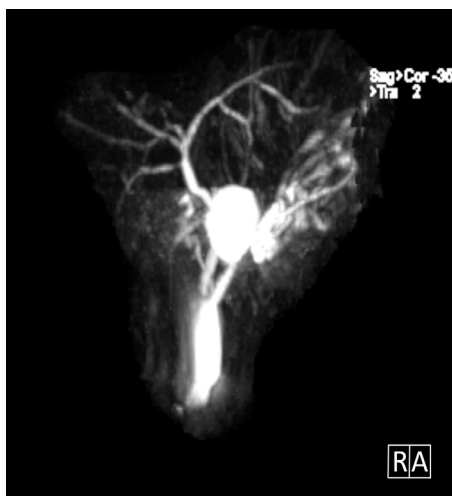


Figure 4. Type 3 configuration in right-lobe donors according to the new classification. From the right lobe, three channels drain into the common hepatic duct.

The distribution of donors according to the Huang, Couinaud, and Karakas classifications are summarized in Table 2, and those of the Ohkubo, and Choi classifications in Table 3.

A total of 56 (28.8%) of the cases were classified as type 1, 136 (67.7%) as type 2, and 7 (3.5%) as type 3. All cases could be classified (Table 3).

The number of participants that could become liver donors was 93. A number of donors were rejected because of hepatos-teatosis, vascular variation, or insufficient remnants; however, crucially, six donors were rejected because of biliary tract variations.

Among these 93 donors, 36 were type 1, 56 were type 2, and 1 was type 3 according to the proposed classification. A single anastomosis was required in 57 donors, two anastomoses were required in 35, and three in 1. A single anastomosis (single Roux-en-Y or single end-to-end) was performed during transplantation in 83.3% of the type 1 cases predicted to require a single procedure, whereas 51.8% of the type 2 cases required additional anastomosis during surgery (double Roux-en-Y or double-end) (Table 1). The participant classified as type 3 was excluded from the analysis because the case did not meet the chi-square test assumptions; however,

Table 1. Distribution of donated liver parts and surgical technique in patients undergoing transplantation

Donated liver	n = 201 (%)
Right lobe	118 (58.7)
Left lobe lateral segment	83 (41.3)
Surgical technique	n = 93 (%)
End-to-end	20 (18.7)
Roux-en-Y	37 (34.5)
Double end-to-end	12 (11.2)
Double Roux-en-Y	23 (21.5)
Triple Roux-en-Y	1 (0.9)

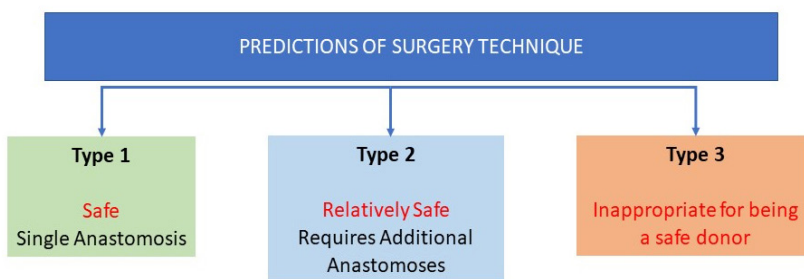


Figure 5. The choice of surgical technique according to the new classification.

this participant required three anastomoses. The donors with radiologically different bile duct types (type 1 and type 2) required a different number of anastomoses (83.3%, and 48.2%, respectively, $P < 0.001$) (Table 4).

Discussion

Right-left hepatic duct bifurcation and length, the total number of bile ducts joining the CHD, and the way the cystic duct joins to the CHD (directly or in the form of the cystohepatic duct) were evaluated in this study, and all cases could be classified as a specific type. A total of 83% of the donors who were predicted to require a single anastomosis (type 1) were found to require this single procedure during transplantation surgery.

Table 2. Distribution of the donors according to the Couinaud, Huang, and Karakas classifications

Couinaud	n (%)
A	106 (52.7)
B	19 (9.5)
C1	22 (10.9)
C2	22 (10.9)
D1	10 (4)
D2	4 (2)
E1	3 (1)
E2	3 (1)
F	4 (1.5)
Unclassified	8 (4)
Huang	
1	106 (52.7)
2	21 (10.4)
3	33 (16.4)
4	30 (14.9)
5	4 (2)
Unclassified	7 (3.5)
Karakas	
K1	28 (13.9)
K2a	77 (38.3)
K2b	21 (10.4)
K3a	21 (10.4)
K3b	11 (5.5)
K4	30 (14.9)
Unclassified	13 (6.5)

In addition, a different surgical technique involving more than one anastomosis was required in >50% of the type 2 cases expected to require additional anastomosis. Only one of the nine participants classified as not being suitable as a donor (type 3) could, in fact, be a donor, because of an urgent transplantation need.

This study aimed to draw attention to distinguishing types 1 and 2 and the relationship between radiological type and the number of anastomoses needed. During surgery, 51.8% of type 2 cases required additional anastomosis. This rate is lower than expected. The main reason for this relates to the tendency of transplantation surgeons to anastomose two separate bile ducts together to reduce the complication risk. Six participants classified as type 1 required two anastomoses. All of these participants underwent

Table 3. Distribution of donors according to the Choi, Ohkubo, and proposed classifications

Choi	n (%)
1	105 (52.2)
2	19 (9.5)
3A	26 (12.9)
3B	35 (17.4)
3C	3 (1.5)
4	2 (1)
5	1 (0.5)
Unclassified	10 (5)
Ohkubo	
A	105 (52.2)
B	19 (9.5)
C	34 (16.9)
D	27 (13.4)
E	2 (1)
F	3 (1.5)
G	3 (1.5)
Unclassified	8 (4)
Proposed classification	
1	58 (28.8)
2	136 (67.7)
3	7 (3.5)
Unclassified	0

two anastomoses because the biliary duct was resected shorter than planned during resection. The main reason for this pertains to the resection technique, which follows the demarcation line that occurs after clamping the hepatic artery and portal vein, with the ultimate goal of protecting the donor.

One of the most critical points in the transplantation process is the evaluation of the donor. This entails identifying the most suitable donor for the recipient and minimizing the risk of complications in the donor. A radiological evaluation in transplantation centers is commenced following a clinical and psychiatric evaluation regarding donor suitability. Multi-detector computed tomography (MDCT) and MRCP are used in the radiological assessment of the donor. The presence of hepatic steatosis, vascular variations (hepatic artery, portal vein, and hepatic vein), and any systemic disease are evaluated using MDCT, whereas volumetric measurements are performed using CT images. The anatomy and variations of the bile ducts are evaluated using MRCP. The donor's biliary tract evaluation for identifying bile duct variations exclude donors with anatomical variations that may cause complications in the recipient after transplantation¹¹ and in planning the surgical resection line and biliary anastomosis technique.^{10,12-14} Therefore, the donor's bile ducts should be investigated using imaging techniques. Among these, MDCT cholangiography and MRCP are non-invasive methods, whereas endoscopic retrograde cholangiopancreatography is an invasive method¹⁵ and is considered the gold standard in biliary imaging. However, this invasive technique can lead to severe complications, such as acute pancreatitis and perforation. For this reason, it is not the preferred option for healthy people such as donor candidates. Because MDCT and MRCP allow for precise definitions of biliary and cystic duct anatomy, they are used preoperatively to identify anatomical variations that require special attention from the surgeon.^{16,17}

In our center, where liver transplantations involving a living donor are performed in high volume, MRCP is used to evaluate bile duct variations. Biliary surgery is complex and challenging. Since the ducts are thin, anastomosis is difficult. In addition, anatomical variations of the biliary tract in the donor alter the surgical technique used for the recipient. The details on anatomy pertain to the field of transplantation radiology, which is a key component of liver transplantation teams. Many classifications, including the Huang, Couinaud, Choi, Ohkubo, and Karak-

Table 4. Association of the donor's biliary anatomy type according to the proposed classification and the number of anastomoses needed during transplantation surgery

Class	Number of anastomoses		
	1	2	P
Type 1	30 (83.3%)	6 (16.7%)	<0.001
Type 2	27 (48.2%)	29 (51.8%)	

as classifications, describe the anatomy of the biliary tract. The first known study of bile duct variation was published in 1957 by Couinaud et al.⁶, an anatomist and surgeon. According to this classification, the liver consists of eight distinct segments, each with its own portal venous supply and hepatic venous drainage system. In each segment, the biliary drainage system is parallel to the portal venous supply.¹⁸ The RHD drains segments of the right liver lobe (V–VIII) and has two main branches: the right posterior hepatic duct (RPHD), which drains the posterior segments, and the right anterior hepatic duct (RAHD), which drains the anterior segments. The RPHD tends to have an almost horizontal course, whereas the RAHD tends to have a more vertical course. The RPHD usually passes behind the RAHD and combines it with a left (medial) approach to form the RHD. The LHD is formed by segmental branches that drain segments II–IV. The RHD and LHD unite to form the CHD. The bile duct draining the caudate lobe usually joins the origin of the LHD or RHD. This normal biliary anatomy is thought to be present in 58% of the population.¹⁹ The Couinaud classification does not consider the accessory ducts; however, 2%–6% of the population have an accessory canal.⁹ Identification of accessory ducts is essential for liver resections and biliary drainage.^{20–23} Unmentioned accessory ducts may be a source of biliary leakage or cholangitis. The basis for our proposed classification is the number of ducts joining the CHD. This is crucial in predicting the surgical techniques and reducing complications. Some variations are not included in Couinaud's classification.

One of the most widely used classifications is the Huang classification, the basis of which is the drainage site of the RPHD.⁷ Although this classification is widely used, its major limitation is that it only evaluates variations of the right biliary tract. In the study by Choi et al.⁹, 300 consecutive donors who underwent intraoperative cholangiography during liver transplantation were examined, and anatomical variations of the intrahepatic bile ducts, their frequency, and their branching patterns were defined. The classification was performed according to the RAHD and RPHD branching pattern, presence of a first-degree branch of the RHD, and presence of an accessory hepatic duct. In the study results, type 1, representing the classical anatomy, was detected in 63% of donors, whereas a variation was observed in the remaining 37%, and 1% could not be classified. The key limitation of this study is that the imag-

ing was intraoperative and invasive. For this reason, the method is unsuitable for routine donor evaluation and has the potential to induce severe complications in donors who undergo this procedure.

The classification proposed in the present study was performed using MRCP, a non-invasive imaging technique that does not require contrast material or cause ionizing radiation exposure. All donors could be classified into a specific type. In short, when the two classifications are compared, our classification is more straightforward and inclusive and uses a non-invasive imaging test.

The purpose of the Ohkubo classification is to evaluate the anatomical variations of the biliary tree in patients undergoing liver transplantation involving a living donor. In this study, the junctional patterns of the intrahepatic bile ducts in the hepatic hilum were evaluated following an examination of extrahepatic bile duct resection and major hepatectomy surgical materials in 165 patients (right-sided hepatectomy in 110 patients and left-sided hepatectomy in 55 patients). A key message of this study is that anatomical variability is a rule rather than an exception in liver surgery. In addition, the importance of correctly identifying biliary tract variations for successful LDLT was emphasized. The segment 4 duct is always considered to be single in the Ohkubo classification.¹⁰ However, more than one accessory duct may join from segment 4. Thus, a significant disadvantage of the Ohkubo classification is that it is impossible to classify a donor in cases where two or more segment 4 ducts joining to the LHD separately are present. In this study, since the number of all channels participating in the CHD was considered, all variations of segment 4 ducts could be included in the classification, meaning there is no unclassified variation.

The classification devised by Karakas et al.⁸ is for liver transplantation involving a living donor, with the donor's compliance with standard surgical techniques evaluated using MRCP, as in the present study. The classification proposed in the present study is, as with the Huang classification, based on the relationship between the RPHD and CHD. However, RHD and LHD lengths have been added to the Huang classification, with those below 1 cm labeled as another subtype, although they are morphologically similar. Length is a further parameter used in the present study. However, in our classification, if the duct length is <5 mm, it is considered

type 2 because this length is generally sufficient for anastomosis, especially in right-lobe donors. However, since this length is affected by the transverse diameter of segment 4 in left-lobe donors, it may not be sufficient for anastomosis in these donors, which is one of the limitations of our proposed classification.

This study has a number of limitations. Crucially, it was impossible to demonstrate that a type 3 donor, classified as unsuitable for surgery, was genuinely unsuitable. One donor of this type, whose donation was accepted under emergency conditions, was operated on, and the recipient required three anastomoses. However, this also supports our results because the increased number of anastomoses is a factor that increases the risk of surgical complications. Another limitation is the retrospective design of the study. The inherent disadvantage of MRCP is that it does not reveal very thin bile ducts because of the low spatial resolution and lack of contrast material. Furthermore, artifacts secondary to intestinal peristalsis and respiration can reduce image quality. In addition, although the study population comprised healthy adults, it may not reflect the general population. Another disadvantage of our study is that only healthy donors were evaluated, and the classification could only be used for liver transplantation involving living donors. There is a need to improve the study and classification in terms of evaluating oncological patient groups. Although a power analysis was performed, the study involves a limited number of cases because the biliary system has an excessive variation. However, all donors in this study could be classified according to our classification, and there is no variation that does not fit into the three types. Ultimately, the proposed classification was inclusive, and there was a high correlation between the number of ducts counted using this classification system and the surgical technique.

In conclusion, the classification proposed in this study includes all anatomical variations. This inclusive classification accurately predicts the surgical technique for LDLT.

Acknowledgments

We extend our thanks to Taha Selbasan and Esin Gündoğmuş for Figure 1.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Adam R, McMaster P, O'Grady JG, et al. Evolution of liver transplantation in Europe: report of the European Liver Transplant Registry. *Liver Transpl.* 2003;9(12):1231-1243. [\[CrossRef\]](#)
- Pascher A, Sauer IM, Walter M, et al. Donor evaluation, donor risks, donor outcome, and donor quality of life in adult-to-adult living donor liver transplantation. *Liver Transpl.* 2002;8(9):829-837. [\[CrossRef\]](#)
- Akyol Y. To assess the role of multislice CT in the evaluation of liver parenchyma, hepatic arterial, portal and venous anatomy, the detection of variations and liver volume measured by manual volumetry in potential liver donors. Istanbul Faculty of Medicine, Radiology and Nuclear Medicine, Master thesis, 2007. [\[CrossRef\]](#)
- Brandhagen D, Fidler J, Rosen C. Evaluation of the donor liver for living donor liver transplantation. *Liver Transpl.* 2003;9(10 Suppl 2):16-28. [\[CrossRef\]](#)
- Valls C, Alba E, Cruz M, et al. Biliary complications after liver transplantation: diagnosis with MR cholangiopancreatography. *AJR Am J Roentgenol.* 2005;184(3):812-820. [\[CrossRef\]](#)
- Couinaud C, Delmas A, Patel J. Le foie: études anatomiques et chirurgicales. Paris: Masson & Cie; 1957:530. [\[CrossRef\]](#)
- Huang TL, Cheng YF, Chen CL, Lee TY. Variants of the bile ducts: clinical application in the potential donor of living-related hepatic transplantation. *Transplant Proc.* 1996;28(3):1669-1670. [\[CrossRef\]](#)
- Karakas HM, Celik T, Alicioglu B. Bile duct anatomy of the Anatolian Caucasian population: Huang classification revisited. *Surg Radiol Anat.* 2008;30(7):539-545. [\[CrossRef\]](#)
- Choi JW, Kim TK, Kim KW, et al. Anatomic variation in intrahepatic bile ducts: an analysis of intraoperative cholangiograms in 300 consecutive donors for living donor liver transplantation. *Korean J Radiol.* 2003;4(2):85-90. [\[CrossRef\]](#)
- Ohkubo M, Nagino M, Kamiya J, et al. Surgical anatomy of the bile ducts at the hepatic hilum as applied to living donor liver transplantation. *Ann Surg.* 2004;239(1):82-86. [\[CrossRef\]](#)
- Champetier J. Les voies biliaires. In: Chevrel JP, ed. Anatomie clinique, Le tronc. Paris: Springer; 1994:416. [\[CrossRef\]](#)
- Kim RD, Sakamoto S, Haider MA, et al. Role of magnetic resonance cholangiography in assessing biliary anatomy in right lobe living donors. *Transplantation.* 2005;79(10):1417-1421. [\[CrossRef\]](#)
- Limanond P, Raman SS, Ghobrial RM, Busuttil RW, Lu DS. The utility of MRCP in preoperative mapping of biliary anatomy in adult-to-adult living related liver transplant donors. *J Magn Reson Imaging.* 2004;19(2):209-215. [\[CrossRef\]](#)
- Larobina M, Nottle PD. Extrahepatic biliary anatomy at laparoscopic cholecystectomy: is aberrant anatomy important? *ANZ J Surg.* 2005;75(6):392-395. [\[CrossRef\]](#)
- Erden A, Kuru Öz D, Erden İ. Ectopic opening of common bile duct into the duodenal bulb: magnetic resonance cholangiopancreatography findings. *Diagn Interv Radiol.* 2022;28:286-293. [\[CrossRef\]](#)
- Turner MA, Fulcher AS. The cystic duct: normal anatomy and disease processes. *Radiographics.* 2001;21(1):3-22. [\[CrossRef\]](#)
- Hirao K, Miyazaki A, Fujimoto T, Isomoto I, Hayashi K. Evaluation of aberrant bile ducts before laparoscopic cholecystectomy: helical CT cholangiography versus MR cholangiography. *AJR Am J Roentgenol.* 2000;175(3):713-720. [\[CrossRef\]](#)
- Gazelle GS, Lee MJ, Mueller PR. Cholangiographic segmental anatomy of the liver. *Radiographics.* 1994;14(5):1005-1013. [\[CrossRef\]](#)
- Puente SG, Bannura GC. Radiological anatomy of the biliary tract: variations and congenital abnormalities. *World J Surg.* 1983;7(2):271-276. [\[CrossRef\]](#)
- Deka P, Islam M, Jindal D, Kumar N, Arora A, Negi SS. Analysis of biliary anatomy according to different classification systems. *Indian J Gastroenterol.* 2014;33(1):23-30. [\[CrossRef\]](#)
- Mazroua JA, Almalki YE, Alaa M, et al. Precision mapping of intrahepatic biliary anatomy and its anatomical variants having a normal liver using 2D and 3D MRCP. *Diagnostics (Basel).* 2023;13(4):726. [\[CrossRef\]](#)
- El Hariri M, Riad MM. Intrahepatic bile duct variation: MR cholangiography and implication in hepatobiliary surgery. *Egypt J Radiol Nucl Med.* 2019;50(78):1-8. [\[CrossRef\]](#)
- Sureka B, Bansal K, Patidar Y, Arora A. Magnetic resonance cholangiographic evaluation of intrahepatic and extrahepatic bile duct variations. *Indian J Radiol Imaging.* 2016;26(1):22-32. [\[CrossRef\]](#)



Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions

Tugba Akinci D'Antonoli¹

Arnaldo Stanzione²

Christian Bluethgen³

Federica Vernuccio⁴

Lorenzo Ugga²

Michail E. Klontzas⁵

Renato Cuocolo⁶

Roberto Cannella⁷

Burak Koçak⁸

¹Cantonal Hospital Baselland, Institute of Radiology and Nuclear Medicine, Liestal, Switzerland

²University of Naples Federico II, Department of Advanced Biomedical Sciences, Naples, Italy

³University Hospital Zurich, University of Zurich, Institute for Diagnostic and Interventional Radiology, Zurich, Switzerland

⁴University Hospital of Padova, Department of Radiology, Padova, Italy

⁵Department of Medical Imaging, University Hospital of Heraklion, Crete, Greece & Department of Radiology, University of Crete, Heraklion, Crete, Greece & Computational Biomedicine Laboratory, Institute of Computer Science, FORTH, Heraklion, Crete, Greece

⁶University of Salerno, Department of Medicine, Surgery and Dentistry, Baronissi, Italy

⁷University of Palermo, Department of Biomedicine, Neuroscience and Advanced Diagnostics, Section of Radiology, Palermo, Italy

⁸University of Health Sciences, Başakşehir Çam and Sakura City Hospital, Clinic of Radiology, İstanbul, Türkiye

Corresponding author: Tugba Akinci D'Antonoli

E-mail: akincidantonoli@unibas.ch

Received 03 August 2023; revision requested 31 August 2023; accepted 18 September 2023.



Epub: 03.10.2023

Publication date: 05.03.2024

DOI: 10.4274/dir.2023.232417

ABSTRACT

With the advent of large language models (LLMs), the artificial intelligence revolution in medicine and radiology is now more tangible than ever. Every day, an increasingly large number of articles are published that utilize LLMs in radiology. To adopt and safely implement this new technology in the field, radiologists should be familiar with its key concepts, understand at least the technical basics, and be aware of the potential risks and ethical considerations that come with it. In this review article, the authors provide an overview of the LLMs that might be relevant to the radiology community and include a brief discussion of their short history, technical basics, ChatGPT, prompt engineering, potential applications in medicine and radiology, advantages, disadvantages and risks, ethical and regulatory considerations, and future directions.

KEYWORDS

Large language models, natural language processing, artificial intelligence, deep learning, ChatGPT

Radiology is one of the most technology-driven medical specialties and has always been closely linked to computer science. In particular, ever since the picture archiving and communication system (PACS) revolution, there have been many examples of emerging new technology that has shaped and reshaped the day-to-day practice of radiologists.¹ More recently, the scientific community has witnessed the remarkable progress of artificial intelligence (AI), and the advances in image-recognition tasks are likely to herald another significant leap forward for radiology practice.² There are potential applications of AI in almost the entire radiology workflow, such as image quality improvement (e.g., reducing image acquisition time and/or radiation dose), image post-processing (e.g., image annotation and segmentation), and image interpretation (e.g., prediction of diagnosis).³ With the advent of natural language processing (NLP) and especially with the development of large language models (LLMs), it is becoming clear that AI applications are not limited to imaging-related tasks in radiology, and LLMs have a potential impact in radiology, as radiologists mainly provide textual reports comprising their interpretations of diagnostic images and their clinical significance.

The origins of LLMs date back to the 1950s, a pivotal decade that witnessed the establishment of AI as an academic discipline and the successful demonstration of machine translation through the Georgetown-IBM experiment.⁴ Before delving into the significant milestones that have led to the remarkable technology of today, it is imperative to establish definitions and introduce key concepts. In essence, a language model is a computer program designed to process human language that varies in size and complexity from small rule-based systems to sophisticated AI-driven models. On the other hand, LLMs represent an exceptional class of language models distinguished by their scale, complexity, and emergent capabilities not found in their smaller-scale counterparts.⁵ These models, built on deep learning architectures and trained on vast data with billions of parameters, excel in a diverse range of NLP tasks, such as summarization, translation, sentiment analysis, and text generation. Put simply, LLMs predict the next word or token in a given sequence of words.

You may cite this article as: Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagn Interv Radiol.* 2024;30(2):80-90.

Among the earliest examples of language models was one of the first “chatbots,” coded in the 1960s and named ELIZA, which was based on a set of predefined rules and used pattern matching to simulate human conversation.⁶ Although ELIZA and the other early language models were limited in their capabilities and struggled to handle the complexity and nuances of human language, research in the field of NLP had begun, and the interest continued to grow.

The breakthrough in LLMs occurred in the 1990s with the emergence of the internet and enhanced computational capabilities, facilitating access to extensive text corpora for training datasets. Notably, the introduction of the long short-term memory (LSTM) network in 1997 can be regarded as a turning point for precursors to present-day LLMs.⁷ The pace of technological advancement gained further momentum, culminating in the groundbreaking publication of “Attention Is All You Need” in 2017, which introduced the transformer network architecture.⁸ Subsequently, in 2018, the release of the generative pre-trained transformer (GPT) and the bidirectional encoder representations from transformers (BERT) marked a turning point in the NLP landscape and ushered in the era of LLMs. From there, LLMs have continued to grow in all respects, gaining popularity within the general population as well as the medical community (Figure 1).⁹

This review article provides an overview of the LLMs that might be relevant to the radiology community, with a brief discussion of the technical basics, the ChatGPT revolution, prompt engineering, potential applications

in medicine and radiology, the advantages, disadvantages, and risks, the ethical and regulatory considerations, and future directions. Readers are advised to first refer to Table 1 for definitions of key terms that are used extensively in this discussion of LLMs.

Technical basics of large language models

Language modeling can be technically divided into the following development stages: statistical language models,¹⁰⁻¹² neural language models,^{13,14} and pre-trained language models (PLMs) (Figure 2).^{15,16} The last one is only trained once with unsupervised learning methods (i.e., they learn patterns from unlabeled data) on a massive amount of text data and can be used for a variety of tasks without being retrained from scratch.¹⁵ With capabilities of zero-shot and few-shot learning, PLMs can generalize and adapt to new tasks and data with no or minimal additional training.¹⁷⁻¹⁹ Research has shown that scaling PLMs in terms of data or model size frequently improves the performance of the model on downstream tasks.²⁰⁻²² These large-sized PLMs then exhibit surprising behavioral differences from smaller PLMs and demonstrate emergent abilities in solving several complex tasks, such as in-context learning, instruction following, and step-by-step reasoning.^{5,23} These large-sized PLMs can produce the desired results through in-context learning without the need for extra training or gradient updates and provide outputs for new tasks with instructions, without providing explicit examples. Thus, the research community coined the term LLMs for these massive PLMs that can contain hundreds of billions of parameters.^{24,25}

Key concepts in LLMs are shown and explained in Figure 3. LLMs are typically based on transformer architecture, which is highly parallelizable from a computational standpoint.²⁶ Transformers are essentially composed of encoders and decoders, each of which has a particular attention mechanism.⁸ The attention mechanism is simply a dot product operation to obtain similarity scores by which it enables the model to pay more attention to some inputs than to others, regardless of their position in the input sequence, and enables the model to comprehend the context of a word better. Furthermore, in contrast to recurrent neural networks, the attention mechanism permits the model to view the entire sentence or even the entire paragraph at once, rather than one word at a time.

For a simple transformer model (Figure 4), a text input, such as a sentence or paragraph, must be tokenized (i.e., split into smaller units) for further processing (Figure 5).²⁷⁻²⁹ These tokens are then encoded numerically and transformed into embeddings (i.e., vector representations that maintain meaning). In addition, the order of the words in the input is positionally encoded. Using these embeddings of all tokens along with position information, the encoder within the transformer then generates a representation. The positionally encoded input representation and output embeddings are processed by the decoder so that output can be generated based on these clues (e.g., an initial input or a new word that was previously generated). During training, the decoder learns how to predict the next word based on the previous words. To accomplish this, the out-

Main points

- A language model is a computer program for processing human language, ranging in size and complexity from small rule-based systems to sophisticated models driven by artificial intelligence (AI).
- Large language models (LLMs) are usually based on a transformer architecture with a particular attention mechanism.
- Two recent accomplishments, namely ChatGPT and GPT-4, have significantly raised the bar for the capabilities of existing AI systems.
- LLMs have proven to be successful in many tasks in radiology; however, further studies are required to investigate the feasibility of their use in medical imaging.
- Unresolved ethical and legal issues should be addressed before LLMs are implemented within radiology practice.

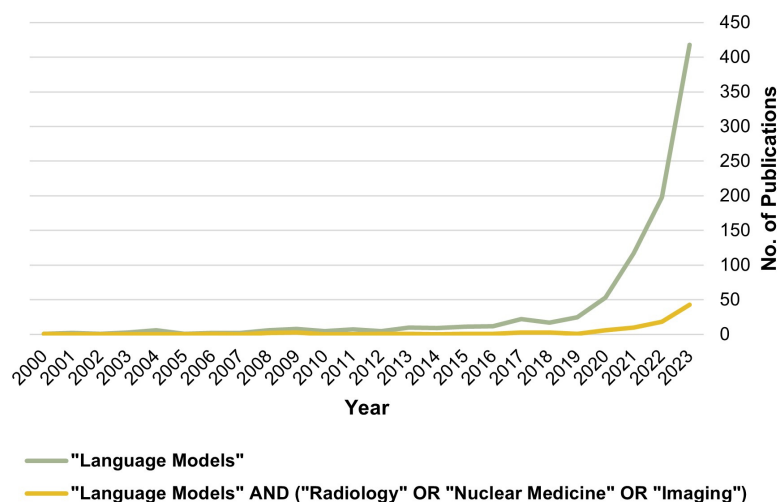


Figure 1. Number of publications about language models in medical publications (green line) and medical imaging (yellow line) including radiology and nuclear medicine. Search date, July 20th, 2023; source, PubMed.

Table 1. Key terminology for large language models	
Terms	Explanations
Application programming interface	An interface that offers a service to other software programs
Attention mechanism	A mechanism that allows the models to focus on certain parts of the input data
Autoregressive	An automatic prediction strategy in which output depends on its own previous values
Chatbot	An application aiming to mimic human-like conversation through text or voice interactions
Decoder	In transformers, it uses the features to produce an output
Deterministic	No randomness is involved
Embedding	Mathematical representation of data (e.g., a word in the form of a string of numbers or vector representation)
Encoder	In transformers, it extracts features from input data
Few-shot learning	A method for making predictions based on a limited number of samples
Few-shot prompting	A task text with examples as context
Fine-tuning	Adjusting models to achieve improved performance on domain-specific tasks
Generative	Algorithms that can create new content
Hallucinations	Fabricated false information generated by models
Long short-term memory networks	A type of recurrent neural network designed to handle sequential data
Natural language processing	Use of machine learning to interpret text
Parallelizable	Being able to do several computations or processes simultaneously, such as those performed by graphics processing units
Pre-trained	A model trained once and intended to be used for a lot of different tasks without re-training from scratch
Prompt	A text that is used to initiate the model's text generation process
Recurrent neural network	A neural network architecture that can deal with sequential data and hold information about the past
Reinforcement learning	A training method based on rewarding and/or punishing
Scaling	Capability of a model to handle increasing amounts of data, workload, or users effectively and efficiently
Stochastic	Random
Stochastic parrots	Irrelevant repetition of information existing in their training data
Token	Basic units of text or code (e.g., word, sub-word)
Transformer	A specific type of self-supervised encoder-decoder deep neural network architecture that can transform one type of input into another type to produce an output, with an attention mechanism
Unsupervised learning	Learning patterns from unlabeled data
Zero-shot learning	A method used for a pre-trained model to classify data according to a new set of labels that were not used to train the model previously
Zero-shot prompting	A task text that should be followed by an answer, without giving more context or examples

put sequence is shifted to the right by one position; thus, the decoder can only utilize the preceding words. After the decoder generates the output embeddings, the linear layer transforms them into the original input space by mapping them to a higher-dimensional space. Then, the softmax function is used to generate a probability distribution for each output, enabling the generation of probabilistic output tokens. This procedure is known as autoregressive generation and is repeated to produce the entire output. Notably, although LLMs are consistent, they are not deterministic but stochastic, meaning they can generate different answers for the same query.³⁰ This is because the model returns a probability distribution over all possible tokens and draws samples from this distribution to produce the output token.

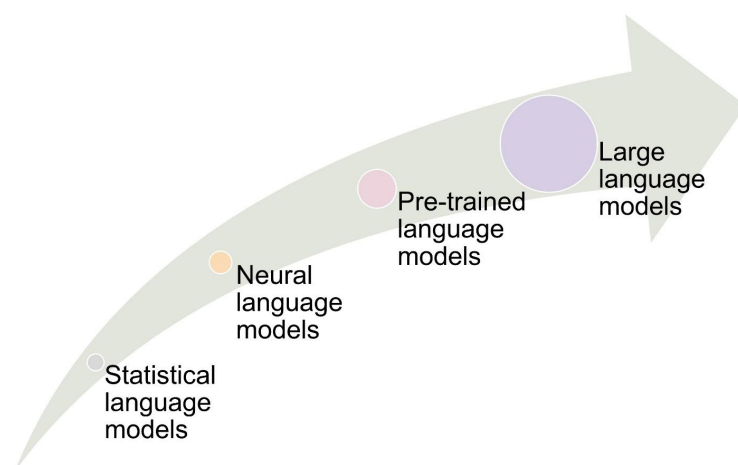


Figure 2. Technical developmental stages of language models.

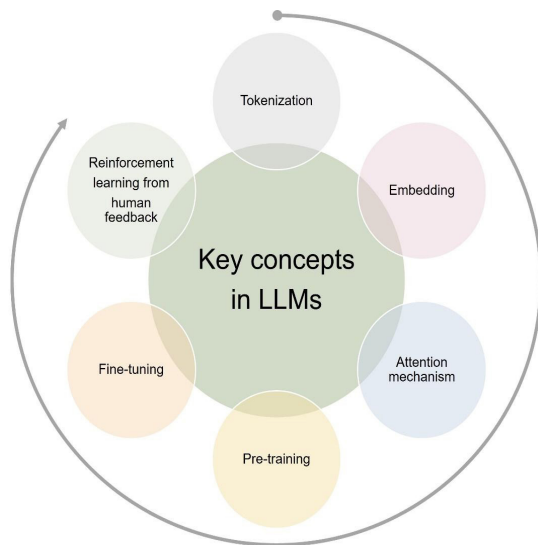


Figure 3. Key concepts in LLMs. Tokenization is the process of splitting text into smaller units (i.e., tokens) that can be processed by language models. Embedding is the mathematical representation of data (e.g., vector representation of a word). The attention mechanism allows the models to focus on certain parts of the input data. Pre-training is the training of a model to be used for a lot of different tasks without re-training from scratch. Fine-tuning is the adjustment of models to achieve improved performance on domain-specific tasks. Reinforcement learning from human feedback is a machine learning approach based on reinforcement learning techniques along with human guidance. LLMs, large language models.

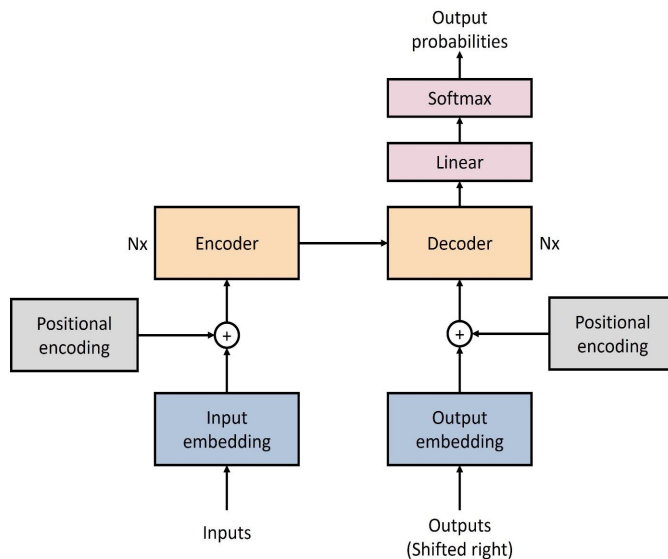


Figure 4. Architecture of transformers. The encoder and decoder are overly simplified in the figure. Both normally include attention mechanisms, feed-forward neural networks, residual connections, and the normalization layer. Transformers utilize multiple layers of encoders and decoders. Nx, number of layers of encoder and decoder parts.

You are reading a paper from Diagnostic and Interventional Radiology.

[1639, 389, 3555, 257, 3348, 422, 31549, 15132, 290, 4225, 20405, 5325, 12371, 13]

Figure 5. Tokenization example. A 10-word sentence with one punctuation sign is tokenized to 14 tokens as shown in the upper panel. The bottom panel shows token identifiers unique to each token. Generated by OpenAI's Tokenizer platform (<https://platform.openai.com/tokenizer/>).

The masked multi-head attention layer is a crucial component that distinguishes the transformer model from the simple encoder-decoder architecture described above.⁸ The attention layer contains the weights learned during training that represent the strength of the relationship between all token pairs in the input sentence. This mechanism guarantees that each token has a direct connection to all tokens that came before it. This is a great achievement considering the gradient issues of older architectures such as recurrent neural networks and LSTM networks, specifically the difficulties in recalling previous tokens when two tokens are far apart.^{31,32} The attention layer is masked, such that the model can only focus on previous tokens or positions in the input sequence. This restriction ensures that the model cannot access information about future tokens, which could result in data leakage or violate the causality of the sequence (i.e., the effects of one part of a sequence on another). The transformer employs a multi-head attention layer because it contains multiple parallel attention layers.

It is important to note that LLMs can use external tools (e.g., calculators, image readers, search engines) to perform tasks that are not best expressed in the form of text (e.g., numerical computation) or to overcome the limitation of being trained on old data that prevents them from capturing current or external information.³³ Furthermore, LLMs can also be used within external tools or applications (e.g., LangChain), which can significantly expand the capabilities of LLMs.

ChatGPT revolution and basics

At the time of writing, the latest text generation tools released by OpenAI are GPT-3.5, GPT-4, and ChatGPT. All these tools are based on the transformer architecture, as the acronym, GPT, indicates. Considering all previous efforts in LLMs, ChatGPT and GPT-4 are two notable accomplishments that have significantly raised the bar for the capabilities of existing AI systems.³⁴ The GPT-3.5 model is a fine-tuned version of the GPT-3 model and was trained as a completion-style model, meaning it can generate relevant words that follow the input words. On the other hand, GPT-4 has an entirely new large multimodal model and is also adjusted with reinforcement learning with human feedback (RLHF) to better align with human expectations.³⁴ Extending text input to multimodal signals is regarded as a significant development. Overall, GPT-4 is superior to GPT-3.5 in its ability to solve complex tasks, as evidenced by a signif-

icant performance increase on various evaluation tasks.³⁵ ChatGPT, based on GPT-3.5 and GPT-4, was optimized for creating conversational responses (i.e., as a conversation-style model) and further fine-tuned using RLHF,³⁶ allowing it to provide human-like responses to user queries or questions. With RLHF, the outputs were ranked by humans, and a reward system was used to improve the model to align the output model based on human expectations, which might be critical for their success, sparking the interest of the AI community ever since its debut because of its exceptional potential for human communication. The implementation of ChatGPT in conversational-style interactions opens up a universe of opportunities for human-computer interaction. Its capacity to comprehend context, create logical responses, and maintain conversational flow makes it a viable tool for a vast array of domains and use cases, such as customer support, brainstorming, content generation, and tutoring. Furthermore, ChatGPT now supports the plugin mechanism, which expands its compatibility with existing tools and applications.³³

Despite the tremendous progress, there remain limitations with these superior LLMs, such as producing “hallucinations” (i.e., fabrication of facts), factual errors, potentially

risky responses in certain contexts, variable source reporting, or changing behaviors or drifts.^{34,37-41} Due to these limitations, they should be used cautiously. The risks related to LLM use are extensively discussed later in this review.

These models could also be used in coding environments and as part of other applications via application programming interfaces (API). Currently, the main issues are token limits and the high usage fees for ChatGPT and various GPT APIs.

Prompt engineering

In the context of LLMs, a prompt is an input provided to the model to steer its output. These prompts are often sequences constructed from natural language but can also be other types of structured information. The prompt’s syntax (e.g., structure, length, ordering) and semantic contents (e.g., words, tone) have a significant impact on the outputs of LLMs.⁴² This poses a challenge, as even slight modifications can lead to substantially different results (“prompt brittleness”).⁴³

Prompt engineering is an emerging field of research that attempts to design prompts that steer LLMs toward a desired output.

In contrast to other methods (e.g., pre-training, fine-tuning), this way of influencing the outputs does not involve updating the weights of LLMs, thus leaving the underlying model unchanged. The currently limited theoretical understanding of why some prompts work better than others makes it challenging to design effective prompts empirically. Therefore, “prompt engineers” often have to resort to extensive experimental work for specific use cases.

A multitude of prompting techniques have been developed (Table 2).^{43,44} The most basic prompts provide a task text that should be followed by an answer, without giving more context or examples (i.e., zero-shot prompting). In-context learning (often an example of few-shot prompting) refers to providing examples of desired input-output pairs in the input prompt (e.g., questions and corresponding answers from the training data) together with a new question that the LLM should respond to following the provided examples. Instruction following requires an LLM that was fine-tuned in a supervised way to follow instructions (e.g., ChatGPT). These types of LLMs can be provided with instructions and one or more examples (similar to in-context learning). Chain-of-thought prompting refers to a strategy of breaking

Table 2. Examples of different prompting techniques aiming for the same answer (e.g., “blue” in the following examples) and resulting outputs generated by different large language models (OpenAI text-ada-001, text-davinci-003; with default parameters, i.e., temperature: 1, top $P = 1$)

Prompting technique	Example input	Example output (model: text-ada-001)	Example output (model: text-davinci-003)
Zero-shot prompting	Question: What color is the sky? Answer: [Output]	The sky is blue.	The sky is usually blue, although it can appear to be a different color depending on the time of day or the weather.
Few-shot prompting	Question: What color is human blood? Answer: red Question: What color is grass? Answer: green Question: What color is the sky? Answer: [Output]	The sky is a different color than human blood.	Blue.
Instruction following	Please answer the following question to determine the usual color of the object that is mentioned. Question: What color is the sky? Answer: [Output]	The sky is blue.	Blue.
Chain-of-thought prompting	Question: What color is the sky at 12:00 pm? Let’s think step-by-step and explain your thoughts. Answer: [Output]	The sky is blue at 12:00 pm.	Step 1: Consider the time of day. At 12:00 pm it is likely to be midday so the sun will be in the highest point in the sky. Step 2: Consider the weather conditions. Generally, the sky is blue during the day, although it can be affected by weather conditions. If there is cloud cover or rain, then the sky may be affected by this. Step 3: Look around and observe the sky. At 12:00 pm, the sky is likely to be a bright blue color!

Outputs provided in the table are model version specific and are only intended as examples. Due to the stochastic nature of large language models, they can generate different answers for the same query if readers try the same input.

down a task into smaller logical subtasks, which can empirically improve the performance of LLMs.²⁵ One simple way to steer the LLM in this direction is to provide the instruction “let’s think step-by-step.”¹⁹ The Tree-of-Thoughts framework is an example of multi-turn prompting that extends this approach by considering multiple reasoning possibilities at each step.⁴⁵

Prompt engineering could also play a valuable role in radiology-specific tasks such as report structuring, summarization, or lan-

guage translation (Table 3). Nevertheless, its true value requires further exploration. Initial results suggest that for tasks such as report summarization, domain adaptation through lightweight fine-tuning may outperform various in-context prompting approaches.⁴⁶ A promising research direction involves enriching initial prompts with information retrieved from external sources (e.g., through API calls to other models, tools, and databases) to augment the capabilities of LLMs and increase the correctness of their outputs.^{33,47,48}

Potential applications in medicine

The application of LLMs is expected to transform medical practice in all fields and in numerous ways. First, LLMs may potentially assist students during their medical training, by providing nonobvious and logical insights into explanations and role-modeling a deductive reasoning process.⁴⁹ Second, LLMs can rapidly develop specialized knowledge for different medical disciplines and generate answers to clinical questions by analyzing large amounts of medical data, and with

Table 3. A practical example to illustrate how prompt engineering can be used to structure a free-text radiology report and also translate the report into different languages (OpenAI API GPT-4)

Example prompt	Output part 1	Output part 2
You are a medical writer specialized in healthcare communications and radiology, given the task to structure the report from the following unstructured radiology report in the following order: indication, technique, liver, gallbladder, pancreas, adrenal gland, spleen, kidneys, vascular. Then translate this structured report into German and print it separately. Radiology report: Indication: Abdominal pain. No history of previous abdominal operations. Technique: Transabdominal ultrasonography. Findings: The pancreas is normal in echotexture and demonstrates no evidence of a dilated pancreatic duct. The liver is normal in echotexture and size. No evidence of dilated intrahepatic ducts. The gallbladder has no evidence of intraluminal echogenicity and the wall measures 2 mm in diameter (<3 mm is normal). The common hepatic duct measures 5 mm in diameter (<6 mm is normal). The hepatic and portal veins demonstrate normal directional flow. The right and left kidneys measure 10.5 cm and 11 cm in length, respectively. No hydronephrosis. No nephrolithiasis. The spleen is normal in echotexture and size. The inferior vena cava measures 11 mm in diameter. The aorta measures 29 mm in diameter. No ascites. Impression: Normal abdominal ultrasound.	Structured radiology report: Indication: Abdominal pain. No history of previous abdominal operations. Technique: Transabdominal ultrasonography. Liver: - The liver is normal in echotexture and size. - No evidence of dilated intrahepatic ducts. Gallbladder: - The gallbladder has no evidence of intraluminal echogenicity. - The wall measures 2 mm in diameter (<3 mm is normal). - The common hepatic duct measures 5 mm in diameter (<6 mm is normal). Pancreas: - The pancreas is normal in echotexture. - No evidence of dilated pancreatic duct. Adrenal Gland: - No specific mention in the original report. Spleen: - The spleen is normal in echotexture and size. Kidneys: - The right kidney measures 10.5 cm in length. - The left kidney measures 11 cm in length. - No hydronephrosis. - No nephrolithiasis. Vascular: - The hepatic and portal veins demonstrate normal directional flow. - The inferior vena cava measures 11 mm in diameter. - The aorta measures 29 mm in diameter. Impression: - Normal abdominal ultrasound.	Translation in German: Strukturierter Radiologiebericht: Indikation: Bauchschmerzen. Keine Vorgeschichte einer früheren Bauchoperationen. Technik: Transabdominelle Ultraschalluntersuchung. Leber: - Die Leber hat eine normale Echotextur und Größe. - Kein Hinweis auf erweiterte intrahepatische Gänge. Gallenblase: - Die Gallenblase zeigt keine intraluminale Echogenität. - Die Wand misst 2 mm im Durchmesser (<3 mm ist normal). - Der gemeinsame Lebergang misst 5 mm im Durchmesser (<6 mm ist normal). Bauchspeicheldrüse: - Die Bauchspeicheldrüse hat eine normale Echotextur. - Kein Hinweis auf erweiterte Bauchspeicheldrüsengänge. Nebenniere: - Keine spezifische Erwähnung im ursprünglichen Bericht. Milz: - Die Milz hat eine normale Echotextur und Größe. Nieren: - Die rechte Niere misst 10.5 cm in der Länge. - Die linke Niere misst 11 cm in der Länge. - Keine Hydronephrose. - Keine Nephrolithiasis. Gefäße: - Die hepatischen und portalen Venen zeigen einen normalen Richtungsfluss. - Die untere Hohlvene misst 11 mm im Durchmesser. - Die Aorta misst 29 mm im Durchmesser. Eindruck: - Normale abdominale Ultraschalluntersuchung.

The adrenals are not mentioned in the free text report and GPT correctly detects this in the structure report. Moreover, there is a typo in the free text report (gallbladder is written with a single “d”) and generative pre-trained transformer (GPT) also corrects this typo in the output. As shown, using large language models for report structuring could reduce the errors of omission and typos. Nevertheless, all outputs should be reviewed by users to make sure they are correct.

the possibility of fine-tuning the generated content based on the most recent published papers, the domain-specific medical literature, and on the reader's background.⁵⁰ In all medical fields, this capability of LLMs could finally translate into enhanced clinical decision support, improved patient engagement, and accelerated medical research.⁵¹⁻⁵³

Regarding enhanced clinical decision support, LLMs are expected to improve diagnostic accuracy and the prediction of disease progression and support clinical decision-making.⁵⁴ As practical examples, the use of PubMedBERT (a pre-trained model based on PubMed abstracts and full-text articles) and ClinicalBERT (a contextual language model trained on PubMed Central abstracts, full-text articles, and fine-tuned on notes from the Medical Information Mart for Intensive Care) resulted in two successful diagnoses: the automatic determination of the presence and severity of esophagitis based on the Common Terminology Criteria for Adverse Events guidelines from notes of patients treated with thoracic radiotherapy,⁵⁵ and the accurate prediction of short-, mid-, and long-term mortality by only using clinical notes within the 24 hours of admission of patients admitted to intensive care units.^{56,57}

With regard to benefits for the patient, LLMs proved to be helpful in providing correct answers to basic questions posed by patients with prostate cancer, rhinologic diseases, and cirrhosis,⁵¹⁻⁵³ and in providing emotional support to patients and caregivers, encouraging proactive steps to manage the diagnosis and treatment strategies.⁵³

Furthermore, LLMs may accelerate medical research by allowing for the identification of high-quality papers within all medical literature, the detection of correlations, and the provision of insights that may aid researchers in accelerating medical advancement.^{58,59}

Moreover, the adoption of LLMs may aid or simplify certain daily tasks, such as text generation, text summarization, and text correction, which can lead to significant time savings and improvements in grammar, readability, and conciseness of written content while maintaining the overall message and context. As an example of their potential in clinical practice, LLMs could output a formal discharge summary in a matter of seconds by analyzing all clinical notes.⁶⁰

Potential applications in radiology

Overall, LLMs have shown promise in several fields, including radiology. They have

proven suitable for a variety of tasks, some of which have already been explored in earlier studies. For example, it has been demonstrated that these models may have a role in patient triage and workflow optimization. Specifically, they can help in the automated determination of the imaging study and protocol based on radiology request forms.⁶¹ In this context, LLMs could be integrated into radiology departments' information technology systems to facilitate patient triage; they could help prioritize imaging studies based on urgency, patient information, and existing imaging data. This could, in turn, streamline the workflow and ensure that critical cases receive prompt attention.

Furthermore, the performance of LLMs in generating impressions from radiology reports has been evaluated. A recent study showed promising results, suggesting the feasibility of LLM use in report generation and summarization, considering coherence, comprehensiveness, factual consistency, and harmfulness.⁶² Another possible use case for LLMs in radiology is their assistance in diagnosis. Indeed, by analyzing the imaging data and considering the patient's medical history, these models can suggest potential diagnoses, differential diagnoses, and possible treatment options.⁶³ In view of this, LLMs could be utilized as AI-powered assistants for radiologists, helping them interpret medical images and providing preliminary assessments.

Moreover, they have proven valuable in answering radiology-related questions, including explanations of specific imaging findings, clarifications regarding radiological procedures, and general information about different types of imaging modalities.^{64,65} Radiologists, trainees, and even patients could interact with these models to obtain answers to questions related to radiology. This aspect is closely linked to the use of LLMs in the context of education and training, as a virtual tutor for radiology residents to understand complex concepts, interpret images, and provide learning resources, fostering self-directed learning and knowledge retention. As evidence of this, it is worth mentioning that, despite no radiology-specific pre-training, ChatGPT almost passed a radiology board-style examination, even when image-based questions were excluded.⁶⁶

In fact, LLMs can be integrated with existing radiology software and systems to assist radiologists in various ways. For example, they can serve as a natural language interface to several radiology tools currently in use.

Radiologists can interact with the system using plain language queries, making it easier to retrieve patient data, reports, and images. By being told to "show me all the MRI reports from last week", the LLM can retrieve and display the relevant information. Furthermore, the LLM can suggest structured report templates and help in ensuring that the report includes all necessary information. Finally, LLMs can be integrated with image analysis tools to provide radiologists with assistance in image interpretation and data extraction and structuring. The LLMs can be customized to fit the specific needs of radiology departments and integrated seamlessly with existing PACS and radiology information systems (RISs). If properly integrated with the electronic health records (EHRs) and RIS, LLMs could automatically identify the radiology reports with recommendations for additional imaging and help ensure the timely performance of clinically necessary follow-ups.

It is important to note that although LLMs can be a valuable tool in radiology, they should complement the expertise of radiologists rather than replace it. One notable issue with ChatGPT is its tendency to maintain unwavering confidence in its responses, even when providing incorrect answers. This characteristic could have adverse consequences in clinical situations.⁶⁷ Ethical considerations, validation, and regulatory compliance are essential aspects to be addressed before deploying AI systems in real-world medical settings. In addition, continuous updating and improvement of the model would be necessary to maintain accuracy and relevance.

Advantages, disadvantages, and risks

There are both advantages and disadvantages of LLMs that are inherent to their structure and capabilities. However, certain aspects are applicable to all LLMs, irrespective of their architecture or application. The most important of these advantages is the fact that they possess advanced NLP capabilities. Advanced language comprehension from LLMs allows the performance of tasks such as text summarization, text translation, and question answering in a manner similar to humans.⁶⁸ Text generated by an LLM is usually free of grammatical mistakes and misspellings, which is important in radiology practice. These NLP capabilities can be applied to radiological reports to convert them into structured text, translate them to other languages, and explain them in a way that is comprehensible to patients.⁶⁹

Another important advantage is that their generative capacity can be used to generate code for medical imaging research. Furthermore, LLMs can be used by people with limited to no coding experience, translating research ideas into useful code.⁷⁰ This code can be used to develop machine learning models for medical imaging research. Combining the NLP capabilities of LLMs and their generative capacity can also allow code debugging and application troubleshooting, enhancing research possibilities in medical image analysis. In the case of the latter, LLMs can be successfully coupled with convolutional neural networks (CNNs) to enable image recognition and the generation of relevant text based on images; CNNs can be used to extract image features, which can be subsequently used by LLMs for image recognition and relevant text generation.⁷¹

Nonetheless, despite the important advantages of LLMs, their use still has significant disadvantages. The most important disadvantage of LLM use in radiological research is related to privacy concerns. Privacy issues can emerge because sensitive patient information can be compromised when uploaded to LLMs.⁷² This important disadvantage can raise ethical concerns when utilizing patient data that includes radiological reports and images. Appropriate data de-identification processes need to be in place to ensure the safe use of patient data in LLMs.

Another disadvantage of LLMs is the possibility of generating information that is artificial and potentially harmful based on their logic (i.e., hallucinations), or the irrelevant repetition of information existing in their training data (i.e., “stochastic parrots”).³⁷ When used to translate reports or to generate information that will be distributed to patients or used to assist diagnostic decisions, the user needs to be extremely careful to avoid cases where LLMs generate fake information. Such fake information can vary from an inaccurate translation of a radiological report to reaching false conclusions related to a disease or a diagnosis. This necessitates the validation of LLM-generated content, especially when used in patient care, a fact that should also be disclosed to patients when receiving such information.³⁷

Given that LLMs can generate fake information, the interpretability and transparency of the models are extremely important. Having the ability to explain why the model has produced a certain output, to identify activated neurons and their weights (interpretability), and to decipher how the model

works, how it is structured, what capabilities and what limitations it has (transparency), are of utmost importance when they are used for medical decision making, as errors can have an impact on patient care. Companies such as OpenAI have attempted to produce tools that enable the interpretability of their models, e.g., GPT-4.⁷³ This can increase the trust of the users in the model output and allow debugging and error identification to ensure that critical errors related to patient management are not repeated.⁷⁴

The quality of an LLM’s output is directly influenced by the information used for its training. To ensure accuracy in LLM responses, the quality and diversity of the training data need to be considered. Therefore, generic LLMs (including GPT-4 and Bard) that have not been trained on medical data may yield inaccurate responses to medically related tasks. On the other hand, medically oriented LLMs such as BioBERT and Med-PaLM2 have been trained on medical data, but the representation of certain information in the model is still unknown.⁷⁵ Moreover, LLMs rely on temporal updates for the training data. For instance, at the time of writing this review, ChatGPT has been trained with data up to September 2021, meaning it can be less reliable when up-to-date medical information is required.^{69,76} With the rapidly evolving knowledge in medicine, this can represent a relevant risk for users and patient care as the LLM may not have access to the latest data and recently published guidelines.⁷⁷

LLMs can be freely used by patients for self-diagnosis or to decode radiological reports. Although LLMs can simplify radiological reports with technical language to a more understandable summary for the patient, there is a risk of overconfidence, with patients not being aware of output errors and assuming that the provided answers are always correct.^{78,79} The risk of missing relevant information in a simplified summary should also be considered in patient care.⁷⁸ The generation of different outputs from the same query can pose a risk of contradictory answers with the difficulty of selecting the correct medical information.⁶⁹

Furthermore, LLMs are not capable of providing ethical insights and evaluating the ethical risks related to the use of the information. The generation of incorrect diagnoses, misinterpretation of the results, or wrong recommendations can induce the risk of medico-legal implications with dangerous information for patient management, which requires specific regulation in the near future.⁸⁰

Last but not least, an important disadvantage of LLMs is the environmental and financial risk of their use. Given that the energy needed to train an LLM can be comparable with that of a trans-Atlantic flight, with energy costs reaching thousands of US dollars,³⁸ the widespread use and training of such models require regulation.

Ethical and regulatory considerations

The recent improvements in LLM performance have also affected the potential use of this technology in healthcare and radiology in particular, with studies proposing novel applications or aimed at demonstrating its medical prowess.^{66,81,82} However, the actual use of LLMs in medical imaging remains controversial due to unresolved ethical and regulatory questions, partly due to inherent technical limitations.⁸³

As with other machine learning models, especially deep learning models, LLMs are highly sensitive to bias embedded within their training data. Although some sources of bias such as age or gender distribution can be easily identified and even addressed, others, such as differences due to the sourcing of the training data, can be less apparent or solvable. For example, most text data used to train LLMs will originate from Western countries and will be written in the English language, simply due to the realities regarding the availability of materials and technology necessary to produce and collect sufficiently large datasets.⁸⁴ Beyond the reduced representation of other areas of the world in this setting, and even within the countries from which this data mainly originates, a lack of fair representation of data produced by all societal components can be expected. Moreover, this imbalance cuts both ways, as the voice of the majority may drown smaller communities, but extremely vocal minorities may also end up being overrepresented within the training data. While efforts to address these issues are ongoing, physicians should be aware that human bias is an integral component of any LLM and should be accounted for rather than ignored.⁸⁵ On the other hand, as more and more data available online are produced by software, from simpler automated bots to LLMs, it is also true that this will also represent a novel source of bias, with the risk of harming the training of future models by further diluting the quality of available data and reducing the models’ ability to meet human needs and expectations.⁸⁶

Regulatory bodies are attempting to address these ethical issues, as well as other limitations of LLMs, such as hallucinations. Both in the United States (US) and the European Union (EU), the use of LLMs in health-care would typically fall under the domain of current medical device regulations.⁸⁷ Even if this prevents their marketing as medical devices, the reality is that LLMs are not currently prevented from answering health-related questions, and the risk of misinformation and even potential harm to patients is not absent. For the EU in particular, it should be noted that medical devices not only require preliminary certification but also continuous surveillance, which poses specific challenges to complex and somewhat unpredictable models such as LLMs.⁸⁸

Future directions

As discussed in the previous sections, there is a huge variety of opportunities to apply LLMs in radiology, and new research is being published every day (Figure 1). All these results already indicate that every aspect of radiology practice will eventually be affected by these new tools.

However, the lack of regulations and ethical uncertainties mean that the rapid implementation of these tools in radiology remains unclear. Regulations should be in place to mitigate potential risks that may be associated with this new technology, and these tools will likely be regulated in the EU and the US in the same way as they are with other clinical decision support tools.⁸⁷ Nevertheless, the ethical issues and their solutions could be use-case specific, which may require ongoing human oversight and is not foreseeable with the current examples.⁹

Some LLMs, such as GPT-4, have shown remarkable potential in various fields and have even signaled that they may carry sparks of artificial general intelligence.³⁵ Looking ahead, we can see some important trends that are likely to shape the future of LLM applications in radiology.

Currently, radiologists must log into EHRs separately to attain more information about the medical history or lab results of patients because the EHR is a separate system from the PACS. Considering that most imaging orders are laconic and do not include a good summary of the medical history, the radiologist usually must switch back and forth between systems, which can be extremely time consuming. This process could be assisted or completely taken over by LLMs, whereby a summary of the patient's history and findings would be presented automatically.⁸⁹

Another application of LLMs is that they could serve as sophisticated clinical decision support systems in which they are fine-tuned with guidelines and recommendations, such as those of the Fleischner Society, and automatically generate evidence-based recommendations from radiology reports, such as follow-up recommendations for solid pulmonary nodules.⁶⁵

Furthermore, LLMs can also play a critical role in training the next generation of radiologists.⁹⁰ Currently, training can be hindered by heavy workload. Through integration into PACS, LLMs could provide a personalized, interactive, and effective learning environment, provide similar examples from the archives to the one the trainee is working on, recommend additional resources for diagnosis, or fully simulate a real clinical scenario to prepare trainees for night shifts.

Although the potential of LLMs in radiology is evident, there are various limitations and problems that must be addressed as research in this subject progresses. One of the most serious issues in using LLMs in medicine is data privacy.^{83,91} To address this issue, continuing research is focusing on building robust approaches for privacy-preserving machine learning.⁹²⁻⁹⁵ The robustness of LLMs, particularly in clinical setting, is a further concern of the utmost importance. These models must consistently and reliably perform across a broad spectrum of demographics, equipment, and scenarios. Ongoing research focuses on enhancing model generalization and minimizing biases to address this issue.^{96,97}

Concluding remarks

Overall, LLMs have the potential to transform the field of radiology, not only in the clinical setting but also in the academic setting. Consequently, radiologists should be familiar with the inner workings and idiosyncrasies of LLMs, such as hallucinations, drifts, and their stochastic nature, as described in this review. Nonetheless, the future of LLMs in radiology appears to be very bright and has the potential to revolutionize patient care, improve outcomes, and enhance radiologists' capabilities. However, these developments should be accompanied by regulations and ethical guidelines to ensure that these tools are used safely and responsibly without compromising patient privacy or data security. The authors hope the overview of the key concepts provided in this article will help improve the understanding of LLMs among the radiology community.

Conflict of interest disclosure

F.V.; none related to this study; received support to attend meetings from Bracco Imaging S.r.l., and GE Healthcare. M.E.K.; meeting attendance support from Bayer. Ro.C.; support for attending meetings from Bracco and Bayer; research collaboration with Siemens Healthcare; co-funding by the European Union - FESR or FSE, PON Research and Innovation 2014–2020 - DM 1062/2021. Burak Koçak, MD, is Section Editor in Diagnostic and Interventional Radiology. He had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Other authors have nothing to disclose.

References

1. Bick U, Lenzen H. PACS: the silent revolution. *Eur Radiol.* 1999;9(6):1152-1160. [\[CrossRef\]](#)
2. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18(8):500-510. [\[CrossRef\]](#)
3. Richardson ML, Garwood ER, Lee Y, et al. Noninterpretive uses of artificial intelligence in radiology. *Acad Radiol.* 2021;28(9):1225-1235. [\[CrossRef\]](#)
4. Ornstein J. Mechanical translation: new challenge to communication. *Science.* 1955;122(3173):745-748. [\[CrossRef\]](#)
5. Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models. *arXiv.* 2022. [\[CrossRef\]](#)
6. Shum HY, He X, Li D. From Eliza to xiaoice: challenges and opportunities with social chatbots. *arXiv.* 2018. [\[CrossRef\]](#)
7. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9(8):1735-1780. [\[CrossRef\]](#)
8. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv.* 2017. [\[CrossRef\]](#)
9. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med.* 2023;29(8):1930-1940. [\[CrossRef\]](#)
10. Gao J, Lin CY. Introduction to the special issue on statistical language modeling. *ACM Trans Asian Low-Resour.* 2004;3(2):87-93. [\[CrossRef\]](#)
11. Chelba C, Mikolov T, Schuster M, et al. One billion word benchmark for measuring progress in statistical language modeling. *arXiv.* 2013. [\[CrossRef\]](#)
12. Rosenfeld R. Two decades of statistical language modeling: where do we go from here? *Proc IEEE.* 2000;88(8):1270-1278. [\[CrossRef\]](#)
13. Bengio Y, Ducharme R, Vincent P, Janvin C. A neural probabilistic language model. *J Mach Learn Res.* 2003;3:1137-1155. [\[CrossRef\]](#)

14. Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S. Recurrent neural network based language model. *Interspeech*. 2010;1045-1048. [\[CrossRef\]](#)
15. Wang H, Li J, Wu H, Hovy E, Sun Y. Pre-trained language models and their applications. *Engineering*. 2022. [\[CrossRef\]](#)
16. Min B, Ross H, Sulem E, et al. Recent advances in natural language processing via large pre-trained language models: a survey. *ACM Comput Surv*. 2023;56(2):1-40. [\[CrossRef\]](#)
17. Wei J, Bosma M, Zhao VY, et al. finetuned language models are zero-shot learners. *arXiv*. 2021. [\[CrossRef\]](#)
18. Gao T, Fisch A, Chen D. Making pre-trained language models better few-shot learners. *arXiv*. 2020. [\[CrossRef\]](#)
19. Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. *arXiv*. 2022. [\[CrossRef\]](#)
20. Chowdhery A, Narang S, Devlin J, et al. PaLM: scaling language modeling with pathways. *arXiv*. 2022. [\[CrossRef\]](#)
21. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. *arXiv*. 2020. [\[CrossRef\]](#)
22. Kaplan J, McCandlish S, Henighan T, et al. Scaling laws for neural language models. *arXiv*. 2020. [\[CrossRef\]](#)
23. Zhao WX, Zhou K, Li J, et al. A survey of large language models. *arXiv*. 2023. [\[CrossRef\]](#)
24. Hoffmann J, Borgeaud S, Mensch A, et al. Training compute-optimal large language models. *arXiv*. 2022. [\[CrossRef\]](#)
25. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. *arXiv*. 2022. [\[CrossRef\]](#)
26. Wolf T, Debut L, Sanh V, et al. Transformers: state-of-the-art natural language processing. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics; 2020:38-45. [\[CrossRef\]](#)
27. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv*. 2018. [\[CrossRef\]](#)
28. Kudo T. Subword regularization: improving neural network translation models with multiple subword candidates. *arXiv*. 2018. [\[CrossRef\]](#)
29. Rust P, Pfeiffer J, Vulić I, Ruder S, Gurevych I. How good is your tokenizer? On the monolingual performance of multilingual language models. *arXiv*. 2020. [\[CrossRef\]](#)
30. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. *OpenAI Blog*. 2019;1(8):9. [\[CrossRef\]](#)
31. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. *arXiv*. 2012. [\[CrossRef\]](#)
32. Landi F, Baraldi L, Cornia M, Cucchiara R. Working memory connections for LSTM. *Neural Netw*. 2021;144:334-341. [\[CrossRef\]](#)
33. Schick T, Dwivedi-Yu J, Dessi R, et al. Toolformer: language models can teach themselves to use tools. *arXiv*. 2023. [\[CrossRef\]](#)
34. OpenAI. GPT-4 Technical Report. *arXiv*. 2023. [\[CrossRef\]](#)
35. Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: early experiments with GPT-4. *arXiv*. 2023. [\[CrossRef\]](#)
36. Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. *arXiv*. 2022. [\[CrossRef\]](#)
37. Li Z. The Dark Side of ChatGPT: legal and ethical challenges from stochastic parrots and hallucination. *arXiv*. 2023. [\[CrossRef\]](#)
38. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*; 2021:610-623. [\[CrossRef\]](#)
39. Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation. *ACM Comput Surv*. 2022;55(12):1-38. [\[CrossRef\]](#)
40. Chen L, Zaharia M, Zou J. How is ChatGPT's behavior changing over time? *arXiv*. 2023. [\[CrossRef\]](#)
41. Mallio CA, Sertorio AC, Bernetti C, Beomonte Zobel B. Large language models for structured reporting in radiology: performance of GPT-4, ChatGPT-3.5, perplexity and Bing. *Radiol Med*. 2023;128(7):808-812. [\[CrossRef\]](#)
42. Lu Y, Bartolo M, Moore A, Riedel S, Stenetorp P. Fantastically ordered prompts and where to find them: overcoming few-shot prompt order sensitivity. *arXiv*. 2021. [\[CrossRef\]](#)
43. Kaddour J, Harris J, Mozes M, Bradley H, Raileanu R, McHardy R. Challenges and applications of large language models. *arXiv*. 2023. [\[CrossRef\]](#)
44. Prompt Engineering | Lil'Log. Accessed July 30, 2023. [\[CrossRef\]](#)
45. Yao S, Yu D, Zhao J, et al. Tree of thoughts: deliberate problem solving with large language models. *arXiv*. 2023. [\[CrossRef\]](#)
46. Van Veen D, Van Uden C, Attias M, et al. RadAdapt: radiology report summarization via lightweight domain adaptation of large language models. *arXiv*. 2023. [\[CrossRef\]](#)
47. Mialon G, Dessi R, Lomeli M, et al. Augmented language models: a survey. *arXiv*. 2023. [\[CrossRef\]](#)
48. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *arXiv*. 2020. [\[CrossRef\]](#)
49. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198. [\[CrossRef\]](#)
50. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180. [\[CrossRef\]](#)
51. Zhu L, Mou W, Chen R. Can the ChatGPT and other large language models with internet-connected database solve the questions and concerns of patient with prostate cancer and help democratize medical knowledge? *J Transl Med*. 2023;21(1):269. [\[CrossRef\]](#)
52. Yoshiyasu Y, Wu F, Dhanda AK, Gorelik D, Takashima M, Ahmed OG. GPT-4 accuracy and completeness against International Consensus Statement on Allergy and Rhinology: rhinosinusitis. *Int Forum Allergy Rhinol*. 2023. [\[CrossRef\]](#)
53. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol*. 2023;29(3):721-732. [\[CrossRef\]](#)
54. Sorin V, Klang E, Sklair-Levy M, et al. Large language model (ChatGPT) as a support tool for breast tumor board. *NPJ Breast Cancer*. 2023;9(1):44. [\[CrossRef\]](#)
55. Chen S, Guevara M, Ramirez N, et al. Natural language processing to automatically extract the presence and severity of esophagitis in notes of patients undergoing radiotherapy. *JCO Clin Cancer Inform*. 2023;7:e2300048. [\[CrossRef\]](#)
56. Alsentzer E, Murphy JR, Boag W, et al. Publicly available Clinical BERT embeddings. *arXiv*. 2019. [\[CrossRef\]](#)
57. Mahbub M, Srinivasan S, Danciu I, et al. Unstructured clinical notes within the 24 hours since admission predict short, mid & long-term mortality in adult ICU patients. *PLoS ONE*. 2022;17(1):e0262182. [\[CrossRef\]](#)
58. Lokker C, Bagheri E, Abdelkader W, et al. Deep learning to refine the identification of high-quality clinical research articles from the biomedical literature: performance evaluation. *J Biomed Inform*. 2023;142:104384. [\[CrossRef\]](#)
59. Zitu MM, Zhang S, Owen DH, Chiang C, Li L. Generalizability of machine learning methods in detecting adverse drug events from clinical narratives in electronic medical records. *Front Pharmacol*. 2023;14:1218679. [\[CrossRef\]](#)
60. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5(3):107-108. [\[CrossRef\]](#)

61. Gertz RJ, Bunck AC, Lennartz S, et al. GPT-4 for automated determination of radiological study and protocol based on radiology request forms: a feasibility study. *Radiology*. 2023;307(5):e230877. [\[CrossRef\]](#)
62. Sun Z, Ong H, Kennedy P, et al. Evaluating GPT4 on impressions generation in radiology reports. *Radiology*. 2023;307(5):e231259. [\[CrossRef\]](#)
63. Kottlors J, Bratke G, Rauen P, et al. Feasibility of differential diagnosis based on imaging patterns using a large language model. *Radiology*. 2023;308(1):e231167. [\[CrossRef\]](#)
64. Bhayana R, Bleakney RR, Krishna S. GPT-4 in radiology: improvements in advanced reasoning. *Radiology*. 2023;307(5):e230987. [\[CrossRef\]](#)
65. Rahsepar AA, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. How AI responds to common lung cancer questions: ChatGPT vs google bard. *Radiology*. 2023;307(5):e230922. [\[CrossRef\]](#)
66. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. *Radiology*. 2023;307(5):e230582. [\[CrossRef\]](#)
67. Jiang ST, Xu YY, Lu X. ChatGPT in radiology: evaluating proficiencies, addressing shortcomings, and proposing integrative approaches for the future. *Radiology*. 2023;308(1):e231335. [\[CrossRef\]](#)
68. Zhu W, Liu H, Dong Q, et al. Multilingual machine translation with large language models: empirical results and analysis. *arXiv*. 2023. [\[CrossRef\]](#)
69. López-Úbeda P, Martín-Noguerol T, Luna A. Radiology in the era of large language models: the near and the dark side of the moon. *Eur Radiol*. 2023. [\[CrossRef\]](#)
70. Li R, Allal LB, Zi Y, et al. StarCoder: may the source be with you! *arXiv*. 2023. [\[CrossRef\]](#)
71. Hu M, Pan S, Li Y, Yang X. Advancing medical imaging with language models: a journey from n-grams to ChatGPT. *arXiv*. 2023. [\[CrossRef\]](#)
72. Coavoux M, Narayan S, Cohen SB. Privacy-preserving neural representations of text. *arXiv*. 2018. [\[CrossRef\]](#)
73. Bills S, Cammarata N, Mossing D, et al. Language Models Can Explain Neurons in Language Models. OpenAI; 2023. Accessed September 6, 2023. [\[CrossRef\]](#)
74. Liao QV, Vaughan JW. AI Transparency in the age of LLMs: a human-centered research roadmap. *arXiv*. 2023. [\[CrossRef\]](#)
75. Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models. *arXiv*. 2023. [\[CrossRef\]](#)
76. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. *Diagn Interv Imaging*. 2023;104(6):269-274. [\[CrossRef\]](#)
77. Elkassam AA, Smith AD. Potential use cases for ChatGPT in radiology reporting. *AJR Am J Roentgenol*. 2023;221(3):373-376. [\[CrossRef\]](#)
78. Li H, Moon JT, Iyer D, et al. Decoding radiology reports: potential application of OpenAI ChatGPT to enhance patient understanding of diagnostic reports. *Clin Imaging*. 2023;101:137-141. [\[CrossRef\]](#)
79. Shahsavari Y, Choudhury A. User Intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. *JMIR Hum Factors*. 2023;10:e47564. [\[CrossRef\]](#)
80. Doo FX, Cook T, Siegel E, et al. Exploring the clinical translation of generative models like chatgpt: promise and pitfalls in radiology, from patients to population health. *J Am Coll Radiol*. 2023. [\[CrossRef\]](#)
81. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023;183(6):589-596. [\[CrossRef\]](#)
82. Korngiebel DM, Mooney SD. Considering the possibilities and pitfalls of generative pre-trained transformer 3 (GPT-3) in healthcare delivery. *NPJ Digit Med*. 2021;4(1):93. [\[CrossRef\]](#)
83. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med*. 2023;6(1):120. [\[CrossRef\]](#)
84. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. *Lancet Digit Health*. 2023;5(6):333-335. [\[CrossRef\]](#)
85. Song R, Giunchiglia F, Li Y, Shi L, Xu H. Measuring and mitigating language model biases in abusive language detection. *Inf Process Manage*. 2023;60(3):103277. [\[CrossRef\]](#)
86. Shumailov I, Shumaylov Z, Zhao Y, Gal Y, Papernot N, Anderson R. The curse of recursion: training on generated data makes models forget. *arXiv*. 2023. [\[CrossRef\]](#)
87. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. *Nat Med*. 2023. [\[CrossRef\]](#)
88. Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts | panel for the future of science and technology (STOA) | European Parliament. Accessed July 25, 2023. [\[CrossRef\]](#)
89. Madden MG, McNicholas BA, Laffey JG. Assessing the usefulness of a large language model to query and summarize unstructured medical notes in intensive care. *Intensive Care Med*. 2023;49:1018-1020. [\[CrossRef\]](#)
90. Lourenco AP, Slanetz PJ, Baird GL. Rise of chatgpt: it may be time to reassess how we teach and test radiology residents. *Radiology*. 2023;307(5):e231053. [\[CrossRef\]](#)
91. Karabacak M, Margetis K. Embracing large language models for medical applications: opportunities and challenges. *Cureus*. 2023;15(5):e39305. [\[CrossRef\]](#)
92. Zhang X, Li S, Yang X, Tian C, Qin Y, Petzold LR. Enhancing small medical learners with privacy-preserving contextual prompting. *arXiv*. 2023. [\[CrossRef\]](#)
93. Liu Z, Yu X, Zhang L, et al. DeID-GPT: Zero-shot medical text de-identification by GPT-4. *arXiv*. 2023. [\[CrossRef\]](#)
94. Wang B, Zhang YJ, Cao Y, et al. Can public large language models help private cross-device federated learning? *arXiv*. 2023. [\[CrossRef\]](#)
95. Hou C, Zhan H, Shrivastava A, et al. Privately customizing prefinetuning to better match user data in federated learning. *arXiv*. 2023. [\[CrossRef\]](#)
96. Zhu K, Wang J, Zhou J, et al. PromptBench: towards evaluating the robustness of large language models on adversarial prompts. *arXiv*. 2023. [\[CrossRef\]](#)
97. Collins KM, Wong C, Feng J, Wei M, Tenenbaum JB. Structured, flexible, and robust: benchmarking and improving large language models towards more human-like behavior in out-of-distribution reasoning tasks. *arXiv*. 2022. [\[CrossRef\]](#)



Diagnostic value of synthetic diffusion-weighted imaging on breast magnetic resonance imaging assessment: comparison with conventional diffusion-weighted imaging

Ebru Yılmaz¹
 Nilgün Göldoğan¹
 Sıla Ulus²
 Ebru Banu Türk¹
 Mustafa Enes Mısırlı³
 Aydan Arslan⁴
 Mustafa Erkin Arıbal³

¹Acıbadem Altunizade Hospital Breast Center,
Department of Radiology, İstanbul, Türkiye

²Acıbadem Ataşehir Hospital, Department of
Radiology, İstanbul, Türkiye

³Acıbadem Mehmet Ali Aydınlar University,
Department of Radiology, İstanbul, Türkiye

⁴University of Health Sciences Türkiye, Ümraniye
Training and Research Hospital, Clinic of Radiology,
İstanbul, Türkiye

PURPOSE

To compare images generated by synthetic diffusion-weighted imaging (sDWI) with those from conventional DWI in terms of their diagnostic performance in detecting breast lesions when performing breast magnetic resonance imaging (MRI).

METHODS

A total of 128 consecutive patients with 135 enhanced lesions who underwent dynamic MRI between 2018 and 2021 were included. The sDWI and DWI signals were compared by three radiologists with at least 10 years of experience in breast radiology.

RESULTS

Of the 82 malignant lesions, 91.5% were hyperintense on sDWI and 73.2% were hyperintense on DWI. Of the 53 benign lesions, 71.7% were isointense on sDWI and 37.7% were isointense on DWI. sDWI provides accurate signal intensity data with statistical significance compared with DWI ($P < 0.05$). The diagnostic performance of DWI and sDWI to differentiate malignant breast masses from benign masses was as follows: sensitivity 73.1% [95% confidence interval (CI): 62–82], specificity 37.7% (95% CI: 24–52); sensitivity 91.5% (95% CI: 83–96), specificity 71.7% (95% CI: 57–83), respectively. The diagnostic accuracy of DWI and sDWI was 59.2% and 83.7%, respectively. However, when the DWI images were evaluated with apparent diffusion coefficient mapping and compared with the sDWI images, the sensitivity was 92.68% (95% CI: 84–97) and the specificity was 79.25% (95% CI: 65–89) with no statistically significant difference. The inter-reader agreement was almost perfect ($P < 0.001$).

CONCLUSION

Synthetic DWI is superior to DWI for lesion visibility with no additional acquisition time and should be taken into consideration when conducting breast MRI scans. The evaluation of sDWI in routine MRI reporting will increase diagnostic accuracy.

KEYWORDS

Breast tumors, image analysis, diagnostic imaging, diffusion magnetic resonance imaging, echo-planar imaging

Corresponding author: Aydan Arslan

E-mail: arslanaydan@gmail.com

Received 21 August 2023; revision requested 18
September 2023; accepted 22 October 2023.



Epub: 27.10.2023

Publication date: 05.03.2024

DOI: 10.4274/dir.2023.232466

Conventional breast magnetic resonance imaging (MRI) has the highest sensitivity for breast cancer detection, staging of known cancer, and evaluation of response to neoadjuvant chemotherapy. In recent years, specifically according to the Dense Tissue and Early Breast Neoplasm Screening trial¹ and the EA1411 Eastern Cooperative Oncology Group–American College of Radiology (ACR) Imaging Network study,² the indication spectrum of breast MRI has widened. The European Society of Breast Imaging now recommends offering screening breast MRI every 2–4 years in women aged 50–70 years with extremely dense breasts.³ However, MRI is limited by high costs, which include the cost of contrast mate-

rial administration, intravenous (IV) supplies, point-of-care renal function screening, and on-site physician coverage for adverse contrast material-related events. Furthermore, prolonged examination time is an additional concern. With the latest concerns regarding the safety of gadolinium-based contrast agents, diffusion-weighted imaging (DWI) has been recommended as an encouraging alternative to dynamic contrast-enhanced MRI for detecting early breast cancer. In addition, an improvement in tumor visibility without contrast injection could improve the cost-effectiveness of MRI. DWI is a fast, widely available, unenhanced MRI technique that provides a unique radiologic image contrast by providing information on the cellular environment of tissues *in vivo*. In recent years, this sequence has been used in addition to conventional sequences and decreases false positivity.

However, DWI has many limitations, such as a decrease in signal-to-noise ratio (SNR) and experiencing eddy current distortions when using a high *b* value; studies have been conducted to overcome these limitations and to improve this sequence.⁴ The principal basis for using DWI for disease detection relies on maximizing the image contrast between diseased tissue and the background. The extent to which this occurs depends on the intrinsic tissue diffusivity, the T2 relaxation time, and the diffusion weighting (*b* value) of the motion probing gradients that are applied. The selection of the *b* value is a key point as it directly affects the image SNR, lesion contrast-to-noise ratio, and apparent diffusion coefficients (ADCs). The *b* value can enhance both the lesion detection sensitivity and specificity, but it also leads to a decreased SNR. Furthermore, the perfusion effect is minimized.^{5,6} However, acquiring images at higher *b* values (>1.000 sec/mm²) leads to more distortion due to susceptibility effects and eddy currents and lengthens

imaging times.⁷⁻¹⁰ One of the new techniques for improving the accuracy of DWI is synthetic DWI (sDWI). sDWI is a mathematical computation technique that builds on previously described principles and calculates a high *b* value (or any *b* value) image from DWI images acquired with at least two different lower *b* values.^{11,12} Once the ADC is known, it can be used to extrapolate the expected signal intensity for each image voxel to any computed *b* value using the equation $S(b) = S(0) e^{-b \cdot \text{ADC}}$, thus generating a computed DWI image.^{6,13,14} The calculation of synthetic high *b* values is a strategy to enhance contrast already present in lower *b* value images and is potentially useful to detect and depict lesions but lacks the power of non-Gaussian diffusion to characterize tissues.

sDWI is superior to conventional DWI with fewer artifacts, no inhomogeneity in fat suppression, and a high SNR. There are many reports¹⁵⁻¹⁸ on the efficacy of sDWI for imaging organs such as the prostate; however, there are limited reports on the evaluation of breast lesions.^{18,19}

The aim of this study is to investigate the feasibility of the sDWI technique for lesion detection and to compare it with conventional DWI.

Methods

Patients selection

This study involved a retrospective analysis of acquired data. The medical Ethics Committee of Acibadem University approved this single institution study (2023-09/303), and informed consent was waived. All the enhancing lesions on breast MRI images between March 2018 and September 2021 were included in the study. The exclusion criteria were as follows: cases with no histopathological diagnosis or 2 years follow-up; cases with biopsy history prior to MRI; cases involving MRI scans following neoadjuvant chemotherapy.

A total of 139 breast MRI scans were evaluated, and four patients were excluded due to artifacts and technical inadequacy (insufficient fat suppression). Consequently, 135 lesions in 128 patients were evaluated (median age: 47.51 ± 11.15 years; age range: 27–79 years). Among these, 117 lesions had histopathological diagnoses, either with core needle biopsy or vacuum-assisted biopsy, and the remaining 18 lesions were stable in the 2-year follow-up and were regarded as benign.

Magnetic resonance imaging technique

All examinations were performed using a 1.5 Tesla (T) MRI device (Aera; Siemens Healthcare) using an 18-channel breast matrix surface receiver coil with prone positioning. Care was taken to perform the MRI scans of premenopausal women between days 5 and 15 of the menstrual cycle. Multi-parametric MRI images (fat-sat STIR T2W sequence, a pre-contrast DWI sequence, and a dynamic contrast-enhanced T1W sequence) were obtained for all the patients. For the dynamic contrast-enhanced sequences, 0.1 mmol/kg of body weight of contrast material (Gadovist; Bayer Healthcare Pharmaceutical, Berlin, Germany) was injected. Diffusion-weighted echo-planar images (time of repetition: 3.000–7.000 ms, time of echo: 50–60 ms, field of view: 260–300 mm, matrix: 192 × 192, number of excitations: 1, sectional thickness: 4 mm with a 1 mm intersection gap) were obtained in the axial plane prior to contrast administration. The DWI was obtained using diffusion gradient *b* values of 50–800 sec/mm². ADC maps were calculated from raw DWI images using all *b* values and applying the standard monoexponential regression approach performed automatically by the scanner software. sDWI images at *b* = 1.500 sec/mm² were automatically constructed in a commercially available workstation using syngo.via VB10 software (Siemens Healthcare, Erlangen, Germany).

Data analysis

The evaluation was performed by three radiologists with at least 10 years of experience in breast radiology. One radiologist evaluated the sDWI images in addition to all sequences, whereas the other two radiologists, who were blinded to the clinical diagnoses and all imaging findings, evaluated only the conventional DWI and sDWI images. Imaging data were analyzed on a dedicated workstation (Multi-Modality Work-Place, Siemens Healthineers). Detection of hyperintense lesions on sDWI and DWI was acknowledged as positive for malignancy suspicion as on DWI, and the readers assigned a qualitative positive or negative assessment.

Lesion size was defined as the largest diameter of the enhancing lesion in the first minute post-contrast T1W sequence. For patients with more than one lesion, the largest lesion was included in the study.

Breast densities were categorized into four groups (type A, B, C, and D) according to the American College of Radiology Breast Imaging Reporting and Data System atlas terminology.²⁰

Main points

- Synthetic diffusion-weighted imaging (sDWI) is a mathematical computation technique to generate a high *b* value DWI image. No extra acquisition time is required.
- Synthetic DWI is superior to DWI sequence due to fewer artifacts, lack of inhomogeneity in fat suppression, and a high signal-to-noise ratio.
- Synthetic DWI is superior to DWI sequences for lesion visibility, particularly in dense breasts. Thus, sDWI should be considered when conducting breast magnetic resonance imaging scans.

Statistical analysis

Data of continuous variables were presented as a mean ± standard deviation, minimum–maximum, and percentile. A comparison of two variables that were independent and not normally distributed was performed using the Mann–Whitney U test. The chi-squared test (or, when appropriate, Fisher’s exact test) was used to investigate the relationship between the categorical variables.

Receiver operating characteristic curve analysis was performed for the determination of ADC cut-off values according to the statistically significant parameters. Subsequently, the diagnostic values and confidence intervals (CIs) were obtained.

The inter-reader agreement was evaluated using intra-class correlation and the Fleiss κ test.

Statistical analysis was performed using MedCalc Statistical Software version 12.7.7 (MedCalc Software bvba, Ostend, Belgium; <http://www.medcalc.org>; 2013). All *P* values of <0.05 were considered statistically significant.

Results

Eighty-two lesions were malignant (confirmed via histopathologic diagnosis) (mean size: 27.4 ± 20.55 mm, range: 3–100 mm), and 53 lesions were benign (35 lesions were confirmed via histopathologic diagnosis, and 18 lesions were stable in the 2-year follow-up and regarded as benign) (mean size: 13.06 ± 7.66 mm, range: 4–36 mm) (Table 1).

Diagnoses of benign lesions included fibrocystic changes (*n* = 10; 18.8%), fibroadenoma (*n* = 7; 13.2%), usual epithelial hyperplasia (*n* = 1; 1.8%), fat necrosis (*n* = 2; 3.7%), adenosis (*n* = 4; 7.5%), stromal fibrosis (*n* = 8; 15.0%), radial scar (*n* = 2; 3.7%), intramammary lymph node (*n* = 1; 1.8%), and stable in the 2-year follow-up (*n* = 18; 33.9%). Malignant lesion subtypes consisted of invasive ductal carcinoma (IDC) (*n* = 70; 85.3%), invasive lobular carcinoma (*n* = 3; 3.6%), metaplastic carcinoma (*n* = 1; 1.2%), and ductal carcinoma *in situ* (DCIS) (*n* = 8; 9.7%).

The diagnostic performance of ADC value to differentiate malignant breast masses from benign masses was as follows: sensitivity 92.68% (95% CI: 84–97), specificity 79.25% (95% CI: 65–89), using an ADC cut-off value of 1.189 × 10^{−3} mm²/sec, which is comparable with the literature data.^{21–24}

Of the 82 malignant lesions, 75 were hyperintense on sDWI, whereas 7 were iso-

intense (Figure 1). Four of the 7 sDWI iso-intense malignant lesions were DCIS. Two of the remaining three IDCs (80 and 10 mm in size and both Luminal B cancers) were located peripherally in the axillary tail, and the lesions were not visualized due to insufficient fat-suppression and distortion in that area (Figure 2). One triple-negative IDC lesion appearing isointense on the DWI image

showed diffusion restriction when evaluated with an ADC map. Of the seven isointense cases, six were also isointense on DWI. There was one case in which the biopsy result of DCIS was hyperintense on DWI with a high ADC value (1.504 × 10^{−3} mm²/sec). Therefore, it was interpreted as no diffusion restriction and DWI hyperintensity may be due to the T2 effect. When the DWIs were evaluated with

Table 1. Summary of characteristics of the study population				
Clinical and radiological characteristics		#		
Patient age, median (range) (years)		47.51 ± 11.15 (27–79)		
Amount of fibroglandular tissue (breast density), number (%) of patients	Type A	1 (0.74%)		
	Type B	24 (18.75%)		
	Type C	75 (58.59%)		
	Type D	35 (27.34%)		
Pathology result, number (%) of patients	Malignant	82 (60.74%)		
	Benign	53 (39.25%)		
Lesion size, mean (range)	Benign	13.06 ± 7.66 mm (4–36 mm)		
		Mass	11.18 mm (4–30 mm)	
		Non-mass	16.15 mm (6–36 mm)	
	Malignant	27.4 ± 20.55 mm (3–100 mm)		
		Mass	25.8 mm (3–80 mm)	
		Non-mass	36.5 mm (9–100 mm)	
Lesion enhancement patterns (mass, non-mass)	Benign	Mass	Non-mass	
		33 (62.26%)	20 (37.73%)	
	Malignant	70 (85.36%)	12 (14.63%)	

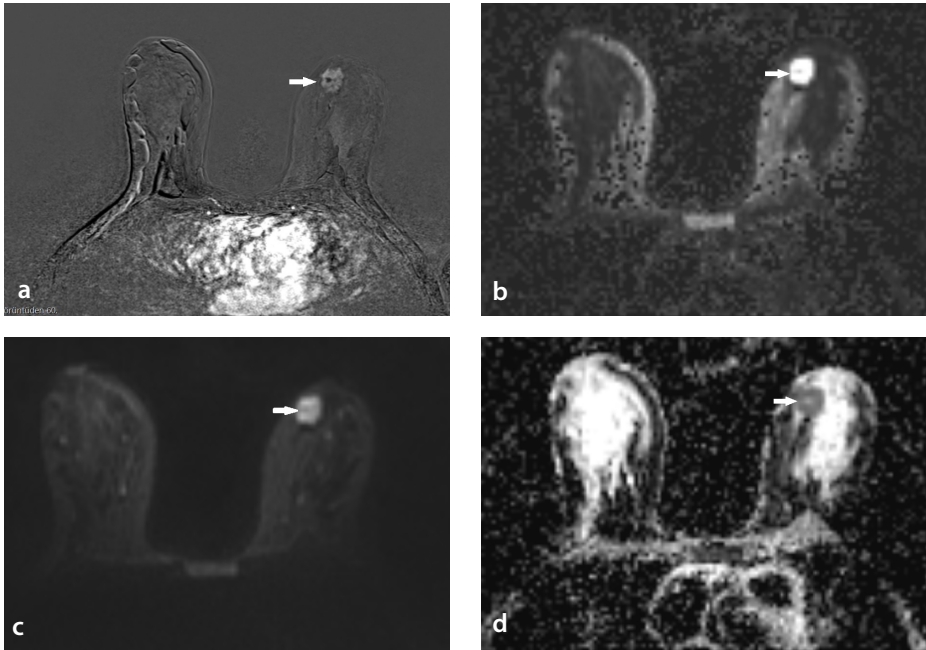


Figure 1. A 44-year-old female patient with a diagnosis of invasive ductal cancer in the inner part of the left breast. In the dynamic contrast-enhanced image (a), there is an irregular mass with heterogeneous enhancement. Synthetic diffusion-weighted image (b) showed markedly hyperintense mass, and conventional diffusion-weighted image (c, d) showed homogeneous diffusion restriction in the mass (hyperintense on DWI and hypointense on ADC map). The ADC values were measured at least three times. The average ADC value was 0.769 × 10^{−3} mm²/sec, which is below the cut-off value (1.189 × 10^{−3} mm²/sec). DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient.

ADC mapping (using a cut-off value of $1.189 \times 10^{-3} \text{ mm}^2/\text{sec}$), three cases showed diffusion restriction, whereas four did not (Table 2).

When the stand-alone DWI sequence was evaluated, 60 of the 82 malignant lesions were hyperintense, and the remaining 22 were isointense. Of these 22 isointense lesions, 6 were isointense on both sDWI and DWI, whereas 16 lesions were hyperintense solely on sDWI. When the DWI sequence was compared with ADC mapping, 15 of 16 lesions (in one lesion ADC value is $1.239 \text{ mm}^2/\text{sec}$) showed restricted diffusion with ADC values less than the cut-off value of $1.189 \times 10^{-3} \text{ mm}^2/\text{sec}$.

Comparison of DWI acquired with $b = 800$ diffusion gradient to sDWI at $b = 1.500$ value (Table 3) in the detection of malignant lesions showed that sDWI was significantly effective ($P < 0.001$).

Of the 53 benign lesions, 38 were isointense on sDWI, whereas 15 were hyperintense (Figure 3). Of these 15 hyperintense lesions, 4 were isointense on DWI and the remaining 11 were hyperintense (Table 4). When compared with ADC mapping, 6 of these 11 lesions had ADC values above the cut-off value of $1.189 \times 10^{-3} \text{ mm}^2/\text{sec}$. The remaining 5 lesions had ADC values below the cut-off value.

For the 53 benign lesions, 33 were hyperintense, and 20 were isointense on DWI images. ADC mapping showed values below the cut-off value of $1.189 \times 10^{-3} \text{ mm}^2/\text{sec}$ in 11 of these 53 benign lesions.

When the sDWI and DWI sequences for benign lesions were compared, the false positivity rates were 28.30% and 62.26%, respectively. If DWI sequences are assessed in conjunction with the ADC map, the rate of false positive results is 20.75%.

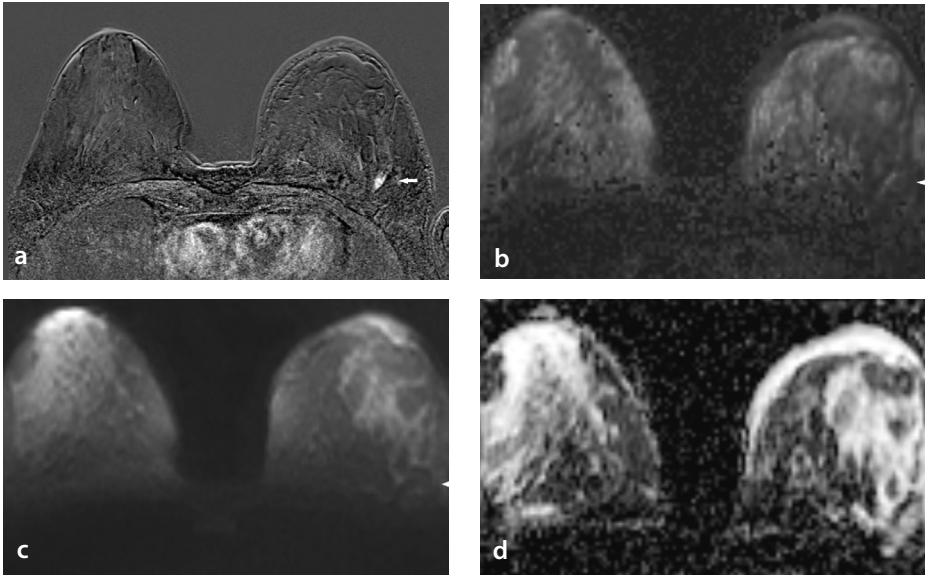


Figure 2. A 64-year-old female patient with a diagnosis of Luminal B IDC in the axillary tail of the left breast is observed. In the dynamic contrast-enhanced image (a), there is an enhanced non-mass lesion, which is a potential malignancy. Synthetic diffusion-weighted image (b) showed lesion isointense. Conventional diffusion-weighted image (c) showed lesion isointense and DWI was evaluated with ADC map (d) there is diffusion restriction in the lesion. The average ADC value was $1.003 \times 10^{-3} \text{ mm}^2/\text{sec}$, which is below the cut-off value ($1.189 \times 10^{-3} \text{ mm}^2/\text{sec}$). IDC, invasive ductal carcinoma; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient.

Table 2. Analysis of the seven false (-) lesions on sDWI				
Maximum diameter	Pathology result	DWI	ADC value ($\times 10^{-3} \text{ mm}^2/\text{sec}$)	sDWI
15 mm	DCIS	Isointense	1.156*	Isointense
10 mm	Luminal B IDC	Isointense	1.003*	Isointense
80 mm	Luminal B IDC	Isointense	1.328	Isointense
10 mm	DCIS	Isointense	1.514	Isointense
12 mm	DCIS	Isointense	1.200	Isointense
10 mm	Triple (-) IDC	Isointense	1.184*	Isointense
23 mm	DCIS	Hyperintense	1.504	Isointense

*The ADC values below the cut-off ($1.189 \times 10^{-3} \text{ mm}^2/\text{sec}$) are marked. DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; sDWI, synthetic diffusion-weighted imaging; IDC, invasive ductal carcinoma; DCIS, ductal carcinoma *in situ*.

Table 3. Comparison of lesion signal intensities of DWI obtained with $b = 800$ diffusion gradient and sDWI with $b = 1.500$									
Pathology result		Benign		Malignant		P value			
		n	%	n	%				
DWI	Hyperintense	33	62.3%	60	73.2%	0.189			
	Isointense	20	37.7%	22	26.8%				
sDWI	Hyperintense	15	28.3%	75	91.5%	<0.001			
	Isointense	38	71.7%	7	8.5%				
Pathology result		Benign		Malignant					
n		%	n	%	Auc	Acc	Sensitivity	Specificity	
DWI	Hyperintense	20	37.7%	22	26.8%	0.555	0.592	0.731 (0.622–0.824)	0.377 (0.248–0.521)
	Isointense	33	62.3%	60	73.2%				
sDWI	Hyperintense	38	71.7%	7	8.5%	0.816	0.837	0.915 (0.832–0.965)	0.717 (0.576–0.832)
	Isointense	15	28.3%	75	91.5%				

DWI, diffusion-weighted imaging; sDWI, synthetic diffusion-weighted imaging; AUC, area under curve; Acc, accuracy.

The diagnostic performance of DWI and sDWI to differentiate malignant breast masses from benign masses was as follows: sensitivity 73.1% (95% CI: 62–82), specificity 37.7% (95% CI: 24–52); sensitivity 91.5% (95% CI: 83–96), specificity 71.7% (95% CI: 57–83), respectively. The diagnostic accuracy of DWI and sDWI was 59.2% and 83.7%, respectively.

Lesion size represents a significant limitation when evaluating lesions on the DWI

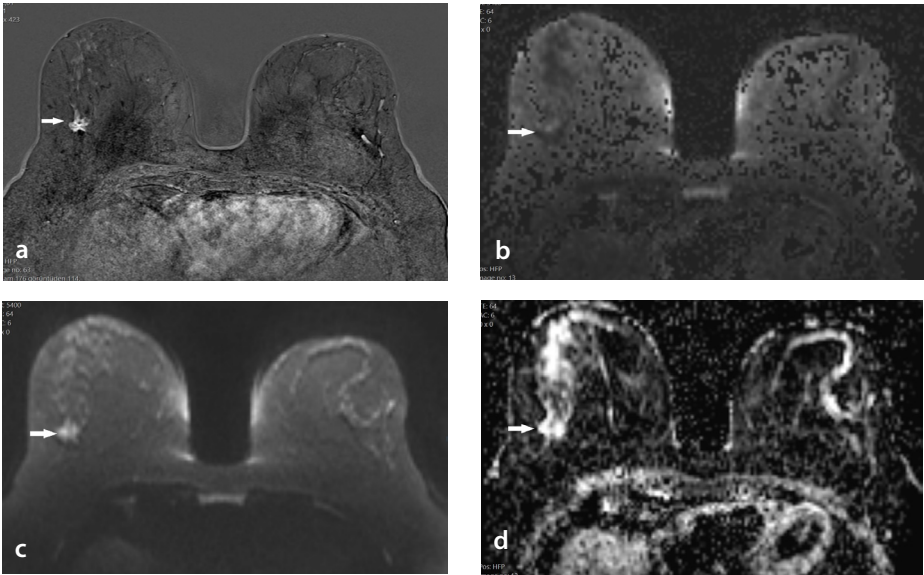


Figure 3. A 42-year-old female patient with a diagnosis of stromal fibrosis in the outer part of the right breast was observed. Structural distortion was observed on routine annual mammography and MRI was suggested. In the dynamic contrast-enhanced image (a), there was an enhanced lesion with an irregular shape, which indicates malignancy. The synthetic diffusion-weighted image (b) showed lesion isointense. The conventional diffusion-weighted image (c) showed a hyperintense signal but when DWI was evaluated with an ADC map (d) there was no diffusion restriction in the lesion. The average ADC value was $1.429 \times 10^{-3} \text{ mm}^2/\text{sec}$, which is above the cut-off value ($1.189 \times 10^{-3} \text{ mm}^2/\text{sec}$). MRI, magnetic resonance imaging; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient.

Table 4. Analysis of the 15 false (+) lesions on sDWI				
Maximum diameter	Pathology result	DWI	ADC ($\times 10^{-3} \text{ mm}^2/\text{sec}$)	sDWI
12 mm	Radial scar	Hyperintense	1.077*	Hyperintense
30 mm	Fibrocystic change	Hyperintense	1.218	Hyperintense
27 mm	Fibrocystic change	Hyperintense	1.614	Hyperintense
9 mm	Stabil -2 years follow up	Hyperintense	1.409	Hyperintense
9 mm	Stabil -2 years follow up	Hyperintense	1.460	Hyperintense
5 mm	Stabil -2 years follow up	Hyperintense	1.357	Hyperintense
8 mm	Intramammary lymph node	Hyperintense	1.100*	Hyperintense
8 mm	Apocrine metaplasia	Hyperintense	1.309	Hyperintense
30 mm	Adenosis	Isointense	1.112*	Hyperintense
12 mm	Stromal fibrosis	Isointense	1.016*	Hyperintense
13 mm	Fibroadenoma	Isointense	1.122*	Hyperintense
30 mm	Sclerosing adenosis	Isointense	1.079*	Hyperintense
36 mm	Apocrine metaplasia.fibrocystic change	Hyperintense	1.033*	Hyperintense
24 mm	Complex sclerosing lesion	Hyperintense	1.152*	Hyperintense
13 mm	Usual epithelial hyperplasia	Hyperintense	0.928*	Hyperintense

*The ADC values below the cut-off ($1.189 \times 10^{-3} \text{ mm}^2/\text{sec}$) are marked. DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; sDWI, synthetic diffusion-weighted imaging.

ing <1 cm (9 mm) that could not be detected in the DWI sequence.

In addition, the study identified a total of 32 lesions falling within the size range of 3–9 mm (<1 cm). Among these, 9 were malignant and 23 were benign. As explained earlier, all but one of these nine malignant lesions were detectable. Among the benign lesions, 11 displayed isointensity in both sequences, 7 showed hyperintensity solely in the DWI sequence, and 6 were hyperintense in both the DWI and sDWI sequences. When comparing the relationship between size and detectability in both benign and malignant lesions, no statistically significant difference ($P = 0.0867$) was found. These findings suggest that size does not significantly impact the assessment of lesions in the DWI and sDWI sequences.

The distribution of the individual amounts of fibroglandular tissue (FGT) is important because difficulties with lack of fat saturation are more common in breasts with a high percentage of fat.¹² The performance of DWI and sDWI sequences can be affected by the distribution of FGT. In this study, the composition of the female group included 24 (18.75%) women within ACR category B, 75 (58.59%) women within category C, and 35 (27.34%) women within category D. Only one case was categorized as “type A,” and this category was thus disregarded (shown in Table 5). No statistically significant difference was found between the breast parenchymal distribution and signal distribution DWI and sDWI sequences regarding type B, type C, and type D (DWI $P = 0.066$; sDWI $P = 0.335$).

Synthetic DWI inter-reader agreement was almost perfect for lesion visibility (shown in Table 6) and was statistically significant (κ : 0.922, $P < 0.001$).

Discussion

DWI has been proposed as an unenhanced option for breast cancer screening via MRI, and synthetic b values may improve lesion visibility without increasing the acquisition time while avoiding the disadvantages of performing DWI at extremely high b values. In this study, DWI was assessed for tumor visibility and breast cancer detection by a combination of acquired b values (800 sec/mm²), ADC maps, and synthetic b values (1.500 sec/mm²). Synthetic b values of 1.500 sec/mm² provided the best lesion conspicuity. Benign lesions were more conspicuous at lower b values, whereas malignant tumors appeared brighter than the surrounding parenchyma at higher b values, particularly in breast composition categories C and D, where lesions can be masked on mammograms by the density of FGT. The study shows that sDWI sequences are superior to DWI sequences for lesion visibility, and the results corroborate those in the recent literature.^{11,18}

In this study, the lesion visibility was assessed using DWI and sDWI. The former was significantly less sensitive than the latter (73.1% vs. 91.5%, $P < 0.001$). Furthermore, the detection rate of cancer was not significantly different when the DWI sequences were

evaluated in conjunction with ADC mapping (92.68%). In two independent studies,^{11,18} the conventional DWI of cancer to parenchyma contrast ratio of malignant lesions was compared with DWI with higher b values. The authors reported that the tumor-to-parenchymal contrast ratio of sDWI was significantly higher than that of conventional DWI, without compromising the cancer detection rate. In another study, Blackledge et al.¹⁶ claimed that sDWI raised the rate of lesion detection in comparison to DWI. O'Flynn et al.¹⁹ also reported that sDWI increased sensitivity for breast cancer. However, these studies involved smaller cohorts, and only malignant lesions were evaluated.

The prime focus of DWI is to differentiate between benign and malignant lesions to prevent unnecessary breast biopsies or enable screening without admission of IV contrast media. Only 19%–36% of the lesions that are biopsied due to MR examination results turn out to be malignant.^{25,26} It is particularly important to differentiate benign lesions, which show contrast enhancement resulting in false positive results. There is limited data on the evaluation of benign lesions on sDWI.¹⁸ Unlike other studies, in this study, both malignant and benign lesions were evaluated. The rate of false positivity of sDWI and DWI was 28.30% and 62.26%, respectively. The high rate of false positivity on DWI has been attributed to the T2 shine-through effect.¹⁶ However, the false positivity was 20.75% when DWI was evaluated with

an ADC map (Table 2). This data indicates that by first examining the sDWI images and then evaluating any questionable lesions using the ADC map, the opportunity exists to achieve a quicker and more precise diagnosis, while also avoiding unnecessary biopsies.

Studies conducted in recent years have shown that the combined evaluation of early phase contrast images and DWI can replace late-phase and kinetic-curve evaluation, meaning results can be achieved with a much shorter imaging time.²⁷ Furthermore, in recent years, an annual or biannual screening MRI has been recommended for high-risk patients with dense breasts, and an abbreviated MRI is aimed at achieving fast and accurate results.²⁸ The primary objective of breast MRI is to enhance the precision of diagnosis while minimizing the likelihood of overdiagnosis. In most of the abbreviated MRI protocols, DWI sequences are included.^{29–31} This study shows that adding sDWI scans, which do not require additional acquisition time in the evaluation, has the potential to increase the diagnostic accuracy of abbreviated MRI evaluations.

The performance of DWI and sDWI sequences can be affected by the distribution of FGT. Fat suppression cannot be made homogeneously in fatty breasts in diffusion-weighted sequences. Furthermore, fatty breasts have a lower ADC value compared with dense breasts in the retroareolar region and upper outer quadrant.^{12,32} This study did not identify any correlation between breast density and the detectability of lesions on either DWI or sDWI. Similarly, prior studies^{32,33} showed that the visibility of breast lesions on DWI was not influenced by breast density. However, since the optimal evaluation is often made with mammography in women with fatty breasts, MRI is rarely required for the evaluation of the patient and thus, the number of fatty breasts in the study group is extremely low.

This study has a number of limitations. First, it is a single-center study. A multicenter study will be valuable to showing the reproducibility of the findings. Second, this study evaluated only lesion detection. Although the image quality in the sDWI series is relatively high, no quantitative assessment was conducted. Furthermore, the role of DWI in the visualization of non-mass enhancement is not definite.^{34–36} In this study, this patient group was not evaluated separately; this evaluation should be conducted in further studies.

Table 5. Cross-evaluation of the relationship between signal distribution and breast parenchyma compositions in DWI and sDWI sequences

		Breast density category							
		Type B		Type C		Type D		P value	
		n	%	n	%	n	%		
B800	Hyperintense	20	83.3	53	70.7	20	57.1	0.066	
	Isointense	4	16.7	22	29.3	15	42.9		
B1500	Hyperintense	16	66.7	48	64.0	26	74.3	0.335	
	Isointense	8	33.3	27	36.0	9	25.7		

DWI, diffusion-weighted imaging; sDWI, synthetic diffusion-weighted imaging.

Table 6. Inter-reader agreement assessment

Readers	sDWI signal assesment	n (number of patients)	%	κ (kappa)	P value
Reader 1	Hyperintense	90	66.7	0.922	<0.001
	Isointense	45	33.3		
Reader 2	Hyperintense	90	66.7		
	Isointense	45	33.3		
Reader 3	Hyperintense	91	67.4		
	Isointense	44	32.6		

sDWI, synthetic diffusion-weighted imaging.

To conclude, the findings indicate that sDWI exhibits significantly greater sensitivity than conventional DWI in assessing both malignant and benign lesions. The results indicate that the inclusion of sDWI image evaluation in the interpretation of breast MRI scans has the potential for a better outcome.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Bakker MF, de Lange SV, Pijnappel RM, et al. Supplemental MRI screening for women with extremely dense breast tissue. *N Engl J Med*. 2019;381(22):2091-2102. [\[CrossRef\]](#)
- Rahbar H, Zhang Z, Chenevert TL, et al. Utility of diffusion-weighted imaging to decrease unnecessary biopsies prompted by breast MRI: a trial of the ECOG-ACRIN Cancer Research Group (A6702). *Clin Cancer Res*. 2019;25(6):1756-1765. [\[CrossRef\]](#)
- Mann RM, Athanasiou A, Baltzer PAT, et al. Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). *Eur Radiol*. 2022;32(6):4036-4045. [\[CrossRef\]](#)
- Baltzer PA, Benndorf M, Dietzel M, Gajda M, Camara O, Kaiser WA. Sensitivity and specificity of unenhanced MR mammography (DWI combined with T2-weighted TSE imaging, ueMRM) for the differentiation of mass lesions. *Eur Radiol*. 2010;20(5):1101-1110. [\[CrossRef\]](#)
- Le Bihan D, Breton E. Imagerie de diffusion in vivo par résonance magnétique. *C R Acad Sci (Paris)*. 1985;301(15):1109-1111. [\[CrossRef\]](#)
- Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*. 1988;168(2):497-505. [\[CrossRef\]](#)
- Katahira K, Takahara T, Kwee TC, et al. Ultrahigh-b-value diffusion-weighted MR imaging for the detection of prostate cancer: evaluation in 201 cases with histopathological correlation. *Eur Radiol*. 2011;21(1):188-196. [\[CrossRef\]](#)
- Kitajima K, Kaji Y, Kuroda K, Sugimura K. High b-value diffusion-weighted imaging in normal and malignant peripheral zone tissue of the prostate: effect of signal-to-noise ratio. *Magn Reson Med Sci*. 2008;7(2):93-99. [\[CrossRef\]](#)
- Nilsson M, Szczepankiewicz F, van Westen D, Hansson O. Extrapolation-based references improve motion and eddy-current correction of high B-value DWI data: application in Parkinson's disease dementia. *PLoS One*. 2015;10(11):e0141825. [\[CrossRef\]](#)
- Amornsiripanitch N, Bickelhaupt S, Shin HJ, et al. Diffusion-weighted MRI for unenhanced breast cancer screening. *Radiology*. 2019;293(3):504-520. [\[CrossRef\]](#)
- Bickel H, Polanec SH, Wengert G, et al. Diffusion-weighted MRI of breast cancer: improved lesion visibility and image quality using synthetic b-values. *J Magn Reson Imaging*. 2019;50(6):1754-1761. [\[CrossRef\]](#)
- Naranjo ID, Lo Gullo R, Saccarelli C, et al. Diagnostic value of diffusion-weighted imaging with synthetic b-values in breast tumors: comparison with dynamic contrast-enhanced and multiparametric MRI. *Eur Radiol*. 2021;31(1):356-367. [\[CrossRef\]](#)
- Blackledge MD, Wilton B, Messiou C, Koh DM, Leach MO, Collins DJ. Computed diffusion weighted imaging (cDWI) for improving imaging contrast [abstr]. In: Proceedings of the International Society for Magnetic Resonance in Medicine. Berkeley, Calif: International Society for Magnetic Resonance in Medicine. 2009; 4005. [\[CrossRef\]](#)
- Blackledge MD, Collins DJ, Koh DM, Leach MO. SNR of high b-value diffusion weighted images can be improved using cDWI [abstr]. In: Proceedings of the International Society for Magnetic Resonance in Medicine. Berkeley, Calif: International Society for Magnetic Resonance in Medicine. 2010; 4707. [\[CrossRef\]](#)
- Kawahara S, Isoda H, Fujimoto K, et al. Additional benefit of computed diffusion-weighted imaging for detection of hepatic metastases at 1.5T. *Clin Imaging*. 2016;40(3):481-485. [\[CrossRef\]](#)
- Blackledge MD, Leach MO, Collins DJ, Koh DM. Computed diffusion-weighted MR imaging may improve tumor detection. *Radiology*. 2011;261(2):573-581. [\[CrossRef\]](#)
- Park JH, Yun B, Jang M, et al. Comparison of the diagnostic performance of synthetic versus acquired high b-value (1500 s/mm²) diffusion-weighted MRI in women with breast cancers. *J Magn Reson Imaging*. 2019;49(3):857-863. [\[CrossRef\]](#)
- Choi BH, Baek HJ, Young Ha J, et al. Feasibility study of synthetic diffusion-weighted MRI in patients with breast cancer in comparison with conventional diffusion-weighted MRI. *Korean J Radiol*. 2020;21(9):1036-1044. [\[CrossRef\]](#)
- O'Flynn EA, Blackledge M, Collins D, et al. Evaluating the diagnostic sensitivity of computed diffusion-weighted MR imaging in the detection of breast cancer. *J Magn Reson Imaging*. 2016;44(1):130-137. [\[CrossRef\]](#)
- Sickles EA, D'Orsi CJ, Bassett LW. ACR BI-RADS Mammography. In: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System, 5th ed. American College of Radiology, Reston, VA; 2013. [\[CrossRef\]](#)
- Şahin, C, Arıbal E. The role of apparent diffusion coefficient values in the differential diagnosis of breast lesions in diffusion-weighted MRI. *Diagn Interv Radiol*. 2013;19(6):457-462. [\[CrossRef\]](#)
- Yilmaz E, Sarı O, Yilmaz A, et al. Diffusion-weighted imaging for the discrimination of benign and malignant breast masses; utility of ADC and relative ADC. *J Belg Soc Radiol*. 2018;102(1):24. [\[CrossRef\]](#)
- Yilmaz E, Erok B, Atca AÖ. Measurement of the apparent diffusion coefficient in discrimination of benign and malignant axillary lymph nodes. *Pol J Radiol*. 2019;84:592-597. [\[CrossRef\]](#)
- Baltzer P, Mann RM, Lima M, et al. Diffusion-weighted imaging of the breast-a consensus and mission statement from the EUSOBI International Breast Diffusion-Weighted Imaging Working Group. *Eur Radiol*. 2020;30(3):1436-1450. [\[CrossRef\]](#)
- Harvey JA, Hendrick RE, Coll JM, Nicholson BT, Burkholder BT, Cohen MA. Breast MR imaging artifacts: how to recognize and fix them. *Radiographics*. 2007;27(Suppl 1):131-145. [\[CrossRef\]](#)
- Strigel RM, Rollenhagen J, Burnside ES, et al. Screening breast MRI outcomes in routine clinical practice: comparison to BI-RADS benchmarks. *Acad Radiol*. 2017;24(4):411-417. [\[CrossRef\]](#)
- Dietzel M, Ellmann S, Schulz-Wendtland R, et al. Breast MRI in the era of diffusion-weighted imaging: do we still need signal-intensity time curves? *Eur Radiol*. 2020;30(1):47-56. [\[CrossRef\]](#)
- Comstock CE, Gatsonis C, Newstead GM, et al. Comparison of abbreviated breast MRI vs digital breast tomosynthesis for breast cancer detection among women with dense breasts undergoing screening. *JAMA*. 2020;323(8):746-756. [\[CrossRef\]](#)
- Jain M, Jain A, Hyzy MD, Werth G. FAST MRI breast screening revisited. *J Med Imaging Radiat Oncol*. 2017;61(1):24-28. [\[CrossRef\]](#)
- Kang JW, Shin HJ, Shin KC, et al. Unenhanced magnetic resonance screening using fused diffusion-weighted imaging and maximum-intensity projection in patients with a personal history of breast cancer: role of fused DWI for postoperative screening. *Breast Cancer Res Treat*. 2017;165(1):119-128. [\[CrossRef\]](#)
- Yamada T, Kanemaki Y, Okamoto S, Nakajima Y. Comparison of detectability of breast cancer by abbreviated breast MRI based on diffusion-weighted images and postcontrast MRI. *Jpn J Radiol*. 2018;36(5):331-339. [\[CrossRef\]](#)
- Hahn SY, Ko ES, Han BK, Lim Y, Gu S, Ko EY. Analysis of factors influencing the degree of detectability on diffusion-weighted MRI and diffusion background signals in patients with invasive breast cancer. *Medicine (Baltimore)*. 2016;95(27):e4086. [\[CrossRef\]](#)

33. Iaconi C, Thakur SB, Dershaw DD, Brooks J, Fry CW, Morris EA. Impact of fibroglandular tissue and background parenchymal enhancement on diffusion weighted imaging of breast lesions. *Eur J Radiol.* 2014;83(2):2137-2143. [\[CrossRef\]](#)
34. Kul S, Eyuboglu I, Cansu A, Alhan E. Diagnostic efficacy of the diffusion weighted imaging in the characterization of different types of breast lesions. *J Magn Reson Imaging.* 2014;40(5):1158-1164. [\[CrossRef\]](#)
35. Imamura T, Isomoto I, Sueyoshi E, et al. Diagnostic performance of ADC for non-mass-like breast lesions on MR imaging. *Magn Reson Med Sci.* 2010;9(4):217-225. [\[CrossRef\]](#)
36. Yabuuchi H, Matsuo Y, Kamitani T, et al. Non-mass like enhancement non contrast-enhanced breast MR imaging: lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. *Eur J Radiol.* 2010;75(1):126-132. [\[CrossRef\]](#)



Head-to-head comparison of ^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI for lymph node metastasis staging in non-small cell lung cancer: a meta-analysis

Min Zhang¹
 Zhikang Liu¹
 Yuhang Yuan¹
 Wenwen Yang¹
 Xiong Cao^{2,3}
 Minjie Ma^{2,3}
 Biao Han^{2,3}

¹Lanzhou University, The First Clinical Medical College, Lanzhou, China

²The First Hospital of Lanzhou University, Department of Thoracic Surgery, Lanzhou, China

³The First Hospital of Lanzhou University, Gansu Province International Cooperation Base for Research and Application of Key Technology of Thoracic Surgery, Lanzhou, China

PURPOSE

The current meta-analysis aimed to compare the diagnostic performance of ^{18}F -fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F -FDG PET/CT) with ^{18}F -FDG PET/magnetic resonance imaging (MRI) in non-small cell lung cancer (NSCLC) lymph node metastasis staging.

METHODS

We searched the PubMed, Web of Science, and Embase databases for relevant articles between November 1992 and September 2022. Studies evaluating the head-to-head comparison of ^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI for lymph node metastasis in patients with NSCLC were included. The quality of each study was assessed using the Quality Assessment of Diagnostic Performance Studies-2 tool.

RESULTS

The analysis includes six studies with a total of 434 patients. The pooled sensitivity of ^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI was 0.78 [95% confidence interval (CI): 0.59–0.90] and 0.84 (95% CI: 0.68–0.93), and the pooled specificity was 0.87 (95% CI: 0.72–0.94) and 0.87 (95% CI: 0.80–0.92), respectively. The accuracy of ^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI was 0.81 (95% CI: 0.71–0.90) and 0.84 (95% CI: 0.75–0.92), respectively. When the pre-test probability was set at 50%, the post-test probability for ^{18}F -FDG PET/CT could increase to 85%, and the post-test probability for ^{18}F -FDG PET/MRI could increase to 87%.

CONCLUSION

^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI have similar diagnostic performance in detecting lymph node metastasis in NSCLC. However, the results of this study were from a small sample study, and further studies with larger sample sizes are needed.

KEYWORDS

^{18}F -FDG PET/CT, ^{18}F -FDG PET/MRI, lymph node metastasis, non-small cell lung cancer, meta-analysis

Corresponding author: Biao Han

E-mail: hanbiao66@163.com

Received 27 April 2023; revision requested 05 June 2023; last revision received 31 July 2023; accepted 16 October 2023.



Epub: 31.01.2024

Publication date: 05.03.2024

DOI: 10.4274/dir.2023.232280

According to the 2020 global cancer incidence and mortality statistics of the Global Cancer Observatory database, lung cancer has the highest mortality rate (approximately 18% of all cancer deaths) and the second highest incidence rate (approximately 11.4% of all new cancer cases).¹ The most common type of lung cancer is non-small cell lung cancer (NSCLC), which accounts for approximately 80% of all lung cancers.^{2,3} The assessment of distant metastases and metastases to mediastinal lymph nodes in patients with NSCLC is critical not only for providing information about the staging of the disease but also for guiding treatment options and determining the patient's prognosis.^{4,5}

Although computed tomography (CT) is the most-used non-invasive modality for assessing mediastinal staging in NSCLC, numerous studies have shown that CT has limited sensi-

tivity and reliability in lymph node staging.⁶⁻⁸ ¹⁸F-fluorodeoxyglucose positron emission tomography/CT (¹⁸F-FDG-PET) has been widely used to evaluate NSCLC over the last decade, as it can distinguish malignant isolated pulmonary nodules from benign lesions, improve staging accuracy, and anticipate histology, treatment response, and prognosis.⁹ ¹⁸F-FDG PET/magnetic resonance imaging (MRI) is a hybrid imaging modality. It provides useful information about metabolic activity as well as tumor cells while reducing radiation exposure and is now increasingly used in the diagnosis of NSCLC.¹⁰ Kajiyama et al.^{11,12} showed that both ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI had more accurate pathological staging results than CT in the diagnosis of hilar and mediastinal lymph node metastases in NSCLC.

Over the past decade, radionuclide imaging techniques, including PET/CT and PET/MRI, have been widely used in the diagnosis of NSCLC and have gained much attention for their better diagnostic performance, compared with CT. However, which diagnostic tool has better diagnostic performance remains controversial. According to one report, PET/MRI may have advantages over PET/CT in terms of radiation dose management and local staging accuracy when evaluating thoracic tumors,¹⁰ whereas another study demonstrated that PET/MRI and PET/CT have equivalent performance when it comes to evaluating the preoperative thoracic staging of NSCLC patients.¹³

Although many studies have reported that ¹⁸F-FDG PET/CT performs well in assessing lymph node metastasis staging in NSCLC, few have quantified its performance in comparison with ¹⁸F-FDG PET/MRI. The purpose

of the current study was to include head-to-head comparison articles comparing the diagnostic efficacy of the two diagnostic modalities for the staging of lymph node metastasis in NSCLC.

Methods

Search strategy

All available literature was searched in the PubMed, Embase, and Web of Science databases between November 1992 and September 2022. The keywords were based on the following: (Carcinoma, Non Small Cell Lung) OR (Carcinomas, Non-Small-Cell Lung) OR (Lung Carcinoma, Non-Small-Cell) OR (Lung Carcinomas, Non-Small-Cell) OR (Non-Small-Cell Lung Carcinomas) OR (Non-Small-Cell Lung Carcinoma) OR (Non Small Cell Lung Carcinoma) OR (Nonsmall Cell Lung Cancer) OR (Non-Small Cell Lung Cancer) OR (NSCLC) OR ("Carcinoma, Non-Small-Cell Lung"[Mesh]) AND (PET-MRI) OR (positron emission tomography/magnetic resonance imaging) OR (PET-MR) OR (positron emission tomography/magnetic resonance).

Inclusion and exclusion criteria

Studies were considered for inclusion if all the following criteria were satisfied: (a) patients with NSCLC who were evaluated for N-stage cancer before starting treatment; (b) head-to-head comparison of ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI; (c) retrospective or prospective original research.

The exclusion criteria were (a) duplicated articles; (b) abstract, case reports, letters, reviews, or meta-analyses; (c) non-English full-text articles; (d) irrelevant titles and abstracts; (e) data unavailable; (d) lesion-based studies.

Two researchers independently reviewed the remaining texts' titles and abstracts, as well as the full-text versions, to determine their eligibility for inclusion in the next stage using the aforementioned inclusion and exclusion criteria. The two researchers resolved disagreements by reaching a consensus.

Quality assessment

The two researchers independently used the Quality Assessment of Diagnostic Performance Studies-2 (QUADAS-2) tool to evaluate the quality of each study.¹⁴ The following criteria were used to evaluate each study: patient selection, index test, reference standard, flow, and timing. Based on the bias risk, these domains were then classified as high, low, or uncertain in terms of applicability. Disagreements that arose during the evalu-

ation process were resolved by a third-party researcher.

Data extraction

Data extracted for all included articles included first author, year, country, study design (retrospective or prospective study), patient characteristics (sample size, mean age), study period, interval between the ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI scans, and reference for lymph node metastasis of NSCLC. The numbers of true-positive, true-negative, false-positive, and false-negative results for ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI on a patient-by-patient basis were also extracted for each study. In addition, data were extracted on technical aspects of each study including scanner modality, ligand dose, and image analysis. All the above data extraction was done independently by two researchers, and any differences were resolved through consensus. This analysis did not require ethics committee or patient approval.

Statistical analysis

The heterogeneity of the threshold effect among pooled studies was assessed using the Spearman correlation coefficient. A value of $P < 0.05$ indicated a statistically significant threshold effect. A bivariate random effects model was used to calculate pooled estimates of sensitivity and specificity. A Fagan diagram was used to evaluate the pre-test and post-test probabilities of the testing tool.

The heterogeneity of non-threshold effects among pooled studies was assessed using inconsistency index (I^2) statistics and the Cochran Q test. A value of $I^2 > 50\%$ or $P < 0.1$ for the Cochran Q test indicated a statistically significant non-threshold effect. Due to the small number of included studies, sensitivity analysis was performed, rather than meta-regression or subgroup analysis.

A Deeks' funnel plot was used to evaluate the publication bias of the included studies. A P value of < 0.05 was deemed to indicate publication bias. The statistical analysis was performed using STATA v15.1 (Stata-Corp, College Station, TX, USA, Review Manager v5.4 (the Nordic Cochrane Centre, Copenhagen, Denmark) and MetaDisc v1.4.

Results

Literature search and study selection

The literature search led to the initial identification of 460 publications. Ninety-three duplicate studies were excluded, 281 studies were excluded by title and abstract, and

Main points

- Our meta-analysis showed that ¹⁸F-fluorodeoxyglucose positron emission tomography/computed tomography (¹⁸F-FDG PET/CT) has good diagnostic potential for non-small cell lung cancer (NSCLC) lymph node metastases, with a pooled sensitivity of 0.78 [95% confidence interval (CI): 0.59–0.90] and a pooled specificity of 0.87 (95% CI: 0.72–0.94).
- ¹⁸F-FDG PET/magnetic resonance imaging (MRI) had a pooled sensitivity of 0.84 (95% CI: 0.68–0.93) and a pooled specificity of 0.87 (95% CI: 0.80–0.92), which had better diagnostic ability for lymph node metastasis in NSCLC.
- ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI have similar diagnostic performance in detecting lymph node metastasis in NSCLC.

72 studies were excluded by article category (review, abstract, case report, meta-analysis). The remaining 14 studies were carefully assessed by full text, and were excluded for the following reasons: not meeting the inclusion criteria (n = 5); data unavailable (n = 3). Finally, 6 articles evaluating head-to-head comparison of ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI for lymph node metastasis in patients with NSCLC were qualified for meta-analysis.¹⁵⁻²⁰ A PRISMA flow diagram of the study selection process is shown in Figure 1.

Study description and quality assessment

The 6 eligible studies contained a total of 434 patients with NSCLC who were evaluated for N-stage cancer before starting treatment, were published between 2014 and 2020, and had a sample size ranging from 22 to 140. Table 1 summarizes the study and patient characteristics. The technical aspects of ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI are shown in Table 2. The QUADAS-2 tool was used to assess the risk of bias in these studies, as shown in Figure 2. None of the studies had a “high” risk of bias, according to the QUADAS-2 suggestions. The included studies were deemed to be of adequate quality.

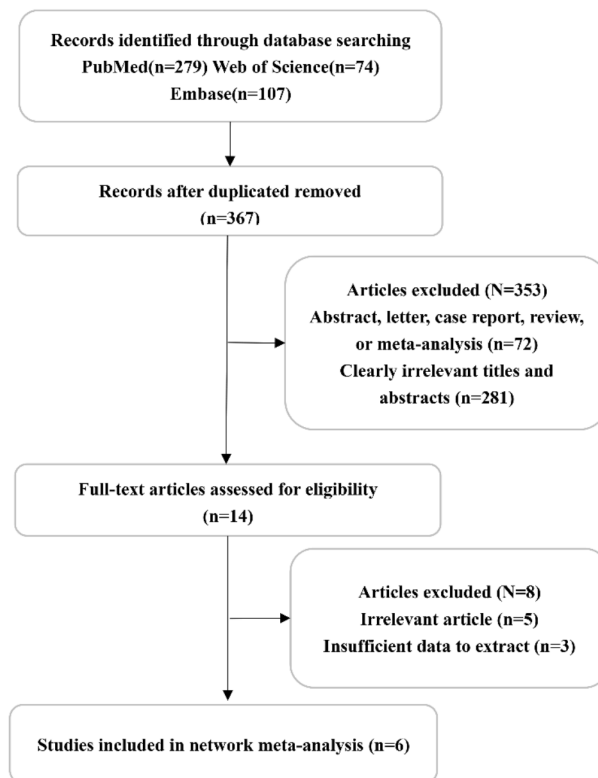


Figure 1. PRISMA flow diagram for study selection.

Table 1. Study characteristics and patient characteristics of the included studies

First author	Year	Country	Study design	Sample size (n)	Age (y) ^a	Study period	Interval between the two imaging tests	Reference standard	¹⁸ F-FDG PET/CT				¹⁸ F-FDG PET/MRI			
									TP	FP	FN	TN	TP	FP	FN	TN
Ohno et al. ²⁰	2020	Japan	Retro	104	71 ± 6.3 (43–85)	2014–2015	<3 wk	PA	23	3	18	60	33	8	8	55
Kirchner et al. ¹⁹	2018	Germany	Pro	84	62.5 ± 9.1	NA	<1 d	PA	42	1	5	36	42	2	5	35
Lee et al. ¹⁸	2016	Korea	Pro	42	62.9 ± 9.9 (35–79)	2013–2014	<1 h	PA	10	9	11	12	8	5	13	16
Huellner et al. ¹⁷	2016	Switzerland	Retro	42	65 (35–89)	2012–2014	<1 h	PA	31	3	1	7	28	4	4	6
Ohno et al. ¹⁶	2015	Japan	Retro	140	72 ± 7.4 (47–83)	2012–2013	<3 wk	PA	48	13	14	65	58	8	4	70
Heusch et al. ¹⁵	2014	Germany	Pro	22	65 ± 9.1	NA	NA	PA	6	2	2	12	7	1	1	13

n, the numbers of patients included in the study; Retro, retrospective; Pro, prospective; ^adata are mean (range) or mean ± standard deviation (range); NA, not available; PA, pathology; TP, true positive; FP, false positive; FN, false negative; TN, true negative; ¹⁸F-FDG PET/CT, ¹⁸F-fluorodeoxyglucose positron emission tomography/computed tomography; MRI, magnetic resonance imaging.

Table 2. Technical aspects of included studies

First author	Year	Scanner modality (PET/CT)	Scanner modality (PET/MRI)	Ligand dose	Image analysis
Ohno et al. ²⁰	2020	GE Healthcare	Canon Medical Systems	3.3 MBq/kg	Quantitative
Kirchner et al. ¹⁹	2018	Siemens, Healthcare GmbH, Erlangen, Germany	Siemens Healthcare GmbH, Erlangen, Germany	275.7 ± 47.4 MBq	Quantitative
Lee et al. ¹⁸	2016	Siemens Healthcare, Erlangen, Germany	Siemens Medical Solutions, Knoxville, TN	5.2 MBq/kg	Quantitative
Huellner et al. ¹⁷	2016	GE Healthcare, Waukesha, WI, USA	GE Healthcare, Waukesha, WI, USA	350 MBq	Quantitative
Ohno et al. ¹⁶	2015	GE Healthcare, Milwaukee, Wis	Toshiba Medical Systems, Otawara, Japan	132–300 MBq	Quantitative
Heusch et al. ¹⁵	2014	Siemens Molecular Imaging	Siemens Healthcare	300 ± 45 MBq	Quantitative

PET/CT, positron emission tomography/computed tomography; MRI, magnetic resonance imaging.

Quantitative synthesis

The analysis includes six studies with a total of 434 patients. For ¹⁸F-FDG PET/CT, the results of the Spearman correlation coefficient demonstrated no threshold effect heterogeneity (Spearman correlation coefficient: -0.200; *P* = 0.704); the forest plot demonstrated a pooled sensitivity of 0.78 (95% CI: 0.59–0.90) and a pooled specificity of 0.87 (95% CI: 0.72–0.94); the heterogeneity results obtained by *I*² were 83.6% for sensitivity and 83.4% for specificity (Figure 3), which was statistically significant in both sensitivity and specificity (*I*² > 50%). The accuracy of ¹⁸F-FDG PET/CT in diagnosing NSCLC lymph node metastasis was 0.81 (95% CI: 0.71–0.90). Furthermore, the Deeks' funnel plot of ¹⁸F-FDG PET/CT revealed no publication bias in the included studies (*P* = 0.802) (Figure 4). The Fagan nomogram indicated that when the pre-test probability was set at 50%, the post-test probability for ¹⁸F-FDG PET/CT could increase to 85% (Figure 5).

For ¹⁸F-FDG PET/MRI, the results of the Spearman correlation coefficient demonstrated no threshold effect heterogeneity (Spearman correlation coefficient: -0.551; *P* = 0.257); the forest plot demonstrated a pooled sensitivity of 0.84 (95% CI: 0.68–0.93) and a pooled specificity of 0.87 (95% CI: 0.80–0.92); the heterogeneity results obtained by *I*² were 86.6% for sensitivity and 56.6% for specificity (Figure 6), which was statistically significant in both sensitivity and specificity (*I*² > 50%). The accuracy of ¹⁸F-FDG PET/MRI in diagnosing NSCLC lymph node metastasis was 0.84 (95% CI: 0.75–0.92). Moreover, the Deeks' funnel plot of ¹⁸F-FDG PET/MRI revealed no publication bias in the included studies (*P* = 0.310) (Figure 7). The Fagan nomogram indicated that when the pre-test probability was set at 50%, the post-test probability for ¹⁸F-FDG PET/MRI could increase to 87% (Figure 8).

Heterogeneity analysis

For ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI, the *I*² for their pooled sensitivity were 83.6% (*P* < 0.001) and 86.6% (*P* < 0.001), and for their pooled specificity were 83.4% (*P* < 0.001) and 56.6% (*P* = 0.042), respectively. This demonstrated that both ¹⁸F-FDG PET/CT and ¹⁸F-FDG PET/MRI had high heterogeneity. For ¹⁸F-FDG PET/MRI, sensitivity analysis by excluding data from Lee et al.¹⁸ demonstrated a pooled sensitivity of 0.88 (95% CI: 0.82–0.93), with acceptable heterogeneity (*I*² = 4.3%), and excluding data from Huellner et al.¹⁷ showed a pooled specificity of 0.88 (95%

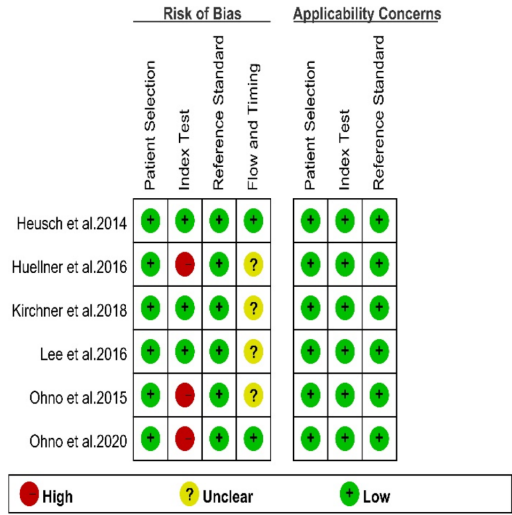


Figure 2. Summary risk of bias and applicability concerns of the included studies.

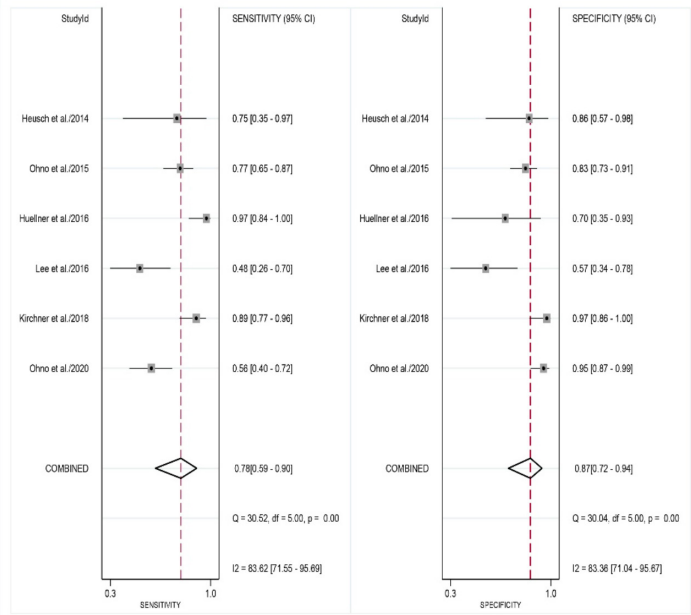


Figure 3. Forest plot showing the pooled sensitivity and specificity of ¹⁸F-fluorodeoxyglucose positron emission tomography/computed tomography.

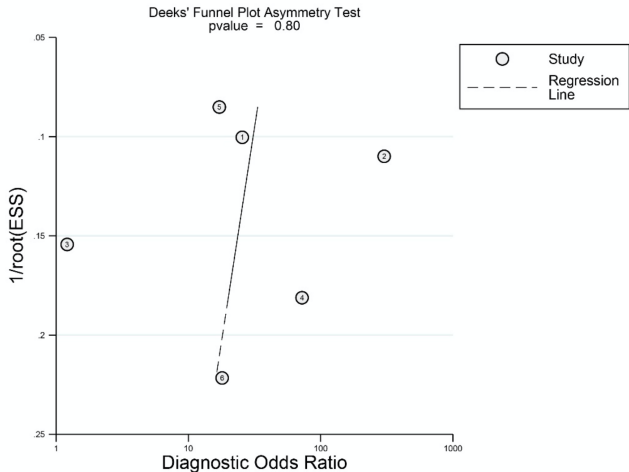


Figure 4. Deeks' funnel plot showing the publication bias of the included studies ¹⁸F-fluorodeoxyglucose positron emission tomography/computed tomography.

CI: 0.81–0.93) with reasonable heterogeneity ($I^2 = 20.8\%$). For ^{18}F -FDG PET/CT, sensitivity analysis was unable to identify the source of heterogeneity. Table 3 shows all the results of the sensitivity analysis.

Discussion

NSCLC has long been an issue of great importance to surgeons, as it is an important factor affecting and determining staging and prognosis.²¹ ^{18}F -FDG PET/CT is increasingly being used to diagnose NSCLC. Since CT can obtain anatomical information about tumor size and location, and FDG-PET can obtain metabolic information about the tissue, this gives ^{18}F -FDG PET/CT a unique advantage in detecting lymph node metastases.^{22,23} MRI has a greater ability to detect pleural and mediastinal involvement and a higher sensitivity to detect brain, liver, and bone metastases. Therefore, ^{18}F -FDG-PET combined with ^{18}F -FDG PET/MRI has also become the mainstream diagnostic tool for chest tumors in the past decade.^{24–26}

To our knowledge, this is one of the few meta-analyses of a head-to-head comparison of ^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI to determine their performance in the diagnosis of lymph node metastasis in NSCLC. According to Kirchner et al.¹⁹, ^{18}F -FDG PET/MRI and ^{18}F -FDG PET/CT have comparable diagnostic performance for T- and N-staging in patients with NSCLC.¹⁹ However, Laffon and Marthan²⁷ suggested that the equivalence between the two imaging techniques reported in the previous study could be due to experimental design, and they concluded that ^{18}F -FDG PET/MRI has greater value for NSCLC chest staging and may even replace ^{18}F -FDG PET/CT. Therefore, a meta-analysis was conducted to compare the performance of the two diagnostic modalities.

In this meta-analysis, we systematically reviewed and compared the ability of two imaging modalities in the detection of lymph node metastases in NSCLC. In the detection of lymph node metastasis in NSCLC, the pooled sensitivity of ^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI were 0.78 (95% CI: 0.59–0.90) and 0.84 (95% CI: 0.68–0.93), respectively. Sun et al.²⁸ reported a pooled sensitivity of 0.68 (95% CI: 0.61–0.75) and a pooled specificity of 0.93 (95% CI: 0.89–0.95) for PET/CT to diagnose lymph node metastasis in NSCLC in a current study. Furthermore, Seol et al.²⁹ showed the pooled sensitivity for ^{18}F -FDG PET was 0.79 (95% CI: 0.70–0.86) and

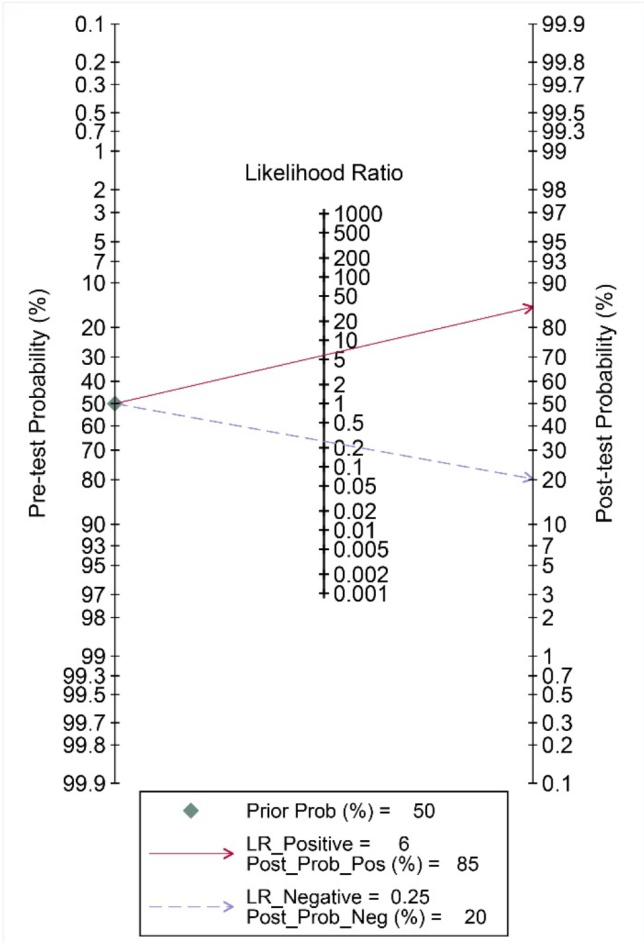


Figure 5. Fagan diagrams showing the pre-test and post-test probabilities of ^{18}F -fluorodeoxyglucose positron emission tomography/computed tomography.

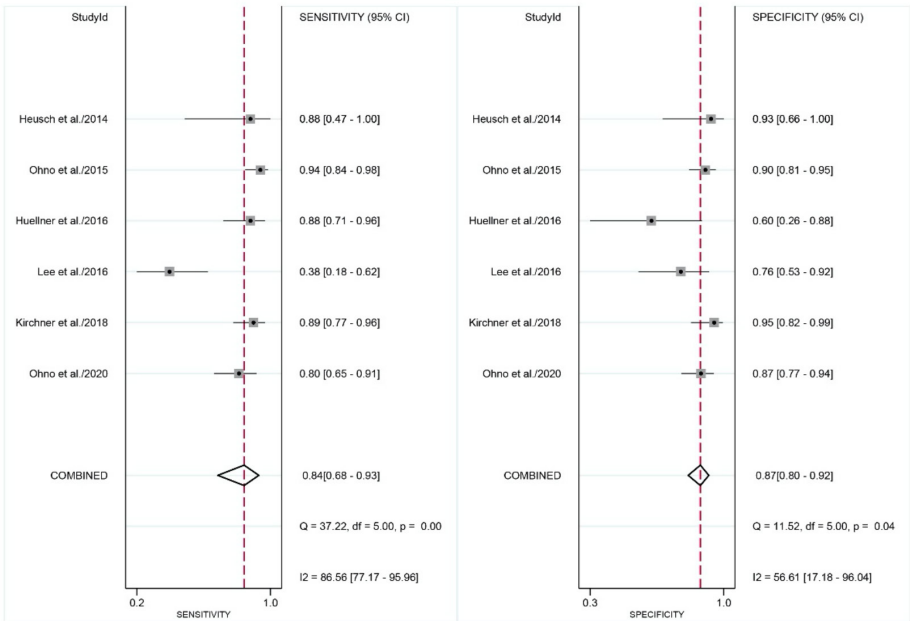


Figure 6. Forest plot showing the pooled sensitivity and specificity of ^{18}F -fluorodeoxyglucose positron emission tomography/magnetic resonance imaging.

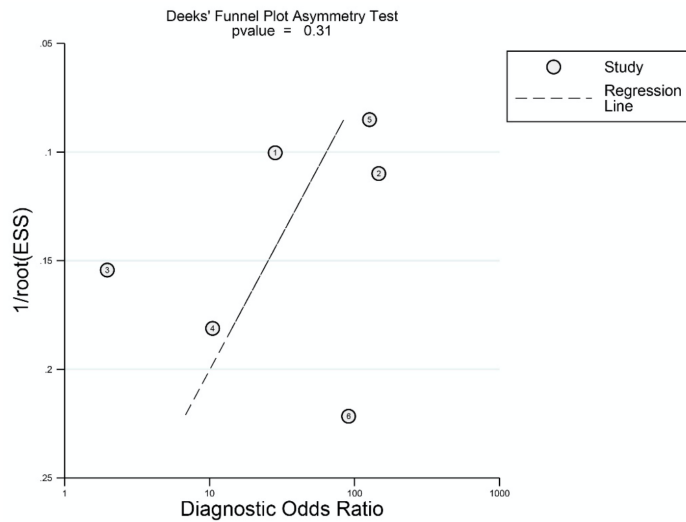


Figure 7. Deeks' funnel plot showing the publication bias of the included studies of ^{18}F -fluorodeoxyglucose positron emission tomography/magnetic resonance imaging.

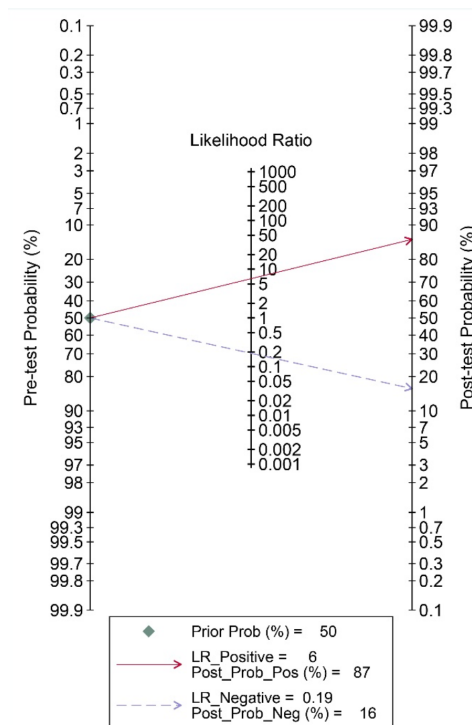


Figure 8. Fagan diagrams showing the pre-test and post-test probabilities of ^{18}F -fluorodeoxyglucose positron emission tomography/magnetic resonance imaging.

a pooled specificity of 0.65 (95% CI: 0.57–0.72) in their study. In terms of sensitivity, our meta-analysis did not differ significantly from previous studies, but it did show higher results in terms of specificity, which could be attributed to the small sample size of our included studies, which required the use of both detection tools in the same patient cohort.

Our meta-analysis showed that both ^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI have high diagnostic performance for lymph node metastases of NSCLC, and we believe that both diagnostic tools have the potential to be used more often in the clinic in the future. However, by reviewing other related studies, we also found that these two diagnostic tools have their shortcomings. There are two key limitations to PET CT: first, it involves a relatively high radiation exposure; second, it has relatively low spatial resolution.³⁰ The following are the primary drawbacks of PET MRI: compared with PET/CT, it needs a specific lung imaging procedure, and the examination is significantly more time-consuming;¹⁵ however, it was introduced relatively recently and has not been studied extensively, and numerous pertinent studies and clinical trials will be required in the future to incorporate it into clinical practice.²⁶ In addition, it has been shown that both diagnostic tools have limited evaluation in the detection of microscopic nodules in the lung.³¹ To produce novel and promising findings, more research comparing these two models head-to-head is required in the future.

In addition, ^{18}F -FDG PET/MRI has been studied extensively as a novel diagnostic tool for applications in other diseases. A study of ^{18}F FDG PET/MRI for the diagnosis of bladder cancer showed that it has a better ability to detect metastatic lesions as well as soft tissue lesions compared with conventional CT, thus allowing better differentiation between primary bladder tumors and pelvic metastases.³² Another review of rectal cancer

Table 3. Sensitivity analysis of overall detection rate for ^{18}F -FDG PET/CT and ^{18}F -FDG PET/MRI

	^{18}F -FDG PET/CT			^{18}F -FDG PET/MRI				
	Sensitivity (95% CI)	I^2	Specificity (95% CI)	I^2	Sensitivity (95% CI)	I^2	Specificity (95% CI)	I^2
Omitting Ohno et al.	0.81 (0.63–0.91)	82.8%	0.84 (0.70–0.93)	78.5%	0.84 (0.64–0.94)	89.5%	0.87 (0.78–0.92)	64.8%
Omitting Kirchner et al. ¹⁹	0.75 (0.52–0.89)	81.8%	0.82 (0.66–0.92)	79.9%	0.82 (0.61–0.93)	87.8%	0.85 (0.78–0.91)	53.0%
Omitting Lee et al. ¹⁸	0.83 (0.65–0.92)	83.0%	0.89 (0.78–0.95)	72.7%	0.88 (0.82–0.93)	4.3%	0.89 (0.83–0.92)	59.4%
Omitting Huellner et al. ¹⁷	0.71 (0.54–0.84)	78.9%	0.88 (0.72–0.96)	84.4%	0.83 (0.62–0.93)	88.7%	0.88 (0.81–0.93)	20.8%
Omitting Ohno et al. ¹⁶	0.79 (0.54–0.92)	86.4%	0.87 (0.69–0.96)	88.7%	0.80 (0.61–0.91)	84.9%	0.86 (0.78–0.92)	63.4%
Omitting Heusch et al. ¹⁵	0.78 (0.56–0.91)	87.0%	0.87 (0.69–0.95)	87.1%	0.83 (0.64–0.93)	88.8%	0.86 (0.79–0.91)	62.3%

I^2 , inconsistency index; ^{18}F -FDG PET/CT, ^{18}F -fluorodeoxyglucose positron emission tomography/computed tomography; MRI, magnetic resonance imaging; CI, confidence interval.

indicated that ^{18}F FDG PET/MRI could be utilized to restage rectal cancer after preoperative chemoradiotherapy or to detect recurrence. Furthermore, because it is more accurate in T-staging and N-staging than PET/CT or MRI, it can be a precise tool for determining which patients to use for rectal preservation rather than standard surgery.³³ In addition, we focused on the concordance between ^{18}F FDG semiquantitative metrics from PET/MRI and PET/CT in the included studies. One of the studies we included showed that the mean difference in standardized uptake value (SUV_{mean} and SUV_{max} for NSCLC from ^{18}F FDG PET/MR imaging and ^{18}F FDG-PET/CT was not statistically significant and showed a high correlation.¹⁵ In contrast, another study showed that the SUV_{max} of PET/CT was significantly higher than that of PET/MR in primary foci,¹⁸ which may be due to the differences in the hardware of PET/CT and PET/MRI devices and the reconstruction software methods used.

According to our meta-analysis, there was a high heterogeneity in the pooled sensitivity and specificity of ^{18}F FDG PET/CT and ^{18}F FDG PET/MRI, and we explored the sources of heterogeneity that would result from the inclusion of studies through sensitivity analysis. For ^{18}F FDG PET/MRI, we discovered that by omitting the data from Lee et al.¹⁸, a reasonable heterogeneity in pooled sensitivity was obtained, and an acceptable heterogeneity in pooled specificity was obtained by excluding the data from Huellner et al.¹⁷, which could be explained by different cut-off thresholds. This may be related to the fact that these two articles included patients with suspected NSCLC, whereas several other studies included patients with NSCLC confirmed by pathologic examination. Nevertheless, other causes, such as changes in the patients, method, and research design, are also possible. Regrettably, we were unable to identify a source of heterogeneity in ^{18}F FDG PET/CT.

However, our meta-analysis has limitations that cannot be ignored. First, we searched only three databases, which may have caused us to omit some studies that were consistent with this study. Second, the number of included studies was too small and they were all small sample size studies, which may be related to the included articles all required the use of ^{18}F FDG PET/CT and ^{18}F FDG PET/MRI in the same patient cohort. Third, half of the included articles were retrospective studies, and more prospective studies are needed in the future. Finally, no consistent source of ^{18}F FDG PET/CT pooled

sensitivity and specificity was found by sensitivity analysis. We must interpret these results cautiously due to these limitations.

Based on the results pooled in the meta-analysis, ^{18}F FDG PET/CT has good diagnostic potential for NSCLC lymph node metastases with a pooled sensitivity of 0.78 (95% CI: 0.59–0.90) and a pooled specificity of 0.87 (95% CI: 0.72–0.94), and ^{18}F FDG PET/MRI had a pooled sensitivity of 0.84 (95% CI: 0.68–0.93) and a pooled specificity of 0.87 (95% CI: 0.80–0.92), which also had better diagnostic ability for lymph node metastasis in NSCLC. Therefore, we conclude that ^{18}F FDG PET/CT and ^{18}F FDG PET/MRI have similar diagnostic performance in detecting lymph node metastasis in NSCLC. Nevertheless, the results of this analysis were from a small sample study, and further studies with larger sample sizes are needed to draw more convincing conclusions.

^{18}F FDG PET/CT and ^{18}F FDG PET/MRI have similar diagnostic performance in detecting lymph node metastasis in NSCLC. Nevertheless, the results of this study were from a small sample study, and further studies with larger sample sizes are needed.

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

Supported by Education Science and Technology Innovation Project of Gan-su Province (2022B-014) and First Hospital of Lanzhou University Intramural Fund (Idyyyn2021-66).

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. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc.* 2008;83(5):584–594. [\[CrossRef\]](#)
3. Zheng X, Cheng M, Fu B, et al. Targeting LUNX inhibits non-small cell lung cancer growth and metastasis. *Cancer Res.* 2015;75(6):1080–1090. [\[CrossRef\]](#)
4. Pieterman RM, van Putten JW, Meuzelaar JJ, et al. Preoperative staging of non-small-cell lung cancer with positron-emission tomography. *N Engl J Med.* 2000;343(4):254–261. [\[CrossRef\]](#)
5. De Leyn P, Doooms C, Kuzdzal J, et al. Revised ESTS guidelines for preoperative mediastinal

lymph node staging for non-small-cell lung cancer. *Eur J Cardiothorac Surg.* 2014;45(5):787–798. [\[CrossRef\]](#)

6. Gould MK, Kuschner WG, Rydzak CE, et al. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: a meta-analysis. *Ann Intern Med.* 2003;139(11):879–892. [\[CrossRef\]](#)
7. Alongi F, Ragusa P, Montemaggi P, Bona CM. Combining independent studies of diagnostic fluorodeoxyglucose positron-emission tomography and computed tomography in mediastinal lymph node staging for non-small cell lung cancer. *Tumori.* 2006;92(4):327–333. [\[CrossRef\]](#)
8. Silvestri GA, Gould MK, Margolis ML, et al. Noninvasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines (2nd edition). *Chest.* 2007;132(3 Suppl):178–201. [\[CrossRef\]](#)
9. Manafi-Farid R, Askari E, Shiri I, et al. (18)F]FDG-PET/CT radiomics and artificial intelligence in lung cancer: technical aspects and potential clinical applications. *Semin Nucl Med.* 2022;52(6):759–780. [\[CrossRef\]](#)
10. Kaseda K. Recent and current advances in FDG-PET imaging within the field of clinical oncology in NSCLC: a review of the literature. *Diagnostics (Basel).* 2020;10(8):561. [\[CrossRef\]](#)
11. Al-Ibraheem A, Hirmas N, Fanti S, et al. Impact of (18)F-FDG PET/CT, CT and EBUS/TBNA on preoperative mediastinal nodal staging of NSCLC. *BMC Med Imaging.* 2021;21(1):49. [\[CrossRef\]](#)
12. Kajiyama A, Ito K, Watanabe H, et al. Consistency and prognostic value of preoperative staging and postoperative pathological staging using (18)F-FDG PET/MRI in patients with non-small cell lung cancer. *Ann Nucl Med.* 2022;36:1059–1072. [\[CrossRef\]](#)
13. Wang ML, Zhang H, Yu HJ, et al. An initial study on the comparison of diagnostic performance of (18)F-FDG PET/MR and (18)F-FDG PET/CT for thoracic staging of non-small cell lung cancer: focus on pleural invasion. *Rev Esp Med Nucl Imagen Mol (Engl Ed).* 2023;42(1):16–23. [\[CrossRef\]](#)
14. 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. [\[CrossRef\]](#)
15. Heusch P, Buchbender C, Köhler J, et al. Thoracic staging in lung cancer: prospective comparison of ^{18}F FDG PET/MR imaging and ^{18}F FDG PET/CT. *J Nucl Med.* 2014;55(3):373–378. [\[CrossRef\]](#)
16. Ohno Y, Koyama H, Yoshikawa T, et al. Three-way comparison of whole-body MR, coregistered whole-body FDG PET/MR, and integrated whole-body FDG PET/CT imaging: TNM and stage assessment capability for non-small cell lung cancer patients. *Radiology.* 2015;275(3):849–861. [\[CrossRef\]](#)

17. Huellner MW, de Galiza Barbosa F, Husmann L, et al. TNM staging of non-small cell lung cancer: comparison of PET/MR and PET/CT. *J Nucl Med*. 2016;57(1):21-26. [\[CrossRef\]](#)
18. Lee SM, Goo JM, Park CM, et al. Preoperative staging of non-small cell lung cancer: prospective comparison of PET/MR and PET/CT. *Eur Radiol*. 2016;26(11):3850-3857. [\[CrossRef\]](#)
19. Kirchner J, Sawicki LM, Schaarschmidt BM, et al. Prospective comparison of 18F-FDG PET/MRI and 18F-FDG PET/CT for thoracic staging of non-small cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2019;437-445. [\[CrossRef\]](#)
20. Ohno Y, Takeshi Y, Takenaka D, Koyama H, Aoyagi K, Yui M. Comparison of diagnostic accuracy for TNM stage among whole-body MRI and coregistered PET/MRI using 1.5-T and 3-T MRI systems and integrated PET/CT for non-small cell lung cancer. *AJR. Am J Roentgenol*. 2020;215(5):1191-1198. [\[CrossRef\]](#)
21. Asamura H, Chansky K, Crowley J, et al. The international association for the study of lung cancer lung cancer staging project: proposals for the revision of the N descriptors in the forthcoming 8th edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2015;10(12):1675-1684. [\[CrossRef\]](#)
22. Gambhir SS. Molecular imaging of cancer with positron emission tomography. *Nat Rev Cancer*. 2002;2(9):683-693. [\[CrossRef\]](#)
23. Steinert HC. PET and PET-CT of lung cancer. *Methods Mol Biol*. 2011;727:33-51. [\[CrossRef\]](#)
24. Landwehr P, Schulte O, Lackner K. MR imaging of the chest: mediastinum and chest wall. *Eur Radiol*. 1999;9(9):1737-1744. [\[CrossRef\]](#)
25. Beiderwellen K, Huebner M, Heusch P, et al. Whole-body [¹⁸F]FDG PET/MRI vs. PET/CT in the assessment of bone lesions in oncological patients: initial results. *Eur Radiol*. 2014;24(8):2023-2030. [\[CrossRef\]](#)
26. Sotoudeh H, Sharma A, Fowler KJ, McConathy J, Dehdashti F. Clinical application of PET/MRI in oncology. *J Magn Reson Imaging*. 2016;44(2):265-276. [\[CrossRef\]](#)
27. Laffon E, Marthan R. Performance of ¹⁸F-FDG PET/MRI and ¹⁸F-FDG PET/CT for T and N staging in patients with non-small-cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2019;46(2):522-523. [\[CrossRef\]](#)
28. Sun J, Li Y, Gong F, et al. The diagnostic value of PET/CT for the lymph node metastasis in Asian patients with non-small cell lung cancer: a meta-analysis. *Hell J Nucl Med*. 2022;25(2):196-204. [\[CrossRef\]](#)
29. Seol HY, Kim YS, Kim SJ. Predictive value of 18F-fluorodeoxyglucose positron emission tomography or positron emission tomography/computed tomography for assessment of occult lymph node metastasis in non-small cell lung cancer. *Oncology*. 2021;99(2):96-104. [\[CrossRef\]](#)
30. Czernin J, Ta L, Herrmann K. Does PET/MR imaging improve cancer assessments? Literature evidence from more than 900 patients. *J Nucl Med*. 2014;55(Suppl 2):59-62. [\[CrossRef\]](#)
31. Truong MT, Ko JP, Rossi SE, et al. Update in the evaluation of the solitary pulmonary nodule. *Radiographics*. 2014;34(6):1658-1679. [\[CrossRef\]](#)
32. Civelek AC, Niglio SA, Malayeri AA, et al. Clinical value of (18)FDG PET/MRI in muscle-invasive, locally advanced, and metastatic bladder cancer. *Urol Oncol*. 2021;39(11):787. [\[CrossRef\]](#)
33. Crimi F, Vallenggia S, Baffoni L, et al. [18F] FDG PET/MRI in rectal cancer. *Ann Nucl Med*. 2021;35(3):281-290. [\[CrossRef\]](#)



Copyright©Author(s) - Available online at dirjournal.org.
Content of this journal is licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License.

Preoperative simulation results and intraoperative image fusion guidance for transjugular intrahepatic portosystemic shunt placement: a feasibility study of nineteen patients

Xin Wei¹
 Hong Hu¹
 Lin Qi²
 Liming Zhong¹
 Yunguo Liao¹
 Jiaqi Pu¹

¹Nanchong Central Hospital, The Second Clinical College of North Sichuan Medical College, Department of Interventional Radiology, Nanchong, China

²North Sichuan Medical College, Nanchong, School of Medical Imaging, Nanchong, China

Corresponding author: Xin Wei

E-mail: wxzyy1998@sina.com

Received 10 June 2022; revision requested 10 September 2022; last revision received 14 November 2022; accepted 29 December 2022.



Epub: 24.03.2023

Publication date: 05.03.2024

DOI: 10.4274/dir.2022.221652

PURPOSE

The purpose is to evaluate the feasibility and efficacy of preoperative simulation results and intraoperative image fusion guidance during transjugular intrahepatic portosystemic shunt (TIPS) creation.

METHODS

Nineteen patients were enrolled in the present study. The three-dimensional (3D) structures of the bone, liver, portal vein, inferior vena cava, and hepatic vein in the contrast-enhanced computed tomography (CT) scanning area were reconstructed in the Mimics software. The virtual Rosch-Uchida liver access set and the VIATORR stent model were established in the 3D Max software. The puncture path from the hepatic vein to the portal vein and the release position of the stent were simulated in the Mimics and 3D Max software, respectively. The simulation results were exported to Photoshop software, and the 3D reconstructed top of the liver diaphragm was used as the registration point to fuse with the liver diaphragmatic surface of the intraoperative fluoroscopy image. The selected portal vein system fusion image was overlaid on the reference display screen to provide image guidance during the operation. As a control, the last 19 consecutive cases of portal vein puncture under the guidance of conventional fluoroscopy were analyzed retrospectively, including the number of puncture attempts, puncture time, total procedure time, total fluoroscopy time, and total exposure dose (dose area product).

RESULTS

The average time of preoperative simulation was about 61.26 ± 6.98 minutes. The average time of intraoperative image fusion was 6.05 ± 1.13 minutes. The median number of puncture attempts was not significantly different between the study group ($n = 3$) and the control group ($n = 3$; $P = 0.175$). The mean puncture time in the study group (17.74 ± 12.78 min) was significantly lower than that in the control group (58.32 ± 47.11 min; $P = 0.002$). The mean total fluoroscopy time was not significantly different between the study group (26.63 ± 12.84 min) and the control group (40.00 ± 23.44 min; $P = 0.083$). The mean total procedure time was significantly lower in the study group (79.74 ± 37.39 min) compared with the control group (121.70 ± 62.24 min; $P = 0.019$). The dose area product of the study group (220.60 ± 128.4 Gy. cm^2) was not significantly different from that of the control group (228.5 ± 137.3 Gy. cm^2 ; $P = 0.773$). There were no image guidance-related complications.

CONCLUSION

The use of preoperative simulation results and intraoperative image fusion to guide a portal vein puncture is feasible, safe, and effective when creating a TIPS. The method is cheap and may improve portal vein puncture, which may be valuable for hospitals lacking intravascular ultrasound and digital subtraction angiography (DSA) equipment equipped with a CT-angiography function.

KEYWORDS

Cirrhosis, portal hypertension, transjugular intrahepatic portosystemic shunt, three-dimensional reconstruction, simulation, image fusion guidance

A transjugular intrahepatic portosystemic shunt (TIPS) has been widely used to treat some complications associated with portal hypertension.¹⁻³ With the increasing experience of surgeons and the continuous development of imaging methods, the incidence of major complications related to a TIPS has decreased significantly in the past few decades.^{4,5} Puncturing the portal vein during a TIPS procedure is no longer a challenge for experienced doctors in large and medium-sized medical centers. However, in developing countries and underdeveloped regions, for medical institutions that are about to carry out a TIPS treatment (or the initial stages of the treatment), the puncture from the hepatic vein to the portal vein is still difficult, and doctors will face potentially fatal puncture-related complications.⁶ Therefore, effective and cheap intraoperative guidance methods may have a significant reference value for these interventional doctors in underdeveloped areas.

The preoperative simulation of a TIPS on a personal computer may provide some useful parameters. The fusion of the simulation results with intraoperative fluoroscopy may be helpful for the portal vein puncture. The purpose of this study is to evaluate the feasibility and effectiveness of carrying out a TIPS procedure under the guidance of fusion images in terms of the number of puncture attempts, puncture time, total procedure time, total fluoroscopy time, and dose area product, and to compare this with a conventional fluoroscopy group.

Methods

Patients

Nineteen consecutive patients who underwent a TIPS procedure because of compli-

cations resulting from cirrhosis-related portal hypertension were enrolled in the study between January 2021 and March 2022. The indication for TIPS creation was recurrent variceal bleeding refractory to endoscopic treatment and drug therapy. According to the Child–Pugh classification, chronic liver disease was categorized as class A in two patients, class B in 16 patients, and class C in one patient. Written informed consent was obtained from all patients before inclusion in this study. The study was approved by the Medical Ethics Committee of Nanchong Central Hospital [approval number: 2021, annual review (048), date: August 24, 2021]. As a control, the last 19 consecutive cases of the TIPS procedure were performed under the guidance of conventional fluoroscopy and analyzed retrospectively, including the number of puncture attempts, puncture time, total procedure time, total fluoroscopy time, and dose area product.

Methods of preoperative simulation

The portal vein phase data of preoperative abdominal-enhanced computed tomography (CT) were imported into the Mimics 10.0 software (Materialise NV, Leuven, Belgium) in Digital Imaging and Communications in Medicine format. The three-dimensional (3D) reconstruction models of the bone, portal vein, hepatic vein, inferior vena cava, and liver were extracted by setting the threshold and combining the functions of “region growing” and “dynamic region growing”.

The above 3D models of the patient were saved in an STL format file and then imported into the 3D Studio Max 7.0 software (Autodesk, San Rafael, California, USA). In the system settings of the software, the modeling unit was set to mm. According to the dimensions in the Rosch–Uchida transjugu-

lar liver access set (RUPS-100; Cook Medical Inc., Bloomington, Indiana, USA) and the VIATORR (W.L. Gore & Associates, Inc., Flagstaff, Arizona, USA) stent instructions, the liver access set, and stent models were established, respectively, by using the functions of line drawing, bending, alignment, and lofting (Figure 1). The liver access set and stent models were incorporated into the patient's 3D model scene. The planning of the puncture path from the hepatic vein to the portal vein, the shaping angle of the puncture needle end, and the simulation operation of the position of the stent in the portal vein were carried out (Figures 2, 3). The liver access set and stent model could also be saved in STL format and imported into the Mimics software for puncture and stent position simulation (Figure 4).

Methods of intraoperative image fusion

In the 3D Studio Max software, the inferior vena cava, portal vein, bone, and liver top models were rendered and saved in the anterior and lateral positions, respectively (image resolution: 1200 × 1200). In the Photoshop software, a transparent background image was established (image resolution: 1200 × 1200). The above-rendered images were imported into the transparent background image to become different layers. In each layer, the blank part was selected and deleted, which formed a combined image containing each part (Figure 5). During the operation, the X-ray fluoroscopy image of the anteroposterior position operation area was collected and saved in BMP or JPG format and then imported into the combined image. The X-ray fluoroscopy layer was scaled, and the ribs and vertebral bodies were overlapped and aligned with the 3D reconstructed bone image. Then, the reconstructed liver top lay-

Main points

- Compared with traditional methods, using preoperative simulation results and intraoperative image fusion to guide portal vein puncture in transjugular intrahepatic portosystemic shunt creation is feasible, safe, and effective.
- The results show that the study group's puncture time and total procedure time were significantly lower than those in the control group guided by fluoroscopy. This method is cheap and may improve portal vein puncture.
- To assess the impact of respiratory movement and the introduction of a stiff puncture needle and sheath on the position and direction of the liver, further study is required to try to find new image-matching reference points.



Figure 1. This image shows three-dimensional (3D) models of the Rups-100 liver access set and the VIATORR stents. In the 3D Max software, the angle of the front end of the liver access set and the bending direction of the VIATORR stent can be adjusted according to the simulated path of the portal vein puncture.

er was associated with the portal vein layer and moved together so that the upper edge of the liver top overlapped with the liver top position of the X-ray fluoroscopy image to form a fusion image (Figure 6); this was the output to the reference display screen in the operation room.

TIPS procedure

The 19 TIPS procedures were performed by the angiography system (Artis Zeego, Siemens Healthcare) and a team of five interventional radiologists (two of whom have more than three years of experience with a total of 76 TIPS procedures). Percutaneous access was achieved by puncture of the right internal jugular vein. A 10-French introducer was inserted, and the operator catheterized the hepatic vein. After introducing the liver access set into the hepatic vein, the intrahepatic puncture was performed according to the fusion image on the reference display screen. The lateral position of the C-arm was adjusted if necessary, and the lateral position fusion image was established. The initial angle of the puncture needle was not shaped in all cases during the first puncture. When the puncture was not successful after three attempts, it was considered that the bending angle of the puncture needle was not appropriate. At this point, the puncture needle was shaped according to the simulated angle. Once access to the intrahepatic portal branches was confirmed, a portogram was acquired, and the portal pressure gradient was measured. In each patient, the parenchymal tract was initially dilated using an 8 mm-diameter angioplasty balloon. A VIA-TORR stent was deployed to cover the entire length of the shunt up to the junction of the hepatic vein and the inferior vena cava. A final portal venogram was acquired, and the portal pressure gradient was measured again after the TIPS procedure.

Analysis methods and definitions

The time required for modeling and simulating puncture and stent release and acquisition of X-ray fluoroscopy images to the completion of the image fusion in each patient were recorded. For each procedure, parameters such as technical success, the number of needle passes, radiographic fluoroscopy time, total procedure time, radiation exposure, and procedural complications were recorded for data analysis. The system automatically recorded the total fluoroscopy time and the dose area product relating to the whole procedure. During the TIPS procedure, digital subtraction angiography (DSA) technicians usually

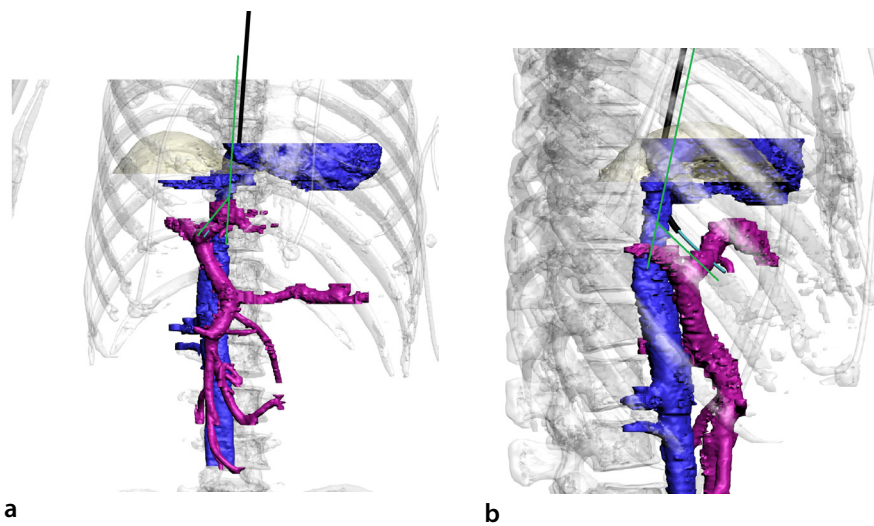


Figure 2. (a, b) The left portal vein puncture was simulated in the three-dimensional Max software. In the rendered anteroposterior position, the image shows the spatial relationship between the puncture needle and the portal vein and the angle of the puncture needle pointing to the portal vein (a). The rendered lateral position image shows the bending angle of the front end of the puncture needle (b). The yellow part of the picture is the liver dome. The additional green line indicates an angle that can be measured.

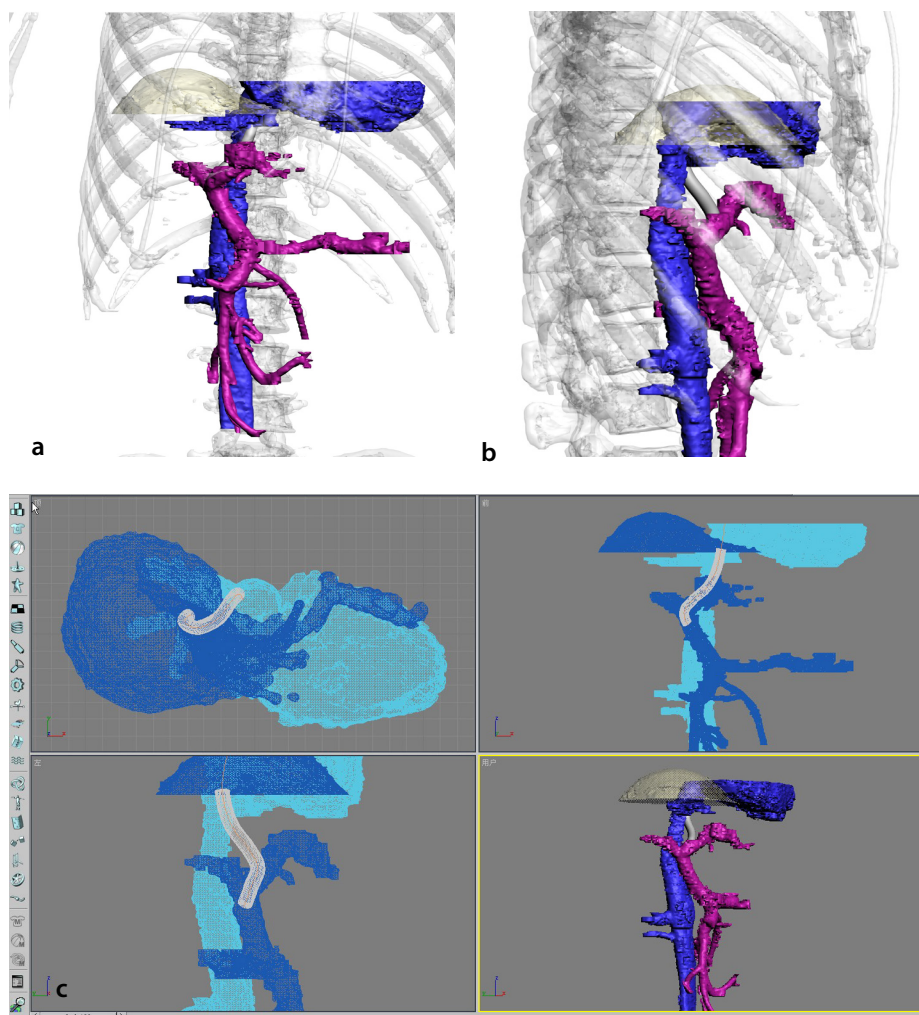


Figure 3. (a-c) The stent release simulation was performed in the three-dimensional (3D) Max software. The 3D rendering results show the relationship between the stent and the blood vessel in the anteroposterior position (a) and the lateral position (b), respectively. The morphology of the intravascular stent is clearly displayed on the 3D Max software interface (c).

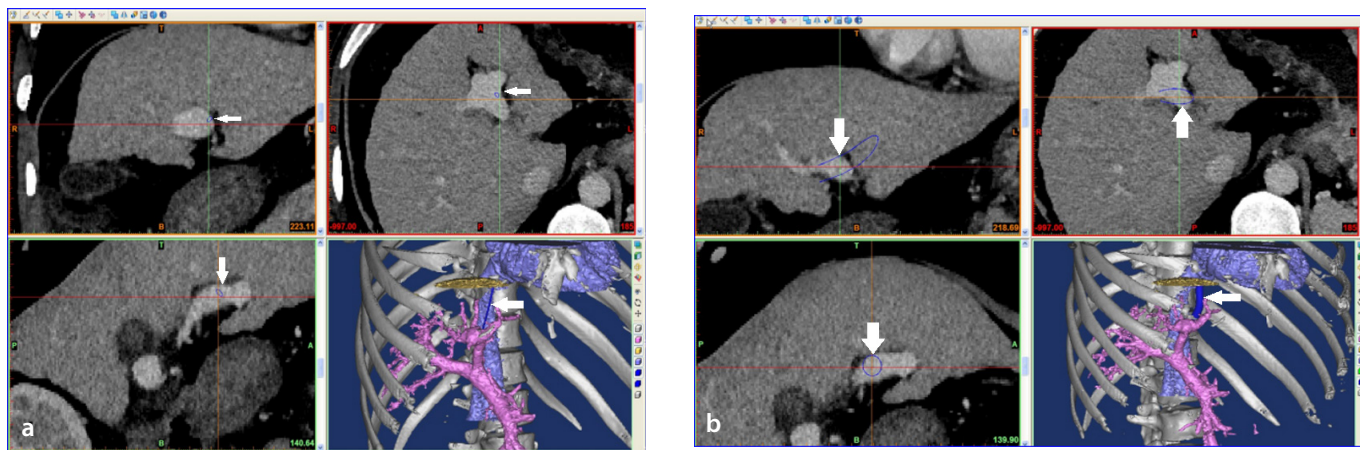


Figure 4. (a, b) Simulated portal vein puncture and stent release were performed in the Mimics software. The software can automatically display the path of the puncture needle (arrows) (a) and the path of the stent (arrows) (b) in transverse, coronal, and sagittal positions.

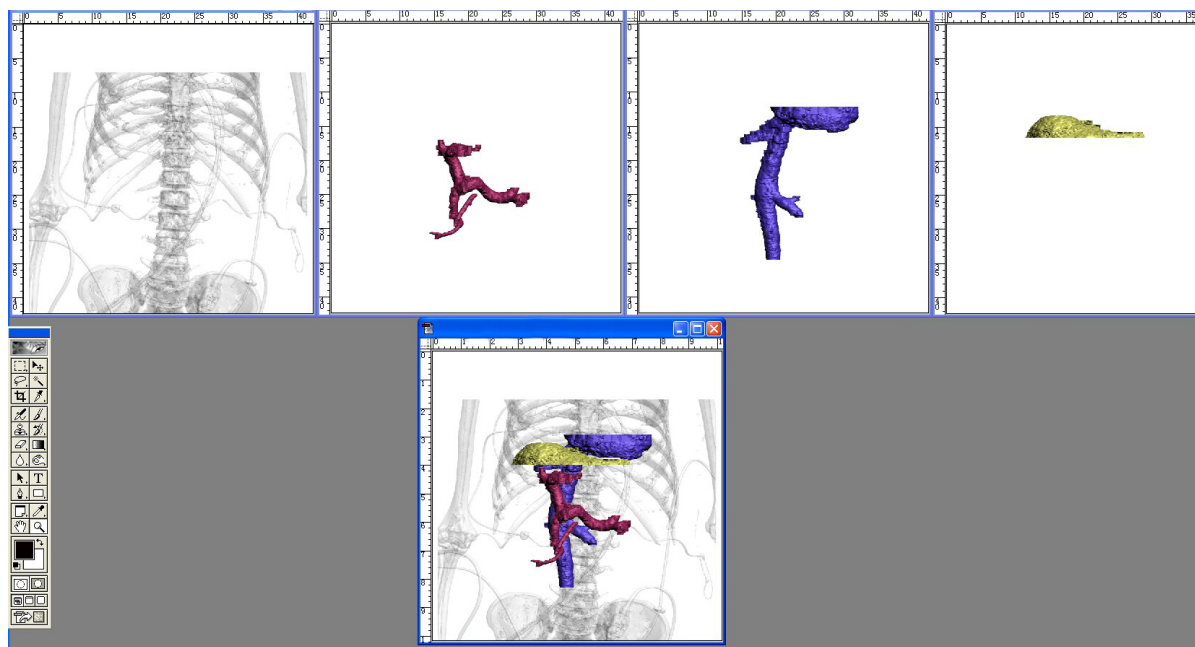


Figure 5. In the Photoshop software, the four images in the upper row show the models of the inferior vena cava, portal vein, bone, and the top of the liver diaphragm rendered in three-dimensional Max software. The following figure shows the combined image after merging the layers of four images.

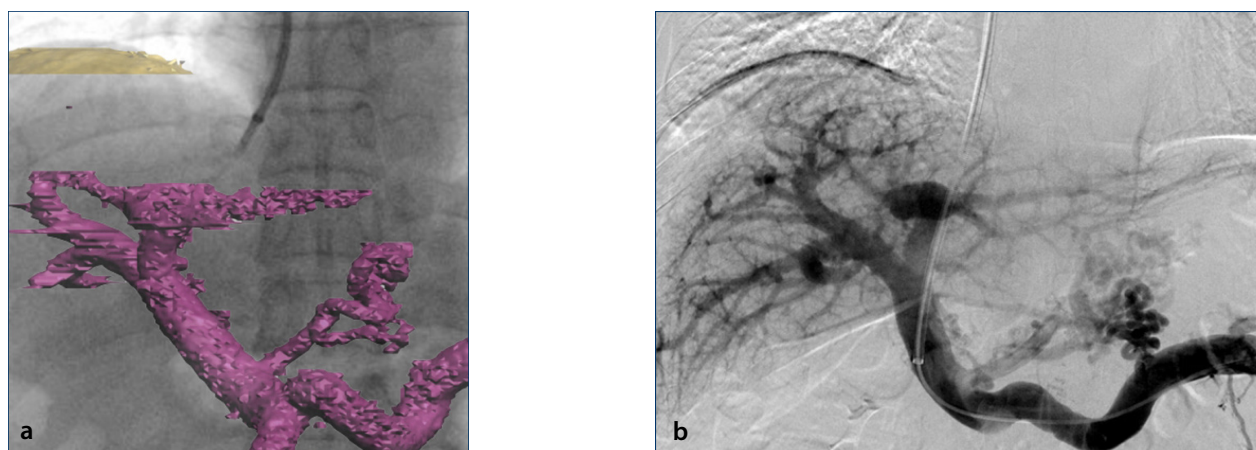


Figure 6. (a, b) A 44-year-old man with liver cirrhosis undergoing intraoperative image fusion guidance assisted transjugular intrahepatic portosystemic shunt to prevent variceal rebleeding. This is an example of relatively good registration accuracy. The yellow part of the picture is the three-dimensional (3D) reconstructed liver dome, which is registered and fused with the diaphragmatic surface of the liver in the intraoperative fluoroscopy (a). The left branch of the portal vein was successfully punctured after one puncture. Portal venography showed that the position and shape of the main portal vein and its left and right branches were consistent with the preoperative 3D reconstruction model (b).

saved the fluoroscopy images for each puncture operation. The portal vein puncture time of the patients in the control group was reconstructed as the time interval from the first image to the recorded image showing successful portal vein puncture. The total procedure time in the control group was reconstructed as the interval from the first to the last documented image. The number of puncture attempts was reconstructed as the number of puncture images saved at different time points before entering the portal vein. This possibly underestimated counting method did not exaggerate the significant difference between the two groups.

Technical success was defined as the successful creation of a shunt between the hepatic vein and the intrahepatic branch of the portal vein. The number of needle passes refers to the number of attempted punctures of the portal vein. Radiographic fluoroscopy time was defined as the period during which the X-ray fluoroscopy was used to guide the whole TIPS procedure. The total radiation dose associated with the whole procedure was automatically recorded by the system. Procedural complications were recorded during hospitalization, including intraabdominal hemorrhage, hepatic artery injury, hemobilia, and stent malposition.

Statistical analysis

All statistical analyses were performed with commercially available software (SPSS 22.0, IBM Corp., Armonk, New York, USA). Graphics were created with GraphPad Prism v.5 (GraphPad Software, San Diego, California, USA). Median and interquartile ranges were given for categorical data. Continuous variables were shown with mean and standard deviation. The Wilcoxon signed-rank test for non-normal distributed data was applied to assess the level of significance. Comparisons between intraoperative image fusion guidance and conventional fluoroscopy guidance were analyzed using the two-sided Mann–Whitney U test for continuous variables and Fisher’s exact test for categorical variables. A *P* value lower than 0.005 was accepted as a significant difference.

Results

All software operations of preoperative simulation and image fusion were completed by an interventional radiologist in the study group. Nineteen patients with a mean age of 49.74 ± 10.93 years (range: 26–66 years) were included in the study. There were 17 men

(89.5%) and two women (10.5%). The establishment of the virtual RUPS-100 liver access set and the VIATORR stent models was completed in the preliminary study, which took about 180 minutes. In this study, the model only needed to be copied and imported into the simulation scene. The average time of preoperative simulation in 19 patients was about 61.26 ± 6.98 (range: 50–75 minutes) minutes (Table 1), including the establishment of an individualized patient model, the simulation of the portal vein puncture, and the stent release in the 3D Studio Max and Mimics software. The operation convenience and the 3D display effect of the simulation process in the 3D Studio Max software were better than those in the Mimics software; the Mimics software had the advantage of observing the relationship between the puncture needle path and the artery and bile duct. Table 2 shows the patients’ data in the control group.

Using the time of the disinfection and the laying of the surgical towel, the rendered images of each reconstructed part of the model were imported into the Photoshop software, and the merged images with different layers were established. When the operator had

Table 1. Data pertaining to preoperative simulation, intraoperative image fusion, and the transjugular intrahepatic portosystemic shunt procedure

Patient no.	Preoperative simulation time (min)	Intraoperative image fusion time (min)	Puncture time (min)	Overall procedure time (min)	No. of needle passes	Radiographic fluoroscopy time (min)		Radiation dose (Gy·cm ²) of whole procedure	No. of stents placed
						Portal vein entry	Whole procedure		
1	57	8	22	90	4	7	30	466.79	1
2	65	6	19	99	2	6	33	151.91	1
3	56	5	34	99	4	11	33	87.99	1
4	64	7	6	42	1	2	14	95.47	1
5	72	8	12	108	2	4	36	391.43	1
6	52	7	36	147	3	12	49	408.53	1
7	68	6	4	30	1	2	10	100.79	1
8	58	7	21	48	4	7	16	106.97	1
9	75	6	5	75	1	2	25	108.71	1
10	64	5	36	162	3	12	54	387.30	1
11	60	5	13	87	3	4	29	300.65	1
12	55	5	4	75	1	2	25	182.65	1
13	63	6	47	87	4	11	29	340.17	1
14	57	5	6	22	1	2	5	126.00	1
15	50	5	16	42	3	5	20	77.84	1
16	72	8	14	116	3	6	42	275.60	2
17	63	6	17	69	3	5	21	269.91	1
18	58	5	22	65	3	4	21	176.33	1
19	55	5	3	52	1	1	14	136.55	1

completed the internal jugular vein puncture, inserted the guide wire and catheter, and introduced the stiff puncture needle and sheath in the hepatic vein for X-ray fluoroscopy, the time required for the image fusion was calculated. The total time was about 6.05 ± 1.13 minutes (range: 5–8 minutes), including saving the X-ray fluoroscopy image, copying the image from the workstation, and importing it into the Photoshop software to form a fusion image.

Technical success in TIPS creation was achieved in all 19 patients in the study group, of which 17 patients received variceal embolization at the same time. Each patient had a 3D reconstruction image of the portal venous system overlaid on the X-ray fluoroscopy to form a fusion image. One case used the preoperative simulated puncture needle as the registration reference point for image fusion to guide the portal vein puncture (Figure 7). Under the guidance of the intraoperative fusion map on the reference screen, the interventional radiologist adjusted the angle pointing to the left or right portal vein branch for punc-

ture and successfully performed the operation on 15 patients without adjusting the bending angle of the front end of the liver access set. Four patients underwent portal vein puncture three times according to the original bending angle of the front end of the liver access set, but all were unsuccessful. Then, according to the simulation results of the bending angle of the liver access set, the bending angle of the puncture needle was increased, and all four cases were successful at the fourth puncture. Intraoperative puncture of the bifurcation of the left and right branches of the portal vein was achieved in two cases, the left branch in 10 cases, and the right branch in seven cases (preoperative simulated puncture of the left branch in 12 cases and the right branch in seven cases). The overall coincidence rate was 89.47% (17/19).

The median number of puncture attempts was not significantly different between the study group ($n = 3$) and the control group ($n = 3$; $P = 0.175$; Figure 8). The mean puncture time in the study group (17.74 ± 12.78 min) was significantly low-

er than that in the control group (58.32 ± 47.11 min; $P = 0.002$; Figure 9). The mean total fluoroscopy time was not significantly different between the study group (26.63 ± 12.84 min) and the control group (40.00 ± 23.44 min; $P = 0.083$; Figure 10). The mean total procedure time was significantly lower in the study group (79.74 ± 37.39 min) compared to the control group (121.70 ± 62.24 min; $P = 0.019$; Figure 11). The dose area product of the study group (220.60 ± 128.4 Gy. cm^2) was not significantly different from that of the control group (228.5 ± 137.3 Gy. cm^2 ; $P = 0.773$; Figure 12). For details, refer to Table 2.

Two patients were punctured at the bifurcation of the left and right branches of the portal vein. The method of releasing the VIATORR stent first and then expanding the balloon in the stent was implemented. No contrast agent extravasation was observed on the portal vein angiography. The condition of these patients remained hemodynamically stable without transfusion. No other major complications or in-hospital deaths were observed in the present study.

Table 2. Data relating to the study group (image fusion guided transjugular intrahepatic portosystemic shunt) and the control group (conventional transjugular intrahepatic portosystemic shunt)

	Study group (n=19)	Control group (n=19)	P value
Gender (male/female)	17/2	12/7	0.124
Age (years)	49.74 ± 10.93	53.58 ± 8.43	0.447
Child–Pugh grade			0.539
A	2	4	
B	16	13	
C	1	2	
Ascites [n (%)]	15 (78.95%)	12 (63.16%)	0.476
Portal vein thrombosis [n (%)]	3 (15.79%)	1 (5.26%)	0.604
Location of portal vein puncture			0.342
Left branch of portal vein	10	15	
Right branch of portal vein	7	3	
Bifurcation of portal vein	2	1	
Variceal embolization [n (%)]	17 (89.47%)	14 (73.68%)	0.405
Assist other guidance methods [n (%)]	0	4 (21.05%)*	0.105
Number of puncture attempts	2.47 ± 1.17	3.74 ± 2.40	0.175
Puncture time (min)	17.74 ± 12.78	58.32 ± 47.11	0.002
Total fluoroscopy time (min)	26.63 ± 12.84	40.00 ± 23.44	0.083
Total procedure time (min)	79.74 ± 37.39	121.70 ± 62.24	0.019
Dose area product (Gy $\times$ cm^2)	220.60 ± 128.4	228.5 ± 137.3	0.773

*Among the four cases requiring other guidance methods, two cases were assisted with indirect portal vein angiography, and two cases were assisted with percutaneous transhepatic portal vein angiography.

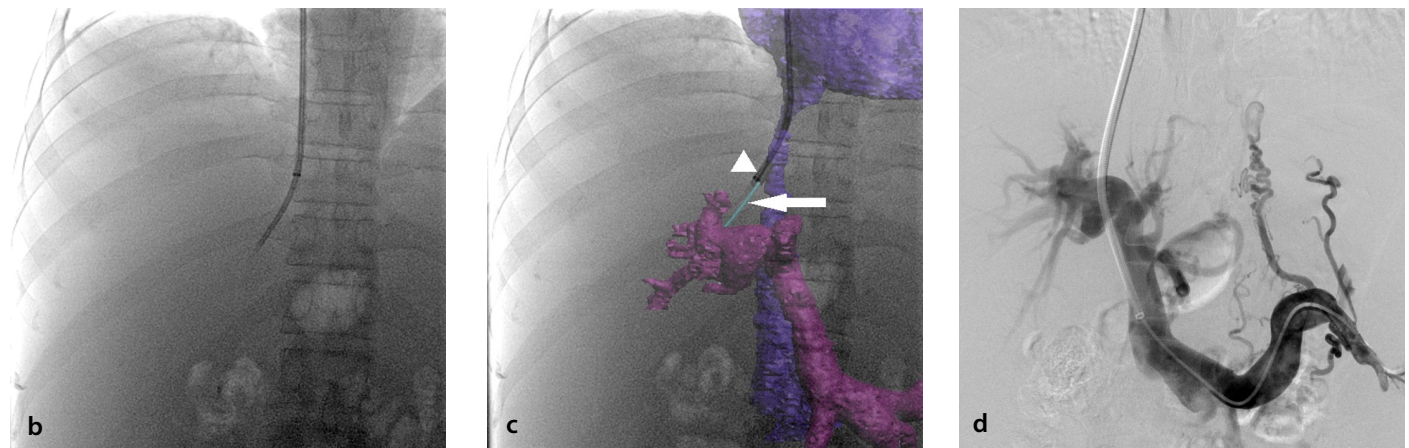
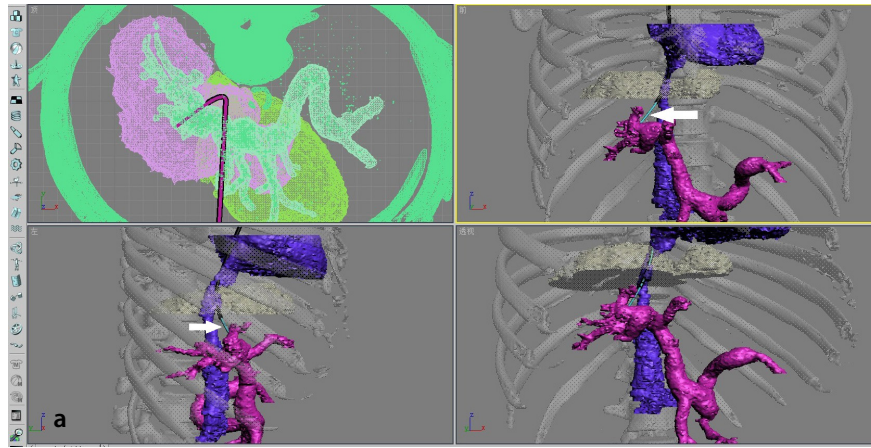


Figure 7. (a-d) An example of better registration accuracy. A 50-year-old man with liver cirrhosis undergoing intraoperative image fusion guidance assisted transjugular intrahepatic portosystemic shunt to prevent variceal rebleeding. The preoperative simulation provided the path for the puncture needle (arrow) to enter the portal vein (a). However, the intraoperative fluoroscopy found that the ascites increased significantly, and the liver diaphragmatic surface originally planned for registration was not clearly displayed (b). Therefore, the simulated puncture needle (arrow) was used as the registration reference point for image fusion with the hard puncture needle (arrowhead) in the fluoroscopy (c). After two punctures, the right branch of the portal vein was successfully punctured. The portogram showed that the position and shape of the main portal vein and its left and right branches were consistent with the preoperative three-dimensional reconstruction model (d).

Discussion

Numerous strategies have been proposed to localize the portal vein during a TIPS procedure, such as the placement of a coil, wire, or snare in or near the portal vein as a target, using a CO₂ wedged hepatic vein portography, and utilizing an image fusion of a preoperative CT and an intraoperative cone beam CT (CBCT) image, including 3D ultrasound (US) and intravascular ultrasound (IVUS).⁷⁻²² The use of IVUS in these strategies seems to be more effective than any other cross-sectional imaging procedure. Most punctures only need to be made once or twice, but they involve special tools, professional skills, and expensive costs,^{20,21} which may be difficult to procure in developing countries and underdeveloped areas. The new generation of fluoroscopy suites that have CT-angiography capability allows a fused image to move along with a fluoroscopy panel detector, creating a live image as obliquity changes, which avoids any preoperative planning and

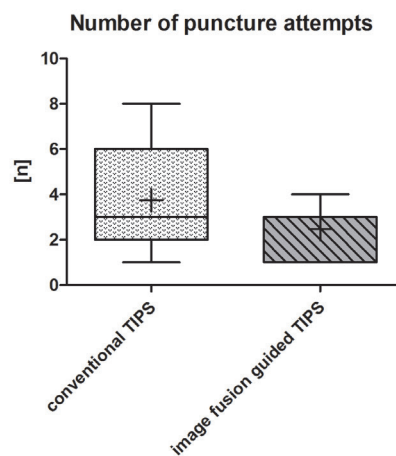


Figure 8. Number of puncture attempts. The median number of puncture attempts was not significantly different between the group of image fusion-guided transjugular intrahepatic portosystemic shunt (TIPS) ($n = 3$) and the group of conventional TIPS ($n = 3$; $P = 0.175$).

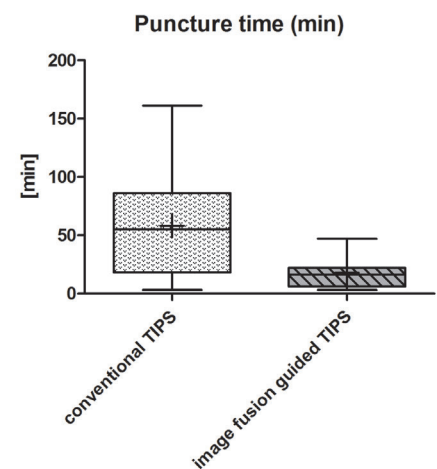


Figure 9. Puncture time. The mean puncture time was significantly different between the group of image fusion-guided transjugular intrahepatic portosystemic shunt (TIPS) (17.74 ± 12.78 min) and the group of conventional TIPS (58.32 ± 47.11 min; $P = 0.002$).

3D image production.^{11,12} The promotion of this excellent method is also limited by the need for expensive equipment in underdeveloped areas. Transabdominal US guidance has proved to be an effective, relatively simple, and non-invasive portal vein puncture guidance method. Its disadvantages include the need for additional equipment and a second clinician familiar with the TIPS procedure and US, leading to additional staff radiation exposure, increased costs, and limitations in personnel planning.^{18,22} Other guidance methods also have some shortcomings, such as indirect portal venography, CO₂ wedge hepatic vein portal venography, and percutaneous liver puncture portal venography, which increase both the operation and material costs.^{7-9,11,19} Additionally, CBCT increases the X-ray irradiation dose¹³⁻¹⁵ and requires a special workstation and software.^{11-15,18} However, the guidance method designed in this study, which integrates preoperative simulation results and intraoperative images to guide portal vein puncture, tries to avoid the shortcomings of the above methods.

Image registration was the difficulty in the fusion of the preoperative CT 3D reconstruction and intraoperative fluoroscopy. In many studies, manual adjustment was performed according to the vertebral body and rib standards.^{11,12,14,15,17,18} Due to the significant difference in the body position and respiratory state between preoperative CT scanning and intraoperative fluoroscopy, it is difficult to register the position of the liver and portal vein with reference to the bone position in the image fusion, but the relative position of the top of the liver diaphragm and portal

vein changes little. In this study, the upper and lower positions of image registration are based on the position of the top of the liver diaphragm, with the left and right positions based on the rib boundary.

Compared with the control group guided by fluoroscopy, the portal vein puncture time and the total procedure time in the study group were significantly lower than those in the control group. In terms of the number of puncture attempts, although the study group avoided assisting other invasive guidance methods, and the quartile Q3 ($n = 3$) was lower than the control group ($n = 6$), there was no significant difference between the two groups. There was also no significant difference between the two groups in total fluoroscopy time and dose product, which may be related to the fact that some cases received variceal embolization at the same time—17 cases in the study group (89.47%) and 14 cases in the control group (73.68%).

Although the present study was designed to evaluate the feasibility and safety of intraoperative image fusion guidance with the use of preoperative simulation results, initial comparisons with the TIPS procedures performed with the use of other imaging modalities could be made regarding the number of needle passes, the radiographic fluoroscopy time, the time required for the whole procedure, and the dose area product. A previous study assessed the usefulness of 3D US during the creation of TIPS. The mean number of needle passes required for portal vein entry was 4.6 in the 3D US group and 10.4 in the conventional TIPS group.²³ Fewer median number of intrahepatic needle passes were

required in the IVUS-guided TIPS creation group compared with the conventional TIPS group (two passes compared with six passes).²⁴ The results of the present study showed that when intraoperative image fusion guidance based on preoperative simulation results was used, the mean number of needle passes could be reduced to 2.47.

The time required for radiographic fluoroscopy during a TIPS procedure ranged from 3.5 to 153.1 minutes, with a mean of 38.7 minutes noted for 135 cases from a previous study.²⁵ Kee et al.²⁶ showed that the mean fluoroscopy time was reduced to 22.3 minutes when hybrid guidance with fluoroscopy and magnetic resonance imaging was used. In comparison, the initial results of the present study showed that preoperative simulation results and intraoperative image fusion guidance required a mean fluoroscopy time of 5.5 minutes for portal vein entry. Seventeen of the 19 cases underwent variceal vein embolization at the same time, so the operation time and exposure dose should be more than in a conventional TIPS. The mean radiographic fluoroscopy time was 26.63 ± 12.84 minutes for the whole TIPS procedure, and the mean radiation dose associated with the entire procedure was 220.60 ± 128.4 Gy·cm². Compared with previous studies,^{11,12,14,15,17,18} this result is at a medium level. However, the above literature does not report the simultaneous implementation of variceal embolization. For details, refer to Table 3.

In two patients in whom it was planned to puncture the left portal vein, the bifurcation of the left and right portal vein was punctured. This may be related to the lower start-

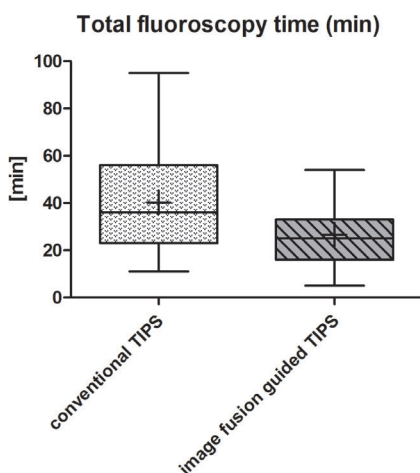


Figure 10. Total fluoroscopy time. The mean total fluoroscopy time was not significantly different between the group of image fusion-guided transjugular intrahepatic portosystemic shunt (TIPS) (26.63 ± 12.84 min) and the group of conventional TIPS (40.00 ± 23.44 min; $P = 0.083$).

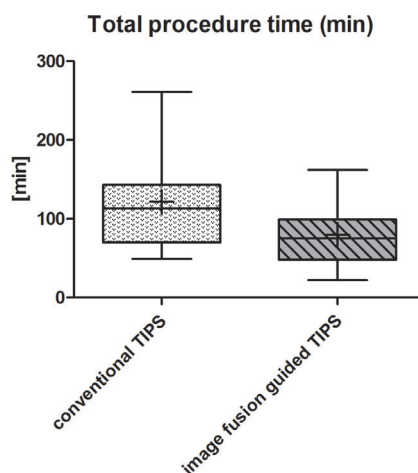


Figure 11. Total procedure time. The mean total procedure time was significantly different between the group of image fusion-guided transjugular intrahepatic portosystemic shunt (TIPS) (79.74 ± 37.39 min) and the group of conventional TIPS (121.70 ± 62.24 min; $P = 0.019$).

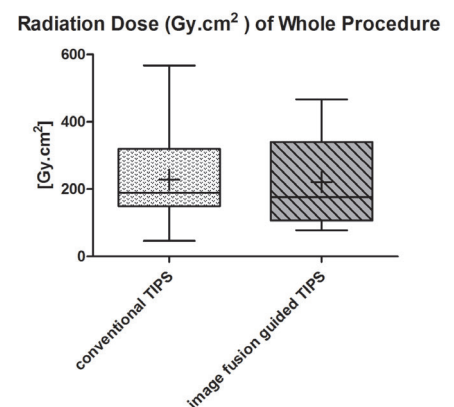


Figure 12. Radiation dose of the whole procedure. The mean radiation dose of the whole procedure was not significantly different between the group of image fusion-guided transjugular intrahepatic portosystemic shunt (TIPS) (220.60 ± 128.4 Gy·cm²) and the group of conventional TIPS (228.5 ± 137.3 Gy·cm²; $P = 0.773$).

Procedural characteristics	The present study	Meine et al. ¹² (2020)	Böning et al. ¹⁸ (2018)	Luo et al. ¹¹ (2018)	Luo et al. ¹⁵ (2017)	Rouabah et al. ¹⁷ (2016)	Ketelsen et al. ¹⁴ (2016)
Number of patients	19	27	21	15	20	18	12
Puncture time (min)	17 ± 12	14 ± 6	32 ± 45	n.a.	n.a.	17 ± 9	14 ± 8
Overall procedure time (min)	79±37*	64±29	115±52	60±13	n.a.	n.a.	66 ± 29
Fluoroscopy time (min)	26 ± 12*	21 ± 12	n.a.	14 ± 4	11 ± 2	n.a.	18 ± 9
Dose area product (Gy x cm ²)	220.60 ± 128.42*	107.48 ± 93.84	563.00 ± 289.00	152.11 ± 86.63	295.50 ± 66.60	258.53 ± 161.41	188.16 ± 121.18

*Among the 19 TIPS patients, 17 patients received variceal embolization at the same time. n.a., not applicable.

ing point of the hepatic vein puncture due to the patient's respiratory movement and the vertical distribution of the left portal vein. This series of cases shows that the VIATORR stent is valuable in avoiding extrahepatic portal vein puncture bleeding.

Using preoperative simulation results and intraoperative image fusion guidance as an auxiliary means of fluoroscopy to create a TIPS has the following characteristics. First, the tools of this study are three common commercial software packages, which are easy to obtain at low cost. No additional surgery and material consumption is needed, and there are no requirements for the configuration parameters of DSA equipment. Although it lacks accuracy compared with IVUS guidance and operability compared with DSA equipment accompanied with CT-angiography function, the method of this study may be more valuable for developing countries and underdeveloped regions.

Second, preoperative simulation can provide the angle of the puncture needle pointing to the portal vein in the anteroposterior position and its lateral bending angle. Because the vertical distance between the hepatic and portal vein is shorter or longer in a few patients, it is necessary to increase or reduce the original bending angle of the front end of the puncture needle before a successful puncture. The preoperative simulation can show this angle, which helps adjust the original bending angle of the front end of the puncture needle during the operation. In this group, four cases were successfully punctured after increasing the bending angle of the puncture needle.

Third, there are many choices of reference points for intraoperative image fusion, such as the rib edge, the vertebral body, the costophrenic angle, the right cardiac margin, and the top of the liver diaphragm. This study used the top of the liver diaphragm as the reference point for image fusion. The image fusion registration of eight cases was perfect

(Table 1), and the number of needle passes required to access the portal branch was one or two. In four cases, although the image fusion registration was also accurate in the anteroposterior position, the puncture was successful only after increasing the original bending angle of the puncture needle. The possible causes of registration deviation in the remaining seven cases include the translation and torsion of the liver after the use of a hard puncture needle and sheath, shortness of breath, and a significant increase or decrease of ascites.

Fourth, for elective surgery, preoperative simulation can simulate the puncture path of the right and middle hepatic veins, which depends on the patient's vascular anatomical characteristics and the needs of the surgeon. In emergency cases, there may not be time to complete the preoperative simulation process, and only the portal vein images in CT data can be extracted for intraoperative image fusion guidance. This process is similar to the method of Rouabah et al.¹⁷

The limitations of this study include the small sample size and the fact that the assessment involves a single institution. The retrospective analysis of the fluoroscopic-guided group resulted in a lack of data concerning the number of puncture attempts, puncture time, and total procedure time. The method of reconstructing the total procedure time and the number of puncture attempts of this group of patients introduced potential inaccuracies, which were accepted due to underestimation. In addition, the preoperative simulation results and the accuracy of intraoperative image fusion guidance technology were not quantitatively analyzed. To account for the impact of respiratory movement and the introduction of a stiff puncture needle and sheath on the position and direction of the liver, attempts to find new image-matching reference points should be made in the future. As with most guidance methods, in this study, it was difficult to replicate guid-

ance effects for patients with chronic portal vein obstruction, while IVUS guidance is effective for patients with completely occluded portal veins with or without cavernous transformation of the portal vein.²⁰

In conclusion, compared with traditional methods, using preoperative simulation results and intraoperative image fusion to guide portal vein puncture in TIPS creation is feasible, safe, and effective. The preoperative simulation method has potential value in TIPS training.

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

Supported by grant S20055 from the Sichuan Medical Association of China.

References

1. Corbett C, Mangat K, Olliff S, Tripathi D. The role of transjugular intrahepatic portosystemic stent-shunt (TIPSS) in the management of variceal hemorrhage. *Liver Int.* 2012;32(10):1493-1504. [\[CrossRef\]](#)
2. Boyer TD, Haskal ZJ. American Association for the Study of Liver Diseases. The role of transjugular intrahepatic portosystemic shunt (TIPS) in the management of portal hypertension. *Hepatology.* 2010;51(1):306. [\[CrossRef\]](#)
3. Colombato L. The role of transjugular intrahepatic portosystemic shunt (TIPS) in the management of portal hypertension. *J Clin Gastroenterol.* 2007;41(Suppl 3):S344-S351. [\[CrossRef\]](#)
4. Freedman AM, Sanyal AJ, Tisnado J, et al. Complications of transjugular intrahepatic portosystemic shunt: a comprehensive review. *Radiographics.* 1993;13(6):1185-1210. [\[CrossRef\]](#)
5. Coldwell DM, Ring EJ, Rees CR, et al. Multicenter investigation of the role of transjugular intrahepatic portosystemic shunt in management of portal hypertension. *Radiology.* 1995;196(2):335-340. [\[CrossRef\]](#)

6. Gaba RC, Khiatani VL, Knuttinen MG, et al. Comprehensive review of TIPS technical complications and how to avoid them. *AJR Am J Roentgenol*. 2011;196(3):675-685. [\[CrossRef\]](#)
7. Theuerkauf I, Strunk H, Brensing KA, Schild HH, Pfeifer U. Infarction and laceration of liver parenchyma caused by wedged CO(2) venography before tips insertion. *Cardiovasc Intervent Radiol*. 2001;24(1):64-67. [\[CrossRef\]](#)
8. Sheppard DG, Moss J, Miller M. Imaging of the portal vein during transjugular intrahepatic portosystemic shunt procedures: a comparison of carbon dioxide and iodinated contrast. *Clin Radiol*. 1998;53(6):448-450. [\[CrossRef\]](#)
9. Roizental M, Kane RA, Takahashi J, et al. Portal vein: US-guided localization prior to transjugular intrahepatic portosystemic shunt placement. *Radiology*. 1995;196(3):868-870. [\[CrossRef\]](#)
10. Solomon SB, Magee C, Acker DE, Venbrux AC. TIPS placement in swine, guided by electromagnetic real-time needle tip localization displayed on previously acquired 3-D CT. *Cardiovasc Intervent Radiol*. 1999;22(5):411-414. [\[CrossRef\]](#)
11. Luo X, Wang X, Yu J, et al. Transjugular intrahepatic portosystemic shunt creation: three-dimensional roadmap versus CO2 wedged hepatic venography. *Eur Radiol*. 2018;28(8):3215-3220. [\[CrossRef\]](#)
12. Meine TC, Dewald CLA, Becker LS, et al. Transjugular intrahepatic portosystemic shunt placement: portal vein puncture guided by 3D/2D image registration of contrast-enhanced multi-detector computed tomography and fluoroscopy. *Abdom Radiol (NY)*. 2020;45(11):3934-3943. [\[CrossRef\]](#)
13. Tacher V, Petit A, Derbel H, et al. Three-dimensional image fusion guidance for transjugular intrahepatic portosystemic shunt placement. *Cardiovasc Intervent Radiol*. 2017;40(11):1732-1739. [\[CrossRef\]](#)
14. Ketelsen D, Groezinger G, Maurer M, et al. Three-dimensional C-arm CT-guided transjugular intrahepatic portosystemic shunt placement: Feasibility, technical success and procedural time. *Eur Radiol*. 2016;26(12):4277-4283. [\[CrossRef\]](#)
15. Luo X, Wang X, Zhao Y, et al. Real-Time 3D CT image guidance for transjugular intrahepatic portosystemic shunt creation using preoperative CT: a prospective feasibility study of 20 patients. *AJR Am J Roentgenol*. 2017;208(1):W11-W16. [\[CrossRef\]](#)
16. Abi-Jaoudeh N, Kruecker J, Kadoury S, et al. Multimodality image fusion-guided procedures: technique, accuracy, and applications. *Cardiovasc Intervent Radiol*. 2012;35(5):986-998. [\[CrossRef\]](#)
17. Rouabah K, Varoquaux A, Caporossi JM, et al. Image fusion-guided portal vein puncture during transjugular intrahepatic portosystemic shunt placement. *Diagn Interv Imaging*. 2016;97(11):1095-1102. [\[CrossRef\]](#)
18. Böning G, Lüdemann WM, Chapiro J, et al. Clinical experience with real-time 3-D guidance based on C-arm-acquired cone-beam CT (CBCT) in transjugular intrahepatic portosystemic stent shunt (TIPSS) placement. *Cardiovasc Intervent Radiol*. 2018;41(7):1035-1042. [\[CrossRef\]](#)
19. Tsao J, Luo X, Ye L, Li X. Three-dimensional path planning software-assisted transjugular intrahepatic portosystemic shunt: a technical modification. *Cardiovasc Intervent Radiol*. 2015;38(3):742-746. [\[CrossRef\]](#)
20. Farsad K, Fuss C, Kolbeck KJ, et al. Transjugular intrahepatic portosystemic shunt creation using intravascular ultrasound guidance. *J Vasc Interv Radiol*. 2012;23(12):1594-1602. [\[CrossRef\]](#)
21. Petersen B. Intravascular ultrasound-guided direct intrahepatic portacaval shunt: description of technique and technical refinements. *J Vasc Interv Radiol*. 2003;14(1):21-32. [\[CrossRef\]](#)
22. Tavare AN, Wigham A, Hadjivassiliou A, et al. Use of transabdominal ultrasound-guided transjugular portal vein puncture on radiation dose in transjugular intrahepatic portosystemic shunt formation. *Diagn Interv Radiol*. 2017;23(3):206-210. [\[CrossRef\]](#)
23. Rose SC, Pretorius DH, Nelson TR, et al. Adjunctive 3D US for achieving portal vein access during transjugular intrahepatic portosystemic shunt procedures. *J Vasc Interv Radiol*. 2000;11(6):611-621. [\[CrossRef\]](#)
24. Kao SD, Morshedi MM, Narsinh KH, et al. Intravascular ultrasound in the creation of transhepatic portosystemic shunts reduces needle passes, radiation dose, and procedure time: a retrospective study of a single-institution experience. *J Vasc Interv Radiol*. 2016;27(8):1148-1153. [\[CrossRef\]](#)
25. Miller DL, Balter S, Cole PE, et al. Radiation doses in interventional radiology procedures: the RAD-IR study: part I: overall measures of dose. *J Vasc Interv Radiol*. 2003;14(6):711-727. [\[CrossRef\]](#)
26. Kee ST, Ganguly A, Daniel BL, et al. MR-guided transjugular intrahepatic portosystemic shunt creation with use of a hybrid radiography/MR system. *J Vasc Interv Radiol*. 2005;16(2 Pt 1):227-234. [\[CrossRef\]](#)



Copyright©Author(s) - Available online at dirjournal.org.
Content of this journal is licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License.

Effect of lipiodol marking before CT-guided cryoablation on the outcome of sporadic renal cell carcinoma

Yasuhiro Ushijima
 Daisuke Okamoto
 Nobuhiro Fujita
 Keisuke Ishimatsu
 Noriaki Wada
 Seiichiro Takao
 Ryo Murayama
 Masahiro Itoyama
 Kousei Ishigami

Kyushu University, Graduate School of Medical
Sciences, Department of Clinical Radiology, Fukuoka,
Japan

PURPOSE

This retrospective study evaluates the impact of preoperative lipiodol marking on the outcomes of computed tomography (CT)-guided cryoablation for histologically diagnosed sporadic renal cell carcinoma (RCC).

METHODS

This study analyzed the data of 173 patients who underwent CT-guided cryoablation for histologically proven sporadic RCC at a single institution between April 2014 and December 2020. The local control rate (LCR), recurrence-free survival rate (RFSR), overall survival rate (OSR), changes in renal function, and complications in patients with (n = 85) and without (n = 88) preoperative lipiodol marking were compared.

RESULTS

The 5-year LCR and 5-year RFSR were significantly higher in patients with lipiodol marking (97.51% and 93.84%, respectively) than in those without (72.38% and 68.10%, respectively) (P value <0.01, log-rank test). There were no significant differences between the two groups regarding the 5-year OSR (97.50% vs. 86.82%) or the deterioration in chronic kidney disease stage (12.70% vs. 16.43%). Grade ≥ 3 complications occurred in patients with lipiodol marking (n = 2, retroperitoneal hematoma and cerebral infarction in 1 patient each) and without (n = 5; urinary fistula in 2, colonic perforation in 2, urinary infection in 1).

CONCLUSION

Lipiodol marking before CT-guided cryoablation for sporadic RCC is a feasible approach to improving local control and RFS while mitigating the decline in renal function. Additionally, it may help reduce complications.

KEYWORDS

Ablation, cryoablation, computed tomography, kidney, oncology, tumor ablation

Corresponding author: Yasuhiro Ushijima

E-mail: ushijima.yasuhiro.538@m.kyushu-u.ac.jp

Received 09 July 2023; revision requested 11 August 2023; last revision received 11 November 2023; accepted 08 December 2023.



Epub: 02.01.2024

Publication date: 05.03.2024

DOI: 10.4274/dir.2023.232381

Renal cell carcinoma (RCC) ranks among the top 10 most common cancers, displaying higher prevalence in men than in women and often peaking between the ages of 60 and 70 years.^{1,2} Advancements in diagnostic imaging modalities have led to a rising incidence of incidentally detected cases.^{3,4} Although surgical resection remains the primary treatment, image-guided ablation techniques such as cryoablation and radiofrequency ablation are emerging as viable alternative therapies.⁵⁻⁷

Cryoablation for RCC has gained prominence as a minimally invasive treatment in recent years. Studies indicate its efficacy in providing high local control and preserving renal function to levels comparable to surgical resection. It is particularly applicable to elderly patients or those with comorbidities or multiple lesions.⁸⁻¹²

Presently, image-guided procedures dominate cryoablation for RCC, utilizing imaging equipment for lesion targeting and therapeutic area monitoring. Computed tomography

(CT), magnetic resonance imaging (MRI), and ultrasonography serve as primary imaging modalities, with CT being the most practical. The treated area is visualized as a low-density “ice ball”.¹³⁻¹⁶

A key limitation of CT-guided cryoablation is its poor soft-tissue contrast, which often makes it difficult to identify the lesion and distinguish it from the normal kidney.¹⁷ This is particularly apparent in small-diameter or embedded lesions, which are common indications for cryoablation treatment. Although intravenously administered contrast improves visibility, the effect is transient and cannot be sustained throughout the time-consuming cryoablation procedure. To address these limitations, preoperative transarterial lipiodol marking has been developed as a preoperative (1 or 2 days) procedure to enhance lesion visibility.¹⁸⁻²¹ Although this technique has shown promising efficacy in small case series, its superiority compared with cryoablation without lipiodol marking remains inadequately established.

Hence, this study aims to evaluate the mid- to long-term outcomes of lipiodol marking performed before cryoablation for RCC, comparing cases with and without lipiodol marking. The analysis includes an assessment of its impact on renal function and the incidence of complications.

Methods

Participants

This study obtained approval from the institution's review board of Kyushu University Hospital and Medical Institutions (no: 21109-00), and the requirement for informed consent was waived. The indication for cryoablation for RCC was determined through discussions among radiologists and urologists while considering factors such as the patient's age, comorbidities, surgical history, and renal function. Cryoablation was especially considered for lesions that would be difficult to surgically resect. Patients were actively engaged in this decision-making

process and provided explicit consent for the procedure.

During the period from April 2014 to December 2020, 336 renal tumors in 280 patients received cryoablation. Patients with sporadic RCC were selected, excluding those with prior RCC or treatment history, multiple lesions, or hereditary diseases such as Von Hippel–Lindau syndrome. Subsequently, out of 221 eligible cases, 173 cases with a histological diagnosis of RCC were included in this study (Figure 1). Lipiodol marking was performed in 85 cases before cryoablation and was not performed in 88 cases. Table 1 summarizes the patients' demographic data and tumor characteristics, revealing no statistically significant differences between the two groups in any variables.

Transarterial lipiodol marking

As described in a previous article,²² transarterial lipiodol marking aimed to enhance tumor visibility before cryotherapy. Under local anesthesia, a 3- to 4-F sheath (Super Sheath, Medikit, Tokyo, Japan) was inserted through the femoral artery. Digital subtraction angiography of the renal arteries and branches was performed. Contrast media use was minimized in patients with impaired renal function and sometimes replaced with carbon dioxide for contrast enhancement. Feeding arteries were identified based on DSA findings and contrast-enhanced CT. Selective catheter insertion was performed

into the feeding arteries, and lipiodol (Guerbet, France) was injected along with a small amount of gelatin sponge for embolization (Serescue, Nippon Kayaku, Tokyo, Japan). Unenhanced CT imaging confirmed lipiodol deposits in the lesion before completing the procedure.

Cryoablation procedure

Cryoablation was performed as described in a previous article.²² Briefly, it was performed under local anesthesia using an interventional radiology-CT system (Aquillion One, Canon, Tokyo, Japan) and a cryoablation system (CryoHit, Galil Medical, Arden Hills, MN, USA). Cryoprobes (IceRod, IceSeed, Galil Medical) were inserted into the tumor, which was frozen in two 10-minute sessions. The ablation area extended with a 5-mm margin from the lesion. Needle biopsies were performed before or during cryoablation.

Follow-up

Post-cryoablation, the patients received inpatient care for several days, followed by outpatient consultations at 3 months post-procedure and every 6 months thereafter. These consultations aimed to assess complications and treatment efficacy, defining local recurrence as a residual or new lesion within or near the ablation area that visibly increased in size during the follow-up period. Further details on the follow-up protocol can be found in our previous article on cryoablation for RCC.²²

Main points

- Lipiodol marking prior to computed tomography-guided cryoablation for sporadic renal cell carcinoma (RCC) improves local control.
- Renal dysfunction caused by lipiodol marking is clinically acceptable.
- Lipiodol marking prior to cryoablation for RCC may reduce complications.

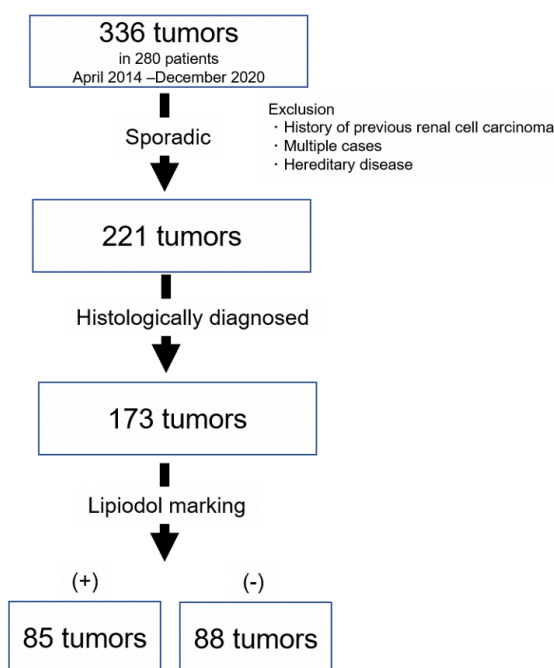


Figure 1. Selection of participants with cryoablation for sporadic renal cell carcinoma with histologic diagnosis.

Evaluations and statistical analysis

Demographic data in patients with sporadic RCCs treated by cryoablation were extracted and included age, sex, renal function [estimated glomerular filtration rate (eGFR), chronic kidney disease (CKD) grade], bilateral/unilateral kidney, presence of diabetes, anticoagulant/platelet medication, and history of other malignant diseases. The extracted characteristics of the tumor treated by cryoablation were size, location (right/left, exophytic/endophytic/hilum, upper/middle/lower, anterior/posterior/x), and histology (clear cell/papillary/chromophobe). The demographic data and tumor characteristics were compared between the groups with and without lipiodol marking using the Student's t-test for continuous variables (e.g., age, eGFR, tumor size).

The chi-square test was used for two categorical variables (e.g., sex, CKD grade, bilateral/unilateral kidney, presence of diabetes, anticoagulant/platelet medication, history

of other malignant diseases, right/left). Furthermore, the Fisher–Freeman–Halton test was used for more than three categorical variables (e.g., exophytic/endophytic/hilum, upper/middle/lower, anterior/posterior/x) (Table 1). Local control was defined as no recurrence in the treated region as identified by CT or MRI after cryoablation.

Recurrence-free survival was defined as being alive without local recurrence in the treated area or distant metastasis by CT or MRI. Overall survival was defined as being alive with or without local recurrence or distant metastasis. The date of cryoablation was used as the starting point for the observation period. The local control rate (LCR), relapse-free survival rate (RFSR), and overall survival rate (OSR) of cryoablation for sporadic RCCs with or without preoperative lipiodol marking were obtained using the Kaplan–Meier method, followed by a log-rank test. Multiple Cox regression analyses of the LCR were performed to assess the effects of lipiodol marking, age, sex, bilateral/unilateral

kidney, renal function, diabetes, anticoagulant/platelet medication, other malignant diseases, tumor size, tumor location, and histology.

For the evaluation of the change in renal function, a comparison of the Δ eGFR and downgrades in CKD stage between the two groups was analyzed using the Student's t-test and chi-square test, respectively. The common terminology criteria for adverse events (CTCAE), version 5, published by the US National Cancer Institute in 2017, was used to evaluate complications on a graded scale. The frequency of complications of CTCAE grade 3 or higher was also compared between the two groups using Fisher's exact test. For all statistical analyses, P values <0.05 were considered significant. All analyses were performed using JMP pro, version 15, software (SAS, Cary, NC, USA). Descriptive statistics are presented as mean $\pm$ standard deviation or median $\pm$ standard error, and frequencies are presented as percentages.

Table 1. Demographic data and tumor characteristics in all patients with sporadic renal cell carcinomas treated by cryoablation

	Lipiodol marking (+) (n = 85)	Lipiodol marking (-) (n = 88)	P value*
Demographic data			
Age, y (mean $\pm$ SD)	71.33 $\pm$ 11.22	71.20 $\pm$ 11.27	0.472
Sex, male/female (%)	60/25 (70.59/29.41)	68/20 (73.12/26.88)	0.316
Renal function			
eGFR mL/min/1.73 m ² (mean $\pm$ SD)	60.92 $\pm$ 23.07	56.91 $\pm$ 22.02	0.244
CKD grade, 2/3/4/5 (%)	49/27/5/4 (57.65/31.76/5.88/4.71)	44/35/6/3 (50/39.77/6.82/3.41)	0.686
Kidney, bilateral/unilateral (%)	84/1 (98.82/1.18)	81/7 (92/04/7.95)	0.064
Diabetes, yes/no (%)	19/66 (22.35/77.65)	17/71 (19.32/80.68)	0.623
Anticoagulant/platelet drug, yes/no (%)	23/62 (27.06/72.94)	29/59 (32.95/67.05)	0.400
Other malignant diseases, yes/no (%)	27/58 (31.76/68.23)	30/58 (34.01/65.91)	0.745
Tumor characteristics			
Tumor size, mm (mean $\pm$ SD)	24.27 $\pm$ 7.47	26.08 $\pm$ 7.92	0.124
Tumor location			
Right/left (%)	39/46 (45.88/54.12)	47/41 (53.41/46.59)	0.322
Exophytic/endophytic/hilum (%)	48/31/6 (56.47/36.47/7.06)	41/39/8 (46.59/44.32/9.09)	0.428
Upper/middle/lower (%)	26/45/14 (30.59/52.94/16.47)	26/41/21 (29.55/46.59/23.86)	0.464
Anterior/posterior/x** (%)	35/41/9 (41.18/48.24/10.59)	37/43/8 (42.04/48.86/9.09)	0.947
Histology, clear cell/papillary/chromophobe (%)	79/5/1 (92.94/5.88/1.18)	82/6/0 (93.18/6.82/0)	0.578

*Categorical data: chi-square test for two variables and Fisher–Freeman–Halton test for more than three variables, continuous variable: Student's t-test. **Unclassifiable as a polar lesion. eGFR, estimated glomerular filtration rate; SD, standard deviation; CKD, chronic kidney disease.

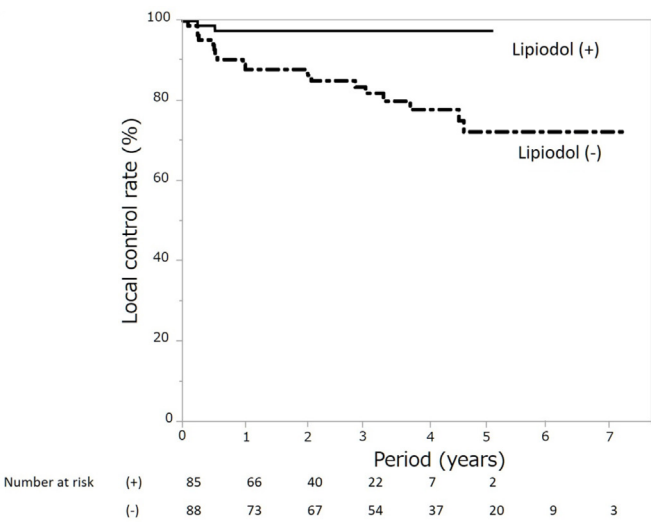
Results

Both the 3-year and 5-year LCRs were 97.51% among patients with preoperative lipiodol marking, compared with 82.00% and 72.38%, respectively, among those without. These differences were statistically significant ($P = 0.005$) (Figure 2). Local recurrence occurred in 20 cases, leading to additional cryoablation in 18 cases. Nephrectomy was performed in one case, and another received no further treatment due to advanced age and concurrent malignancy. The median follow-up duration for local tumor control was 2.01 ± 0.13 years with lipiodol marking and 3.12 ± 0.21 years without. Patients with lipiodol marking exhibited 3-year and 5-year RFSRs of 93.84%, whereas those without the marking had rates of 80.97% and 68.10%, respectively.

These differences were statistically significant ($P = 0.020$) (Figure 3). Four cases developed lung metastases. The 3-year and 5-year OSRs were both 97.50% for patients with lipiodol and 96.08% and 86.82%, respectively, for those without. However, these values did not constitute statistically significant differences (Figure 4). All deaths were attributed to diseases other than RCC. Multiple Cox regression analyses revealed significant effects of lipiodol marking ($P = 0.012$) and tumor size ($P = 0.045$) on the LCR (Table 2).

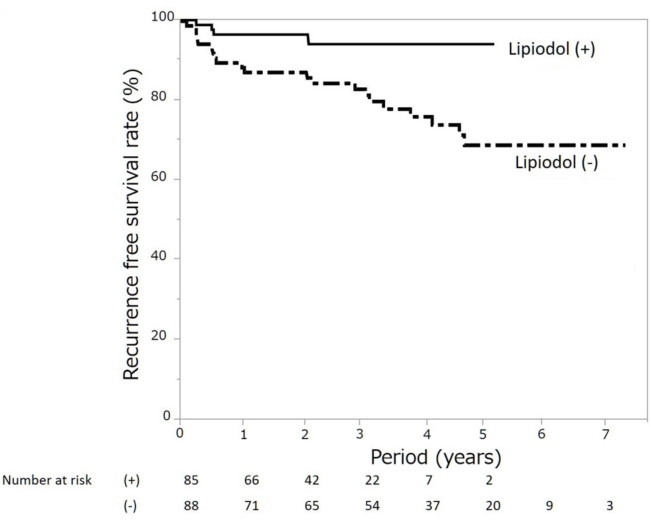
To evaluate renal function changes, $\Delta eGFR$ was calculated. Patients with preoperative lipiodol marking exhibited a mean $\Delta eGFR$ of 4.34 ± 8.08 (mL/min/1.73 m²), whereas those without had 3.48 ± 3.48 . The downgrading of CKD status was observed in 12.70% (8 of 63) of the patients with lipiodol marking and 16.43% (12 of 73) of those without, revealing no statistically significant differences between groups (Table 3).

Grade 3 or higher CTCAE complications occurred in only 2 cases (retroperitoneal hematoma in 1, cerebral infarction in 1) with lipiodol marking and 5 cases (urinary fistula in 2, colonic perforation in 2, urinary infection in 1) without. There was no statistically significant difference in the frequency of complications of CTCAE grade 3 or higher between the two groups (Table 4). The most common complications after cryoablation were fever, pain, and hematuria, almost all of which were grade 2 or lower.



	Period (years) median ± SE	LCR 3 years	LCR 5 years	P value
Lipiodol (+)	2.01 ± 0.13	97.51%	97.51%	0.005
Lipiodol (-)	3.12 ± 0.21	82.00%	72.38%	

Figure 2. The local control rate (LCR) of sporadic renal cell carcinomas treated by cryoablation with or without preoperative lipiodol marking. The LCR at 3 years was 97.51% with and 82.00% without preoperative lipiodol marking; at 5 years, those values were 97.51% and 72.38%, respectively. There was a statistically significant difference between them (log-rank test, $P = 0.005$). SE, standard error.



	Period (years) median ± SE	LCR 3 years	LCR 5 years	P value
Lipiodol (+)	2.01 ± 0.13	93.84%	93.84%	0.020
Lipiodol (-)	3.12 ± 0.21	80.97%	68.10%	

Figure 3. The recurrence-free survival rates (RFSRs) of all patients with sporadic renal cell carcinomas treated by cryoablation with or without preoperative lipiodol marking. The RFSRs at 3 years were 93.84% with and 80.97% without preoperative lipiodol marking; at 5 years, those values were 93.84% and 68.10%, respectively. There was a statistically significant difference between them (log-rank test, $P = 0.020$). SE, standard error; LCR, local control rate.

Discussion

This retrospective study compared cryoablation outcomes, renal function changes, and complications in patients with and without lipiodol marking to clarify the usefulness of lipiodol marking prior to cryoablation for sporadic RCC. Despite its retrospective nature, no significant differences were observed in demographic data or tumor characteristics between groups, establishing the study as a reliable comparison. The LCR was significantly higher in patients with lipiodol marking than in those without. Previous reports have also reported the usefulness of lipiodol marking and effective short-term local control.¹⁸⁻²¹ Moreover, there have been no reports of mid- to long-term treatment effects compared with groups without lipiodol marking, as in this study, and the results of this study further support the usefulness of lipiodol marking in the therapeutic effects of cryoablation.

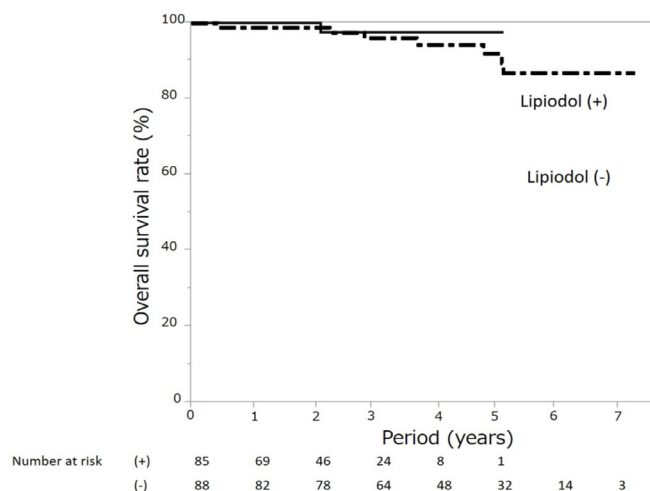
The first factor that improves local control and other treatment effects is improved lesion visibility with lipiodol marking (Figure 5). CT has lower soft-tissue resolution than MRI, and RCCs with small tumor diameters and embedded lesions can be difficult to recognize through CT fluoroscopy.¹⁹ The improved visibility of the lesion and more accurate puncture of the cryoprobe were obvious factors contributing to the improved outcome of the treatment.

In addition, the decrease in blood flow in the lesion may have been a factor affecting the treatment outcome. RCC is a tumor with abundant blood flow, and this heat sink is known to attenuate the freezing effect.²³ The small amount of embolic material in lipiodol marking is thought to induce a degree of ischemia in the lesion, and this may have reduced the heat sink effect and contributed to the therapeutic effect of cryoablation.

Transarterial lipiodol marking prior to cryoablation is associated with concerns regarding damage to the normal kidney and impaired renal function. In this study, there was no significant difference in change in renal function (eGFR) or CKD stage after cryoablation between the groups with and without lipiodol marking. Previous reports have also reported minimal deterioration in renal function when lipiodol marking was performed prior to cryoablation.²⁰ It is necessary to limit the use of contrast media in patients with severely impaired renal function and to be careful about the extensive injection of lipiodol into the normal renal

parenchyma. However, lipiodol marking performed before cryoablation seems acceptable from the perspective of preserving renal function.

Complications were similar to those in previous reports, with most being minor (CTCAE grade 2 or lower) in groups with and without lipiodol marking.^{8-10,18-21}



	Period (years) median ± SE	LCR 3 years	LCR 5 years	P value
Lipiodol (+)	2.01 ± 0.13	97.50%	97.50%	0.489
Lipiodol (-)	4.60 ± 0.18	96.08%	86.82%	

Figure 4. Overall survival rate (OSR) of all patients with sporadic renal cell carcinomas treated by cryoablation with or without preoperative lipiodol marking. The OSR at 3 years was 97.50% with and 96.08% without preoperative lipiodol marking; at 5 years, those values were 97.50% and 86.82%, respectively. There was no statistically significant difference between them (log-rank test, $P = 0.489$). SE, standard error; LCR, local control rate.

Table 2. Results of multiple Cox regression analysis of local control in patients with sporadic renal cell carcinomas treated by cryoablation

	HR (95% CI)	P value
Lipiodol marking (+/-)	0.145 (0.032, 0.657)	0.012
Age	1.385 (0.162, 11.866)	0.766
Sex	0.662 (0.209, 2.091)	0.482
Renal function		
eGFR	1.019 (0.969, 1.071)	0.469
CKD grade	2.118 (0.378, 12.105)	0.870
Kidney	4.024 (0.589, 27.496)	0.156
Diabetes	1.941 (0.541, 6.966)	0.309
Anticoagulant/platelet medication	0.456 (0.136, 1.525)	0.203
Other malignant disease	0.552 (0.189, 1.613)	0.277
Tumor size	1.077 (1.002, 1.158)	0.045
Tumor location		
Right/left	0.435 (0.162, 1.169)	0.100
Exophytic/endophytic/hilum*	1.232 (0.240, 6.309) 1.935 (0.419, 8.941)	0.802 0.398
Upper/middle/lower*	0.610 (0.091, 4.094) 1.675 (0.283, 9.928)	0.611 0.570
Anterior/posterior*/X	0.374 (0.115, 1.225) <0.001 (0, 0)	0.104 1.000
Histology	2.884 (0.302, 27.505)	0.357
Clear cell/papillary*/other	94.934 (0, 0)	1.000

*Reference category, P value of the Cox regression model: 0.022. HR, hazard ratio; eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease; CI, confidence interval.

Table 3. Change of renal function 1 year after cryoablation for sporadic renal cell carcinomas			
	Lipiodol marking (+)	Lipiodol marking (-)	P value*
Δ eGFR (mL/min/1.73 m ²) (SD)	Mean 4.34 ($\pm$ 8.08)	Mean 3.48 ($\pm$ 8.64)	0.547
Downgrade of CKD stage (n)	12.70% (8/63)	16.43% (12/73)	0.600

* Δ eGFR, Student's t-test, downgrade of CKD stage: chi-square test. eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease; SD, standard deviation.

Table 4. Complications of common terminology criteria for adverse events grade 3 or above following cryoablation for sporadic renal cell carcinomas		
Lipiodol (+) (n = 85)	Lipiodol (-) (n = 88)	P value*
Retroperitoneal hematoma: n = 1 Cerebral infarction: n = 1 n = 2 (2.35%)	Urinary fistula: n = 2 Urinary tract infection: n = 1 Colonic perforation: n = 2 n = 5 (5.68%)	0.444

*Fisher's exact test.

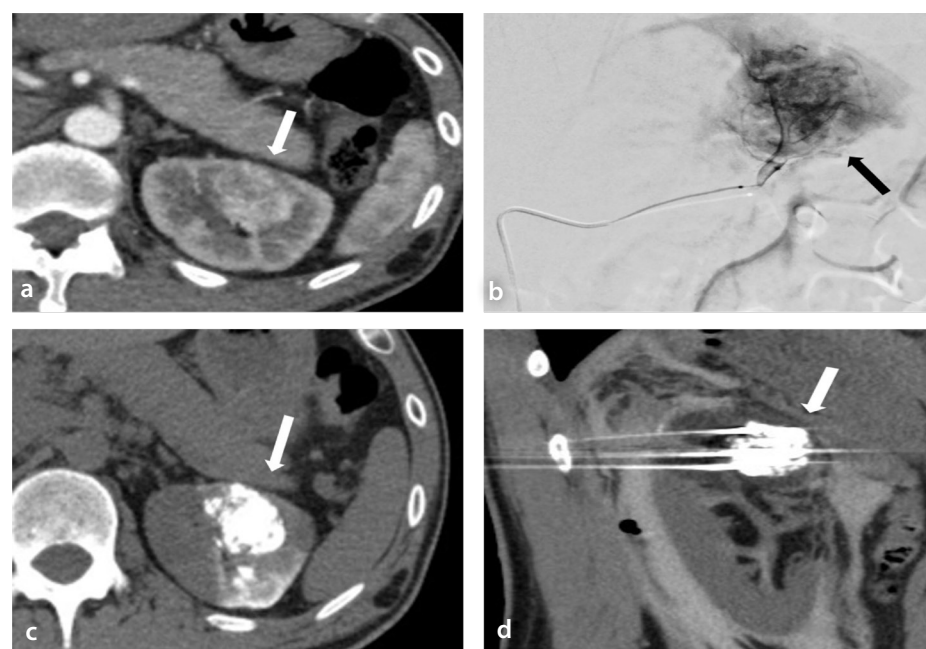


Figure 5. Cryoablation for a sporadic renal cell carcinoma of a 47-year-old man with preoperative lipiodol marking. (a) Contrast-enhanced CT showed an endophytic 3.3-cm mass at the ventral side of the left kidney (arrow). (b) Digital subtraction angiography from the upper branch of the left renal artery showed a hypervascular lesion representing the renal mass (arrow). (c) CT after transarterial lipiodol marking showed well-infused lipiodol in the renal mass (arrow). (d) CT-guided cryoablation was performed on the lipiodol-infused mass (arrow) using cryoprobes. After 4 years of follow-up, there was no recurrence, the Δ eGFR was 2, there was no decrease in CKD grade, and there were no grade 3 or higher complications. CT, computed tomography; eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease.

The two complications that were grade 3 or higher in the group with lipiodol marking were an iliopsoas hematoma in a patient with severely impaired coagulation and a perioperative cerebral infarction in a patient on anticoagulation and platelet medication.

These events could have occurred regardless of whether or not lipiodol marking was performed. The two cases of intestinal perforation among the grade 3 or higher complications in the group without lipiodol marking may have been caused by the intestine entering the cryoablation area because

the cryoprobes were inserted while the positional relationship between the lesion to be treated and the surrounding intestine was poorly visualized (Figure 6).

No intestinal perforation occurred in the group with lipiodol marking in this study, suggesting that lipiodol marking may reduce the risk of intestinal perforation. Although no statistically significant differences were found, our data suggest that improved lesion visibility may result in less frequent complications from cryoablation.

Transarterial lipiodol marking as a pre-treatment for cryoablation for RCC is widely used in Japan,¹⁸⁻²⁰ but it is not well recognized in other countries. Selective catheterization of the target vessel by a transarterial approach, as well as infusion and embolization, is not a difficult procedure for many interventional radiologists and does not require special training. It should be more widely recognized as a method to improve the quality of cryoablation for RCC.

The limitation of this study is that it is a single-center, retrospective study. The number of cases was not sufficiently large. Although there have been reports on the efficacy of cryoablation treatment in larger patient populations,^{8,11,12} there have been no reports on a larger group of participants in studies of lipiodol marking in cryoablation. Furthermore, cases with no tissue diagnosis on biopsy were excluded from this study. It was assumed that lesions that are difficult to diagnose by biopsy (small diameter, endophytic, hilum type, etc.) may not be easily treated by cryoablation, which may have affected the results of this treatment, the changes in renal function, and the occurrence of complications.

Finally, in the group with lipiodol marking, there was a concern that the lipiodol deposited in the lesions may interfere with the early detection of recurrent lesions by contrast-enhanced CT. In this study, the observation period for the group with lipiodol marking was shorter than that for the group without lipiodol marking, suggesting that further follow-up is needed.

Lipiodol marking prior to CT-guided cryoablation for sporadic RCC is a feasible approach to improve local control and RFS while lessening the decline in renal function, and it may be able to reduce complications.

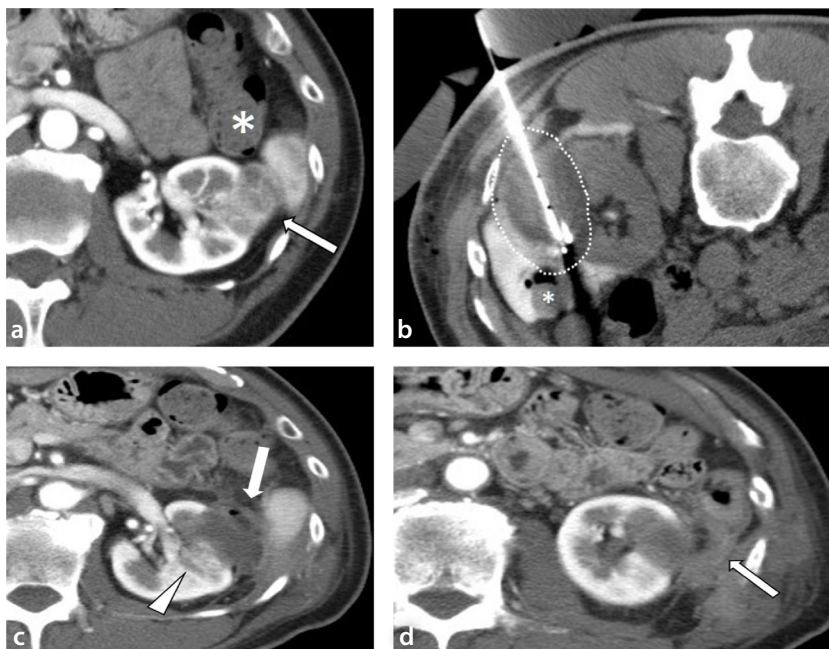


Figure 6. Cryoablation for a sporadic renal cell carcinoma of a 71-year-old man without preoperative lipiodol marking. (a) Contrast-enhanced CT showed an exophytic 2.9-cm mass at the anterolateral side of the left kidney (arrow). The mass and the descending colon (*) were in close contact. (b) CT-guided cryoablation was performed for the mass using cryoprobes with hydrodissection. However, part of the cryoablation area (dotted circle) extended over the colon (*). (c, d) Contrast-enhanced CT 1 month after cryoablation showed abscess formation (c, arrow) due to colon perforation (d, arrow), and part of the renal mass remained due to an insufficient cryoablation effect (c, arrowhead). Surgical resection was performed for the colon perforation, and additional cryoablation was performed for the residual lesions. CT, computed tomography.

Conflict of interest disclosure

The authors declared no conflicts of interest.

Funding

This work was supported by JSPS KAKEN-HI grant number: JP20K07999.

References

- Capitanio U, Montorsi S. Renal cancer. *Lancet*. 2016;387(10021):894-906. [\[CrossRef\]](#)
- Shuch B, Amin A, Armstrong AJ, et al. Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities from biologic complexity. *Eur Urol*. 2015;67(1):85-97. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Gill IS, Aron M, Gervais DA, Jewett MA. Clinical practice. Small renal mass. *N Engl J Med*. 2010;362:624-634. [\[CrossRef\]](#)
- Ljungberg B, Albiges L, Bedke J, et al. EAU Guidelines on Renal Cell Carcinoma. *European Association of Urology* 2022. [\[CrossRef\]](#)
- NCCN Clinical Practice Guidelines in Oncology: Kidney Cancer 2.2014 National Comprehensive Cancer Network. [\[CrossRef\]](#)
- Kanesvaran R, Porta C, Wong A, et al. Pan-Asian adapted ESMO Clinical Practice Guidelines for the diagnosis, treatment and follow-up of patients with renal cell carcinoma. *ESMO Open*. 2021;6(6):100304. [\[CrossRef\]](#)
- Thompson RH, Atwell T, Schmit G, et al. Comparison of partial nephrectomy and percutaneous ablation for cT1 renal masses. *Eur Urol*. 2015;67(2):252-259. [\[CrossRef\]](#)
- Uhlig J, Strauss A, Rücker G, et al. Partial nephrectomy versus ablative techniques for small renal masses: a systematic review and network meta-analysis. *Eur Radiol*. 2019;29(3):1293-1307. [\[CrossRef\]](#)
- Yoshimatsu R, Yamagami T, Nishimori M, et al. Time-dependent effects of cryoablation for renal tumor on overall and split renal function. *J Vasc Interv Radiol*. 2019;30(3):460-465. [\[CrossRef\]](#)
- Breen DJ, King AJ, Patel N, Lockyer R, Hayes M. Image-guided cryoablation for sporadic renal cell carcinoma: three- and 5-year outcomes in 220 patients with biopsy-proven renal cell carcinoma. *Radiology*. 2018;289(2):554-561. [\[CrossRef\]](#)
- Lim E, Kumar S, Seager M, et al. Outcomes of renal tumors treated by image-guided percutaneous cryoablation: immediate and 3- and 5-year outcomes at a regional center. *AJR Am J Roentgenol*. 2020;215(1):242-247. [\[CrossRef\]](#)
- Uppot RN, Silverman SG, Zagoria RJ, Tuncali K, Childs DD, Gervais DA. Imaging-guided percutaneous ablation of renal cell carcinoma: a primer of how we do it. *AJR Am J Roentgenol*. 2009;192(6):1558-1570. [\[CrossRef\]](#)
- Silverman SG, Tuncali K, Adams DF, Nawfel RD, Zou KH, Judy PF. CT fluoroscopy-guided abdominal interventions: techniques, results, and radiation exposure. *Radiology*. 1999;212(3):673-681. [\[CrossRef\]](#)
- Mogami T, Harada J, Kishimoto K, Sumida S. Percutaneous MR-guided cryoablation for malignancies, with a focus on renal cell carcinoma. *Int J Clin Oncol*. 2007;12(2):79-84. [\[CrossRef\]](#)
- Seager M, Kumar S, Lim E, Munneke G, Bandula S, Walkden M. Renal cryoablation - a practical guide for interventional radiologists. *Br J Radiol*. 2021;94(1118):20200854. [\[CrossRef\]](#)
- Fennessy FM, Tuncali K, Morrison PR, Tempny CM. MR imaging-guided interventions in the genitourinary tract: an evolving concept. *Radiol Clin North Am*. 2008;46(1):149-166. [\[CrossRef\]](#)
- Michimoto K, Shimizu K, Kameoka Y, Sadaoka S, Miki J, Kishimoto K. Transcatheter arterial embolization with a mixture of absolute ethanol and iodized oil for poorly visualized endophytic renal masses prior to CT-guided percutaneous cryoablation. *Cardiovasc Intervent Radiol*. 2016;39(11):1589-1594. [\[CrossRef\]](#)
- Hongo F, Yamada Y, Ueda T, et al. Preoperative lipiodol marking and its role on survival and complication rates of CT-guided cryoablation for small renal masses. *BMC Urol*. 2017;17(1):10. [\[CrossRef\]](#)
- Kajiwaru K, Yoshimatsu R, Komoto M, et al. Efficacy and safety of CT-guided cryoablation after lipiodol marking and embolization for RCC. *Minim Invasive Ther Allied Technol*. 2022;31(6):923-929. [\[CrossRef\]](#)
- Lopez O, Chevallier O, Guillen K, et al. selective arterial embolization with N-butyl cyanoacrylate prior to CT-guided percutaneous cryoablation of kidney malignancies: a single-center experience. *J Clin Med*. 2021;10(21):4986. [\[CrossRef\]](#)
- Ushijima Y, Asayama Y, Nishie A, et al. Cryoablation for secondary renal cell carcinoma after surgical nephrectomy. *Cardiovasc Intervent Radiol*. 2021;44(3):414-420. [\[CrossRef\]](#)
- Young JL, Clayman RV. Cryoprobe isotherms: a caveat and review. *J Endourol*. 2010;24(5):673-676. [\[CrossRef\]](#)



Copyright©Author(s) - Available online at dirjournal.org.
Content of this journal is licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License.

Reproducibility of rCBV in glioblastomas using T2*-weighted perfusion MRI: an evaluation of sampling, normalization, and experience

Sabahattin Yüzkan
 Samet Mutlu
 Mehmet Karagülle
 Merve Şam Özdemir
 Hamit Özgül
 Mehmet Ali Arıkan
 Burak Koçak

University of Health Sciences, Başakşehir Çam and Sakura City Hospital, Clinic of Radiology, İstanbul, Türkiye

PURPOSE

The reproducibility of relative cerebral blood volume (rCBV) measurements among readers with different levels of experience is a concern. This study aimed to investigate the inter-reader reproducibility of rCBV measurement of glioblastomas using the hotspot method in dynamic susceptibility contrast perfusion magnetic resonance imaging (DSC-MRI) with various strategies.

METHODS

In this institutional review board-approved single-center study, 30 patients with glioblastoma were retrospectively evaluated with DSC-MRI at a 3.0 Tesla scanner. Three groups of reviewers, including neuroradiologists, general radiologists, and radiology residents, calculated the rCBV based on the number of regions of interest (ROIs) and reference areas. For statistical analysis of feature reproducibility, the intraclass correlation coefficient (ICC) and Bland-Altman plots were used. Analyses were made among individuals, reader groups, reader-group pooling, and a population that contained all of them.

RESULTS

For individuals, the highest inter-reader reproducibility was observed between neuroradiologists [ICC: 0.527; 95% confidence interval (CI): 0.21–0.74] and between residents (ICC: 0.513; 95% CI: 0.20–0.73). There was poor reproducibility in the analyses of individuals with different levels of experience (ICC range: 0.296–0.335) and in reader-wise and group-wise pooling (ICC range: 0.296–0.335 and 0.397–0.427, respectively). However, an increase in ICC values was observed when five ROIs were used. In an analysis of all strategies, the ICC for the centrum semiovale was significantly higher than that for contralateral white matter ($P < 0.001$).

CONCLUSION

The inter-reader reproducibility of rCBV measurement was poor to moderate regardless of whether it was calculated by neuroradiologists, general radiologists, or residents, which may indicate the need for automated methods. Choosing five ROIs and using the centrum semiovale as a reference area may increase reliability for all users.

KEYWORDS

Cerebral blood volume, dynamic susceptibility contrast, glioblastoma, magnetic resonance imaging, observer variation

Corresponding author: Burak Koçak

E-mail: drburakkocak@gmail.com

Received 27 July 2023; revision requested 02 September 2023; accepted 17 September 2023.



Epub: 03.10.2023

Publication date: 05.03.2024

DOI: 10.4274/dir.2023.232442

In neuroradiology practice, dynamic susceptibility contrast perfusion magnetic resonance imaging (DSC-MRI) is used extensively as an advanced method for the diagnosis, grading, and post-treatment follow-up of glioblastomas.^{1,2} The DSC-MRI technique depends on a susceptibility-caused signal loss on T2*-weighted images resulting from a bolus passage of gadolinium-based contrast media. Cerebral blood volume (CBV) is the most commonly used parameter of DSC-MRI and it defines the area under the concentration–time curve.³ CBV is basically an absolute value however, it has some assumptions/conditional requirements. Therefore, to obtain a relative quantification, it is usually rated to a reference point, such as contralateral white matter, the centrum semiovale, or arterial input function.⁴ The relative CBV (rCBV) is the most robust and commonly used DSC-MRI parameter for the radiological characteri-

You may cite this article as: Yüzkan S, Mutlu S, Karagülle M, et al. Reproducibility of rCBV in glioblastomas using T2*-weighted perfusion MRI: an evaluation of sampling, normalization, and experience. *Diagn Interv Radiol.* 2024;30(2):124-134.

zation of glioblastomas.^{3,5} It has been proven that the rCBV value is highly correlated with tumor grade, vascularity, and prognosis.^{2,3} Moreover, rCBV has been shown to be useful in distinguishing tumor progression from its mimickers, such as pseudoprogression and radiation necrosis.^{3,6,7}

To calculate rCBV, radiologists draw regions of interest (ROIs) in the most hyperperfused area of a tumor and a normal-appearing reference area.⁸ In ROI analysis, determining the hyperperfused area on CBV map images and selecting the normal-appearing reference area are two important operator-dependent subjective issues. Despite its operator dependency and inter-observer variability, this is the most widely used method in clinical practice.⁹⁻¹²

Radiologists with different experience levels can potentially assess DSC-MRI. However, the assessment of DSC-MRI parameters by radiologists with different levels of experience, such as radiology residents, general radiologists, and neuroradiologists, may lead to inconsistent evaluations in the diagnosis and treatment processes. It is crucial to improve a reliable and standard analysis method for rCBV measurement to eliminate incompatibility between different users. It has been recommended that DSC-MRI measurements should be reviewed by two experienced radiologists and an adjudicator should be consulted in the event of disagreement.¹³ However, this recommendation is time-consuming and not always applicable; therefore, it is impractical for clinical practice. In clinical practice, the assessment of DSC-MRI by a single radiology resident (when preparing reports), a general radiologist, or an experienced neuroradiologist is not uncommon. Residents and general radiologists do not have the opportunity to consult an experienced neuroradiologist in all DSC-MRI examinations.

To our knowledge, there is limited literature on the reproducibility aspects of rCBV

in terms of sampling, normalization location, and reader experience. Therefore, we sought to fill this knowledge gap in the literature by assessing the inter-reader reproducibility of rCBV measurements in glioblastoma cases in pre-surgical settings, calculated either by radiology residents, inexperienced general radiologists, or experienced neuroradiologists. Second purpose was to investigate the effect of the number of ROI and the selection of reference areas on inter-reader reproducibility.

Methods

Ethics

For this retrospective study, institutional review board approval was acquired from the Local Medical Ethics Committee of Basaksehir Cam and Sakura City Hospital (no: 2023-18; decision date: 11/01/2023). The requirement for written informed consent was waived by the ethics committee due to the study's retrospective design. The study adhered to the principles of the Helsinki Declaration.

Study population

In this single-center retrospective study, 30 patients with newly diagnosed glioblastoma were consecutively included between July 2021 and January 2023. The reason for including 30 cases was to meet the minimum requirement for a reliability analysis.¹⁴ This study was conducted in a tertiary academic hospital.

Eligibility criteria

The inclusion criteria were as follows: *i*, a definitive histopathologic diagnosis of glioblastoma according to the World Health Organization's 2021 classification;¹⁵ *ii*, ≥18 years old at the time of DSC-MRI.

The following criteria determined exclusion: *i*, history of cranial surgical resection or biopsy, or radiotherapy, chemotherapy, corticosteroid, or anti-angiogenic treatment before imaging; *ii*, MRI scans with severe artifacts that impeded the diagnostic evaluation.

Magnetic resonance imaging acquisition protocol and processing

All MRI scans were performed at a 3.0 Tesla system (Ingenia, Philips Healthcare) using 32-channel phased-array head coils in the supine position. The routine DSC-MRI examination protocol included axial spin-echo T1-weighted imaging [repetition time (TR)/echo time (TE): 600/10 ms; field of view

(FOV): 230 mm; slice thickness: 4 mm; matrix: 208 × 165; number of excitations (NEX): 1], a 3D axial fluid-attenuated inversion recovery sequence (TR/TE: 4,800/340 ms; inversion time: 1,650 ms; FOV: 230 mm; slice thickness: 4 mm; matrix: 272 × 243; NEX: 1), subsequent DSC-MRI data, and finally, 3D postcontrast T1-weighted imaging (TR/TE: 600/10 ms; FOV: 230 mm; slice thickness: 4 mm; matrix: 208 × 165; NEX: 1).

DSC-MRI was obtained on an axial plan with a gradient-echo echoplanar imaging technique using the following parameters: TR/TE: 1,500/30 ms; FOV: 237 × 237 mm; matrix: 128 × 128; section thickness: 3 mm; flip angle: 60°; voxel size: 2.33 × 2.39 × 4.00 mm. An intravenous bolus injection of gadolinium-based contrast agent was given at a dose of 0.1 mmol/kg and a speed of 5 mL/s, followed by a 20-mL saline flush. Before the dynamic phase, a saturation pre-bolus of contrast agent was administered as a preload to reduce contaminating T1 effects from contrast agent leakage. The DSC-MRI protocol of this study was in line with consensus recommendations.¹⁶

Dynamic susceptibility contrast perfusion magnetic resonance imaging processing and observer setting

The DSC-MRI datasets were processed using the IntelliSpace Portal (Philips). Maximum rCBV values were calculated independently by three groups of readers (six readers in total). Individual readers were denoted as R1, R2, R3, R4, R5, and R6.

Readers R1 and R2 were neuroradiologists with 5 years of experience in neuroimaging in research settings. Readers R3 and R4 were general radiologists with 4 years of experience in general radiology who were working in all divisions of radiology. Readers R5 and R6 were third-year radiology residents who had completed neuroradiology rotations at the beginning of this study. All readers were blinded to the clinical information and histopathological results of patients. All data were anonymized.

Inter-reader reproducibility was assessed using various strategies as follows: *i*, between individuals with the same experience levels (R1 vs. R2, R3 vs. R4, and R5 vs. R6); *ii*, between individuals with different experience levels (R1 vs. R3, R3 vs. R5, and R1 vs. R5). Since two readers in each group had similar experience levels and to avoid complicating the analysis, only one reader was randomly selected from each group and compared with the other readers; *iii*, between reader groups with dif-

Main points

- Reproducibility of relative cerebral blood volume (rCBV) measurement among readers with different experiences is a concern.
- The inter-reader reproducibility of rCBV measurement was poor to moderate when using the hotspot method, regardless of whether it was calculated by neuroradiologists, general radiologists, or radiology residents.
- Sampling five regions of interest and selecting the centrum semiovale for normalization improved reproducibility.

ferent experience levels (neuroradiologists vs. general radiologists, neuroradiologists vs. residents, and general radiologists vs. residents); *iv*, inter-reader reproducibility of group-wise (neuroradiologists vs. general radiologists vs. residents) and reader-wise (R1 vs. R2 vs. R3 vs. R4 vs. R5 vs. R6) pooling.

Region of interest analysis

The reviewers were encouraged to place five different ROIs within the tumor that visually appeared as mostly hyperperfused on colored relative CBV map images (hotspot method). The CBV value of the first ROIs (CBV₁), the highest CBV value among the first three ROIs (CBV₃), and the highest CBV value among five ROIs (CBV₅) were recorded. Only the highest CBV value among the three and five ROIs was used (not their means). Then, the reviewers were instructed to place an ROI of the same size on the contralateral normal-appearing white matter in the same axial section as the tumor's ROI and the contralateral normal-appearing centrum semiovale, which are known to be the most reliable reference areas.¹³ The CBV values obtained from CBV₁, CBV₃, and CBV₅ were rated to these reference areas to obtain a normalized rCBV. Statistical analyses were performed separately for each reference area.

All circular ROIs were drawn manually by the readers on CBV map images and ranged between 40 and 60 mm². In the ROI analyses, care was taken to avoid hemorrhagic, necrotic, or cystic regions, normal grey matter, and intraslesional non-tumor large vessels that might affect the values. A multi-ROI analysis of a glioblastoma case is represented in Figure 1.

Statistical analysis

The statistical analysis was performed using R 4.3 (main packages: "MKinfer" and "ggpubr" for reliability analyses and inferential statistics; tool: JASP for descriptive statistics only)¹⁷ and Python 3.7 (main package: pingouin for reliability analyses) environments.¹⁸ To assess feature reliability, the mean and 95% CI values of the intraclass correlation coefficient (ICC) were calculated. The ICC was based on the type of ICC (2,1) according to Shrout and Fleiss's convention.¹⁹ The interpretation scale for the ICC was as follows: <0.5: poor; ≤0.5 to <0.75: moderate; ≤0.75 to <0.9: good; and ≥0.9: excellent.¹⁴ In addition to the ICC analysis, non-parametric Bland–Altman analyses were performed to evaluate the differences in measurements and the limits of agreement, relying on median and 2.5–97.5th

percentiles, respectively. The Shapiro–Wilk test was used to determine the normality of continuous variables. Depending on the group distributions, a paired non-parametric test, the Wilcoxon signed-rank test, was used to assess pairwise statistical differences in continuous variables. Statistical results were considered significant if $P < 0.050$. In the case of multiple comparisons, the results were considered significant if the adjusted P values were <0.050 after multiplicity correction using the Bonferroni method.

Results

Patient characteristics

In total, 30 consecutive glioblastoma cases were enrolled, 17 were male, and 13 were female. The mean age of the patients (standard deviation) was 61.1 (9.7) years (range: 38–78 years).

According to the number of ROIs and reference areas, the median rCBV values ranged between 13.7 and 20.1 for neuroradiologists, 18.1 and 22.1 for general radiologists, and 10.1 and 12.8 for residents. The median and the interquartile range (IQR) of the rCBV values of all readers, which were calculated by using the hotspot method, are presented in Supplementary Table 1.

Inter-reader reproducibility for individuals with similar experience levels

The inter-reader reproducibility of the rCBV measurements was poor to moder-

ate (ICC range: 0.288–0.527). The highest inter-reader reproducibility of the rCBV measurements was obtained between neuroradiologists (R1 vs. R2) using one ROI and normalization with white matter (ICC: 0.527; 95% CI: 0.21–0.74) and between residents (R5 vs. R6) using one ROI and normalization with the centrum semiovale (ICC: 0.513; 95% CI: 0.20–0.73). The ICC value of general radiologists (R3 vs. R4) increased from 0.312 to 0.370, with a higher number of ROIs (from one to five) using the centrum semiovale as a reference area. The ICC value of each analysis is presented in detail in Table 1.

Inter-reader reproducibility for individuals with different experience levels

In all analyses, the inter-reader reproducibility of the rCBV measurements was poor (ICC range: 0.296–0.335). However, an increase in ICC values was observed when five ROIs were used instead of one or three ROIs, even if contralateral white matter or the centrum semiovale is used for normalization (Supplementary Table 2).

Inter-reader reproducibility for reader groups with different experience levels

In the analysis of reader groups, the inter-reader reproducibility of rCBV measurements was moderate (ICC range: 0.566–0.640) for neuroradiologists vs. general radiologists. On the other hand, inter-reader reproducibility was poor for all the other group-based analyses. The ICC values ranged

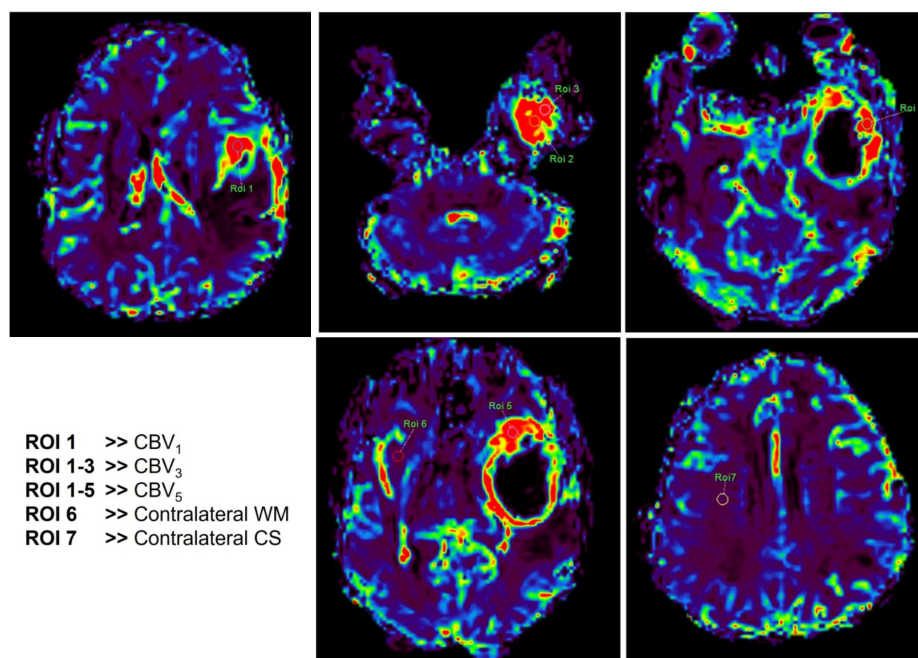


Figure 1. Sampling of the different regions of interest (ROIs) on color maps of cerebral blood volume (CBV). CBV₁, CBV value of the first ROI; CBV₃, highest CBV value among the first three ROIs; CBV₅, highest CBV value among five ROIs; WM, white matter; CS, centrum semiovale.

between 0.350 and 0.422 for neuroradiologists vs. residents and between 0.254 and 0.334 for general radiologists vs. residents (Table 2).

Inter-reader reproducibility for group-wise and reader-wise pooling

In both the group-wise (all groups pooled) and reader-wise (all readers pooled) pooling analyses, inter-reader reproducibility was poor. The ICC ranged between 0.397 and 0.427 and between 0.296 and 0.335 for group-wise pooling and reader-wise pooling, respectively. Similar to the results of individuals with different experience levels, an increase in ICC values was observed when five ROIs were used. All the ICC values are presented in Table 3.

Analysis of the overall reader population of perfusion magnetic resonance imaging

An additional analysis, that included all the readers, groups, and pooled analyses, was performed ($n = 66$). The ICC values of all the analyses performed are summarized in Figure 2.

While the inter-reader reproducibility of the general radiologists (R3 vs. R4) was poor (Table 1), six of the top 10 most reproducible analyses involved neuroradiologists vs. general radiologists. Among all the analyses, the top two in terms of inter-reader reproducibility was for neuroradiologists vs. general radiologists using the centrum semiovale, with an ICC value of 0.640 (95% CI: 0.32–0.82) and 0.583 (95% CI: 0.29–0.78) for one ROI and five ROIs, respectively. The top 10 most reproducible results among all the analyses are presented in Table 4.

In this analysis, the median of the ICC value was 0.349 (IQR: 0.116) for the centrum semiovale and 0.305 (IQR: 0.107) for white matter. The inter-reader agreement was higher, with a statistically significant difference for the centrum semiovale ($P < 0.001$). Comparisons were performed according to the number of ROIs. The median of the ICC value was 0.335 (IQR: 0.117) for five ROIs, 0.321 (IQR: 0.128) for one ROI, and 0.316 (IQR: 0.112) for three ROIs. There was a significant difference between three and five ROIs, including all reference areas ($P < 0.001$). In addition, a significant difference was observed between three and five ROIs when using one of the two reference areas ($P < 0.010$). There was no significant difference in the use of one or three ROIs ($P > 0.050$). All results are summarized in Figure 3.

Figures 4 and 5 show Bland–Altman plots of the readers based on centrum semiovale normalization with the same and different experience levels, respectively. Supplemen-

tary Figures 1 and 2 show Bland–Altman plots based on white matter normalization. In all the Bland–Altman analyses, the vast majority of ROI measurements were within

Table 1. Intraclass correlation coefficients for readers with similar experience levels

Analysis	Location of ROIs for normalization	Number of ROIs	ICC	LB of 95% CI	UB of 95% CI
R1 vs. R2	Centrum semiovale	One	0.489	0.15	0.72
	Centrum semiovale	Three	0.437	0.11	0.68
	Centrum semiovale	Five	0.457	0.13	0.70
	White matter	One	0.527	0.21	0.74
	White matter	Three	0.474	0.14	0.71
	White matter	Five	0.497	0.17	0.73
R3 vs. R4	Centrum semiovale	One	0.312	−0.09	0.63
	Centrum semiovale	Three	0.349	−0.03	0.64
	Centrum semiovale	Five	0.370	0	0.65
	White matter	One	0.301	−0.05	0.59
	White matter	Three	0.254	−0.06	0.54
	White matter	Five	0.254	−0.07	0.54
R5 vs. R6	Centrum semiovale	One	0.513	0.20	0.73
	Centrum semiovale	Three	0.464	0.14	0.70
	Centrum semiovale	Five	0.472	0.15	0.71
	White matter	One	0.288	−0.05	0.57
	White matter	Three	0.312	−0.02	0.59
	White matter	Five	0.314	−0.01	0.59

ROI, region of interest; ICC, intraclass correlation coefficient, LB, lower bound; UB, upper bound; CI, confidence interval; R1–R6, readers 1 to 6.

Table 2. Intraclass correlation coefficients for reader groups with different experience levels

Analysis	Location of ROIs for normalization	Number of ROIs	ICC	LB of 95% CI	UB of 95% CI
GenRad vs. Res	Centrum semiovale	One	0.286	−0.10	0.63
	Centrum semiovale	Three	0.319	−0.10	0.66
	Centrum semiovale	Five	0.334	−0.11	0.67
	White matter	One	0.254	−0.09	0.56
	White matter	Three	0.299	−0.09	0.62
	White matter	Five	0.308	−0.09	0.62
NeuRad vs. GenRad	Centrum semiovale	One	0.640	0.32	0.82
	Centrum semiovale	Three	0.566	0.27	0.77
	Centrum semiovale	Five	0.583	0.29	0.78
	White matter	One	0.571	0.26	0.77
	White matter	Three	0.577	0.28	0.77
	White matter	Five	0.573	0.27	0.77
NeuRad vs. Res	Centrum semiovale	One	0.422	−0.10	0.76
	Centrum semiovale	Three	0.350	−0.10	0.68
	Centrum semiovale	Five	0.398	−0.10	0.72
	White matter	One	0.377	−0.05	0.67
	White matter	Three	0.380	−0.05	0.68
	White matter	Five	0.404	−0.04	0.70

ROI, region of interest; ICC, intraclass correlation coefficient; LB, lower bound; UB, upper bound; CI, confidence interval; NeuRad, neuroradiologist; GenRad, general radiologist; Res, resident.

the upper and lower agreement limits. In the analysis based on white matter normalization, most of ROI measurements by neuroradiologists were quite close to the zero line (Supplementary Figure 1). The same condition was observed for residents using centrum semiovale normalization (Figure 4). In general, ROI measurements were far from the zero line for general radiologists (Figure 4 and Supplementary Figure 1).

Discussion

In the present study, we investigated the inter-reader reproducibility of rCBV measurements in patients with untreated glioblastomas. Generally, a moderate inter-reader agreement was observed when analyses were made between individuals with the similar experience level. In contrast, there was poor inter-reader reproducibility when the analyses were made between different experience levels. According to all the analyses performed, the inter-reader agreement of rCBV measurements when using the centrum semiovale as a reference area was significantly higher than when using contralateral white matter. In addition, the ICC values for the placement of five ROIs were significantly higher than with one or three ROIs.

In clinical radiology practice, T2*-weighted perfusion MRI is often used for gliomas. The radiology community pays relatively little attention to the reproducibility of the derived parameters, despite their importance. In this study, we focused on preoperative glioblastoma cases to make the findings more evident, i.e., to assess the reliability of obviously high perfusion values. In clinical practice, reproducible perfusion parameters are essential for the consistent target area selection of gliomas in stereotactic biopsies and for establishing consistent baseline perfusion parameter values for use in post-treatment follow-up scans. Additionally, the reproducibility of these parameters is necessary for research consistency to ensure that the results of different studies are comparable, which may increase the validity of the conclusions drawn from pooled data and meta-analyses.

In previously published reliability studies, only normal-appearing contralateral white matter^{20,21} or the contralateral centrum semiovale¹³ has been generally selected as a reference area. In these studies, three observers were selected to assess DSC-MRI, including only neuroradiologists^{13,21} or neuroradiologists and a resident.^{20,22} We noticed that there was sparse radiological research

literature on inter-reader reproducibility analyses of rCBV measurements for a population that includes neuroradiologists, general radiologists, and residents in the same study. According to our experience and knowledge, in clinical practice, general radiologists and residents (when preparing reports to present to neuroradiologists) may have to evaluate

DSC-MRI, although this is not as common as with neuroradiologists. Therefore, inter-reader reproducibility of general radiologists and residents within themselves and between other reviewers is a matter of concern.

Definitive interpretations of DSC-MRI in patients with glioblastoma should be con-

Table 3. Intraclass correlation coefficients for group-wise and reader-wise pooling

Analysis	Location of ROIs for normalization	Number of ROIs	ICC	LB of 95% CI	UB of 95% CI
All groups pooled	Centrum semiovale	One	0.424	0.06	0.69
	Centrum semiovale	Three	0.401	0.09	0.66
	Centrum semiovale	Five	0.426	0.10	0.68
	White matter	One	0.397	0.12	0.64
	White matter	Three	0.419	0.13	0.66
	White matter	Five	0.427	0.14	0.66
All readers pooled	Centrum semiovale	One	0.321	0.14	0.53
	Centrum semiovale	Three	0.313	0.15	0.51
	Centrum semiovale	Five	0.335	0.17	0.53
	White matter	One	0.296	0.15	0.49
	White matter	Three	0.297	0.15	0.48
	White matter	Five	0.305	0.16	0.49

ROI, region of interest; ICC, intraclass correlation coefficient; LB, lower bound; UB, upper bound; CI, confidence interval.

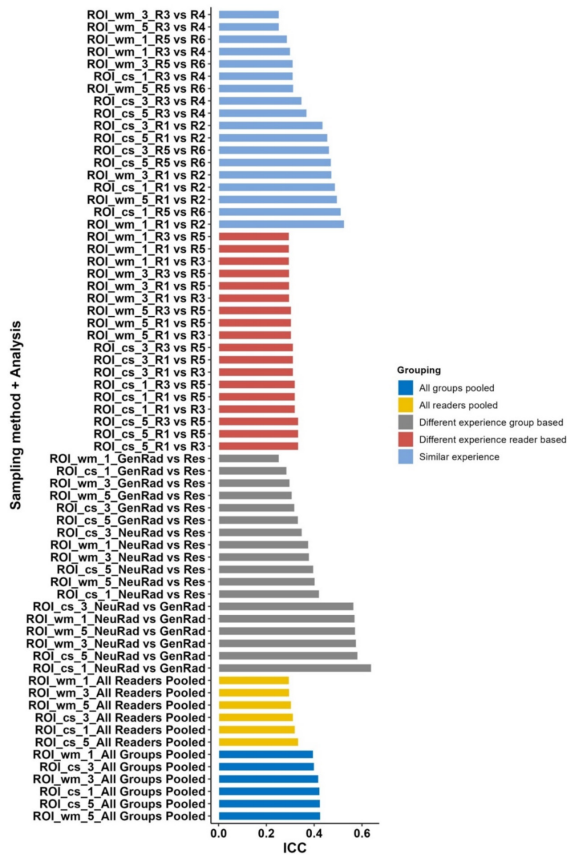


Figure 2. Intraclass correlation coefficient values of all analyses. wm, white matter; cs, centrum semiovale; ROI, region of interest; ICC, intraclass correlation coefficient; NeuRad, neuroradiologist; GenRad, general radiologist; Res, resident; R1–R6, readers 1–6.

ducted and interpreted by specialized neuroradiologists. The most valuable advantage and originality of our paper is that it included two different groups in addition to neuroradiologists in the same study. For all analyses, normalization with the centrum semiovale acquired better inter-reader reproducibility than with white matter. According to our observation, the reason for this difference could be explained by the fact that the centrum semiovale is a large homogenous area that is mostly visible in more than two axial slices, and it suffers less from partial volume artifacts compared with contralateral white matter. In a recently published retrospective study, Roques et al.²⁰ evaluated the inter-observer reproducibility of rCBV measurements in 27 cases of untreated glioblastoma. In that study, three observers (two neuroradiologists and a radiology resident) calculated the maximum rCBV values independently using the hotspot method, similar to our study. However, differently, they used only contralateral white matter as a reference to normalize the CBV value. Their inter-observer reproducibility for maximum rCBV value measurements was fair [ICC: 0.46 (0.22–0.67)], but their inter-observer reproducibility was found to be poor to fair (ICC range: 0.30–0.47) when a resident was added to the analysis. Our research confirms the variability of the hotspot method, similar to the results

of Roques et al.²⁰ study. However, their study did not include general radiologists and did not assess inter-observer reliability among radiology residents or general radiologists, which are the main differences of our study. Another advantage of our research is the use of leakage correction with a gadolinium preload to avoid the underestimation of CBV values. Furthermore, our study assessed the effect of selecting two different reference areas on inter-observer reliability.

The present study has the following limitations: a small sample size, a retrospective

nature, and involvement of only a single institution. In this research, we only studied cases with untreated glioblastomas, in which DSC–MRI perfusion is anticipated to be increased. This study was conducted only on the initial DSC–MRI examinations of untreated glioblastomas and did not include the evaluation of post-treatment perfusion MRI features. Including only the baseline DSC–MRIs in our study may have contributed to reliability.

In conclusion, there is poor to moderate inter-reader reproducibility of rCBV measure-

Table 4. Top 10 most reproducible analyses according to intraclass correlation coefficient

Analysis	Location of ROIs for normalization	Number of ROIs	ICC	LB of 95% CI	UB of 95% CI
NeuRad vs. GenRad	Centrum semiovale	One	0.640	0.32	0.82
NeuRad vs. GenRad	Centrum semiovale	Three	0.566	0.27	0.77
NeuRad vs. GenRad	Centrum semiovale	Five	0.583	0.29	0.78
NeuRad vs. GenRad	White matter	One	0.571	0.26	0.77
NeuRad vs. GenRad	White matter	Three	0.577	0.28	0.77
NeuRad vs. GenRad	White matter	Five	0.573	0.27	0.77
R1 vs. R2	Centrum semiovale	One	0.489	0.15	0.72
R1 vs. R2	White matter	One	0.527	0.21	0.74
R1 vs. R2	White matter	Five	0.497	0.17	0.73
R5 vs. R6	Centrum semiovale	One	0.513	0.20	0.73

ROI, region of interest; ICC, intraclass correlation coefficient; LB, lower bound; UB, upper bound; CI, confidence interval; NeuRad, neuroradiologist; GenRad, general radiologist; R1–R6, readers 1–6.

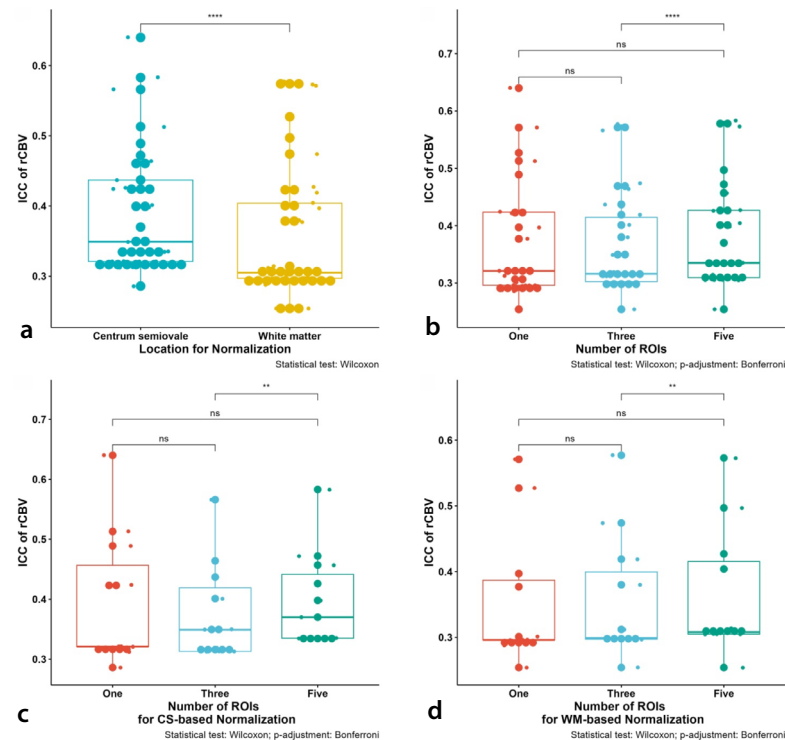


Figure 3. Box plots and statistical comparisons for distribution of mean intraclass correlation coefficient (ICC) of relative cerebral blood volume (rCBV) values. Comparisons are performed according to location (a) and the number of regions of interest (ROIs) (b–d) used in the measurements. The analysis is based on all reliability analyses (n = 66) combined. ns, $P > 0.050$; **, $P < 0.010$; ****, $P < 0.0001$. ns, not significant.

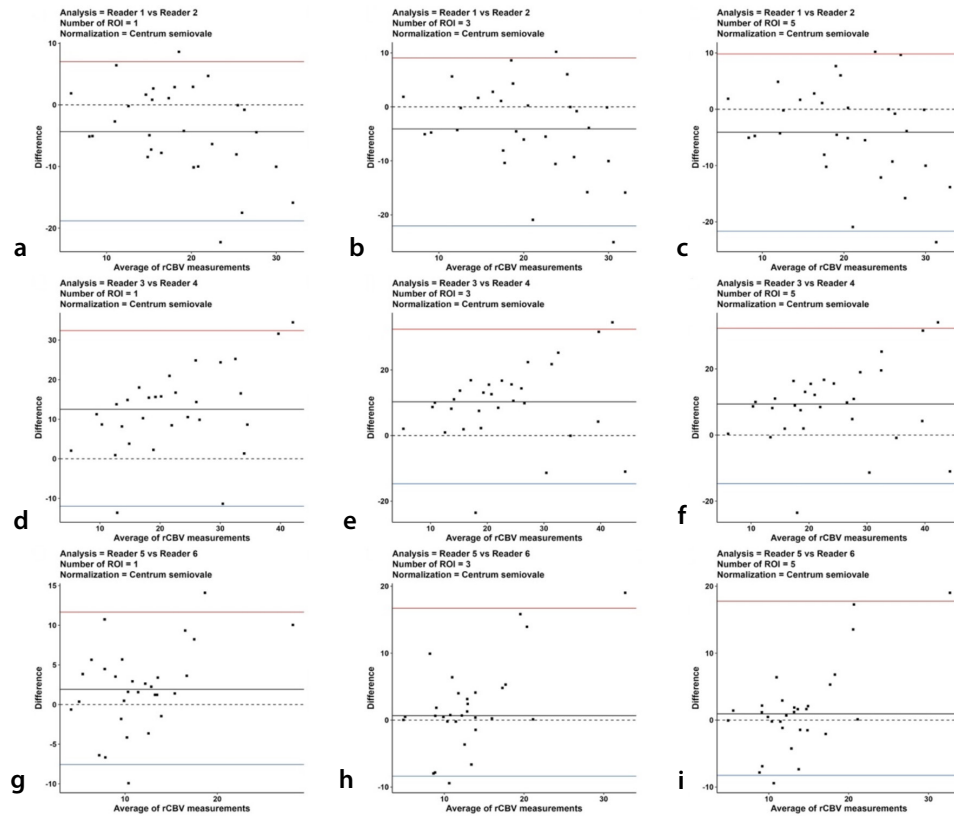


Figure 4. Bland–Altman plots for readers with similar experience levels (a–c for neuroradiologists, R1 vs. R2; d–f for general radiologists, R3 vs. R4; g–i for residents, R5 vs. R6). Relative cerebral blood volume (rCBV) values are based on centrum semiovale normalization. The analysis is non-parametric and relies on the median. Solid black, red, and blue lines show the medians of difference, the upper level of agreement bound (97.5th percentile), and the lower level of agreement bound (2.5th percentile), respectively. The dashed line stands for no difference. ROI, region of interest.

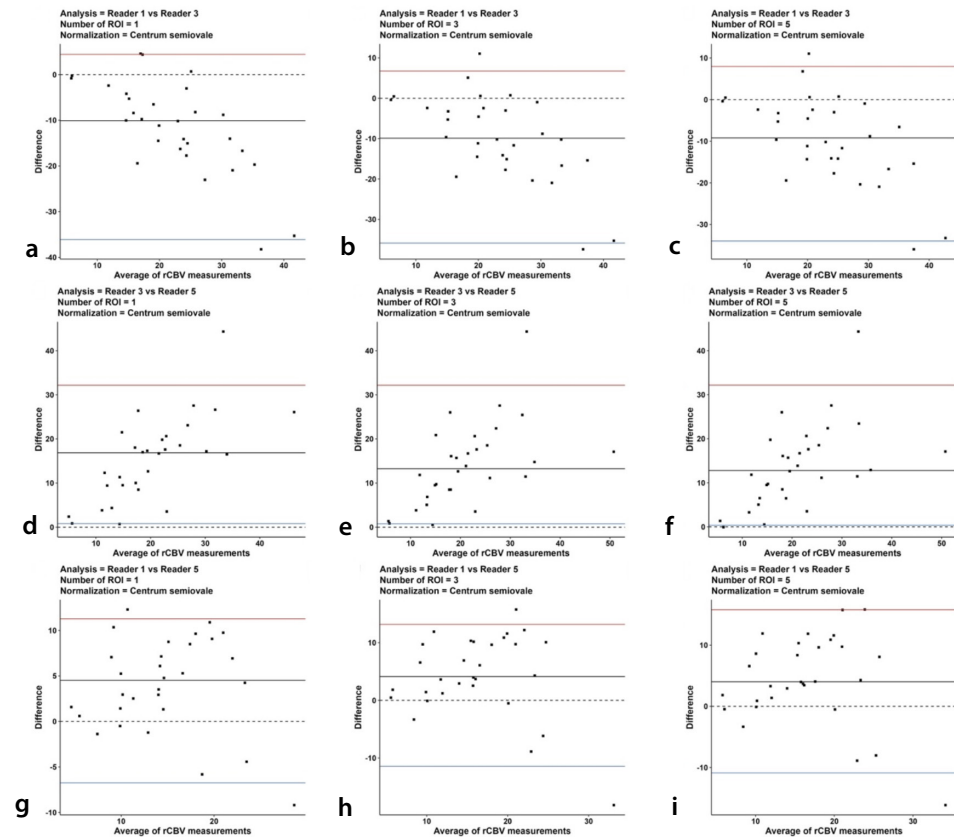


Figure 5. Bland–Altman plots for readers with different experience levels (a–c for neuroradiologist vs. general radiologist; d–f for general radiologist vs. resident; g–i for neuroradiologist vs. resident). Relative cerebral blood volume (rCBV) values are based on centrum semiovale normalization. The analysis is non-parametric and relies on the median. Solid black, red, and blue lines represent the medians of difference, the upper level of agreement bound (97.5th percentile), and the lower level of agreement bound (2.5th percentile), respectively. The dashed line stands for no difference. ROI, region of interest.

ments using the hotspot method, regardless of whether they are calculated by neuroradiologists, general radiologists, or radiology residents. This may indicate the need for automated methods. Selecting five ROIs and using the centrum semiovale as a reference area for normalization may increase the inter-reader reproducibility of measurements.

Conflict of interest disclosure

Burak Koçak, MD, is Section Editor in Diagnostic and Interventional Radiology. He had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Other authors have nothing to disclose.

References

- Shiroishi MS, Castellazzi G, Boxerman JL, et al. Principles of T2*-weighted dynamic susceptibility contrast MRI technique in brain tumor imaging. *J Magn Reson Imaging*. 2015;41(2):296-313. [\[CrossRef\]](#)
- Thust SC, Heiland S, Falini A, et al. Glioma imaging in Europe: a survey of 220 centres and recommendations for best clinical practice. *Eur Radiol*. 2018;28(8):3306-3317. [\[CrossRef\]](#)
- Essig M, Shiroishi MS, Nguyen TB, et al. Perfusion MRI: the five most frequently asked technical questions. *AJR Am J Roentgenol*. 2013;200(1):24-34. [\[CrossRef\]](#)
- Saini J, Gupta RK, Kumar M, et al. Comparative evaluation of cerebral gliomas using rCBV measurements during sequential acquisition of T1-perfusion and T2*-perfusion MRI. *PLoS One*. 2019;14(4):e0215400. [\[CrossRef\]](#)
- Cha S, Knopp EA, Johnson G, Wetzel SG, Litt AW, Zagzag D. Intracranial mass lesions: dynamic contrast-enhanced susceptibility-weighted echo-planar perfusion MR imaging. *Radiology*. 2002;223(1):11-29. [\[CrossRef\]](#)
- Law M, Yang S, Wang H, et al. Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging. *AJNR Am J Neuroradiol*. 2003;24(10):1989-1998. [\[CrossRef\]](#)
- Boxerman JL, Ellingson BM, Jeyapalan S, et al. Longitudinal DSC-MRI for distinguishing tumor recurrence from pseudoprogression in patients with a high-grade glioma. *Am J Clin Oncol*. 2017;40(3):228-234. [\[CrossRef\]](#)
- Sadeghi N, D'Haene N, Decaestecker C, et al. Apparent diffusion coefficient and cerebral blood volume in brain gliomas: relation to tumor cell density and tumor microvessel density based on stereotactic biopsies. *AJNR Am J Neuroradiol*. 2008;29(3):476-482. [\[CrossRef\]](#)
- Rani N, Singh B, Kumar N, et al. Differentiation of recurrent/residual glioma from radiation necrosis using semi quantitative 99mTc MDM (bis-methionine-DTPA) brain SPECT/CT and dynamic susceptibility contrast-enhanced MR perfusion: a comparative study. *Clin Nucl Med*. 2018;43(3):74-81. [\[CrossRef\]](#)
- Santarosa C, Castellano A, Conte GM, et al. Dynamic contrast-enhanced and dynamic susceptibility contrast perfusion MR imaging for glioma grading: preliminary comparison of vessel compartment and permeability parameters using hotspot and histogram analysis. *Eur J Radiol*. 2016;85(6):1147-1156. [\[CrossRef\]](#)
- Dijkstra H, Sijens PE, van der Hoorn A, van Laar PJ. Inter-observer reproducibility of quantitative dynamic susceptibility contrast and diffusion MRI parameters in histogram analysis of gliomas. *Acta Radiol*. 2020;61(1):76-84. [\[CrossRef\]](#)
- Oei MTH, Meijer FJA, Mordang JJ, et al. Observer variability of reference tissue selection for relative cerebral blood volume measurements in glioma patients. *Eur Radiol*. 2018;28(9):3902-3911. [\[CrossRef\]](#)
- Smits M, Bendszus M, Collette S, et al. Repeatability and reproducibility of relative cerebral blood volume measurement of recurrent glioma in a multicentre trial setting. *Eur J Cancer*. 2019;114:89-96. [\[CrossRef\]](#)
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. [\[CrossRef\]](#)
- Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol*. 2021;23(8):1231-1251. [\[CrossRef\]](#)
- Boxerman JL, Quarles CC, Hu LS, et al. Consensus recommendations for a dynamic susceptibility contrast MRI protocol for use in high-grade gliomas. *Neuro Oncol*. 2020;22(9):1262-1275. [\[CrossRef\]](#)
- R Core Team. R: A Language and Environment for Statistical Computing. 2022. [\[CrossRef\]](#)
- Van Rossum G, Drake FL. Python 3 Reference Manual. 2009. [\[CrossRef\]](#)
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428. [\[CrossRef\]](#)
- Roques M, Raveneau M, Adam G, et al. Reproducibility of volume analysis of dynamic susceptibility contrast perfusion-weighted imaging in untreated glioblastomas. *Neuroradiology*. 2022;64(9):1763-1771. [\[CrossRef\]](#)
- Jung SC, Choi SH, Yeom JA, et al. Cerebral blood volume analysis in glioblastomas using dynamic susceptibility contrast-enhanced perfusion MRI: a comparison of manual and semiautomatic segmentation methods. *PLoS One*. 2013;8(8):e69323. [\[CrossRef\]](#)
- Kouwenberg V, van Santwijk L, Meijer FJA, Henssen D. Reliability of dynamic susceptibility contrast perfusion metrics in pre- and post-treatment glioma. *Cancer Imaging*. 2022;22(1):28. [\[CrossRef\]](#)

Supplementary Table 1. Median of relative cerebral blood volume measurements, with the interquartile range for each group and reference area

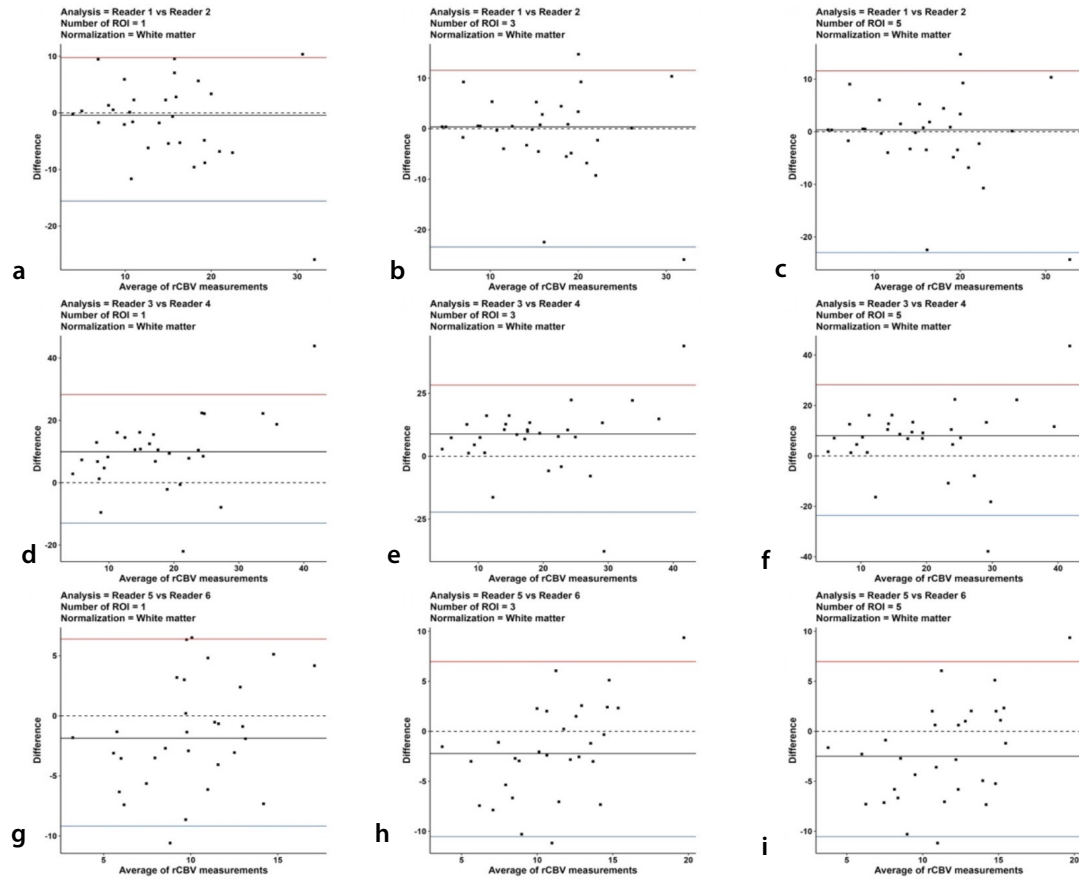
Value	Number of ROIs	NeuRad		Res		GenRad		Total	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
rCBV for WM	One	13.7	8.7	10.1	5.5	18.1	13.5	12.2	9.8
	Three	15.7	9.7	11.6	6.0	18.9	13.0	13.6	10.0
	Five	15.7	9.9	11.7	6.2	19.3	13.1	14.2	10.7
rCBV for CS	One	17.4	11.8	11.4	5.4	19.8	17.2	14.8	12.1
	Three	20.1	11.3	12.4	4.7	20.3	16.5	16.3	13.0
	Five	20.1	11.4	12.8	4.9	22.1	16.3	16.7	12.7

rCBV, relative cerebral blood volume; WM, contralateral white matter; CS, contralateral centrum semiovale, ROI, region of interest; IQR, interquartile range; NeuRad, neuroradiologist; GenRad, general radiologist; Res, resident.

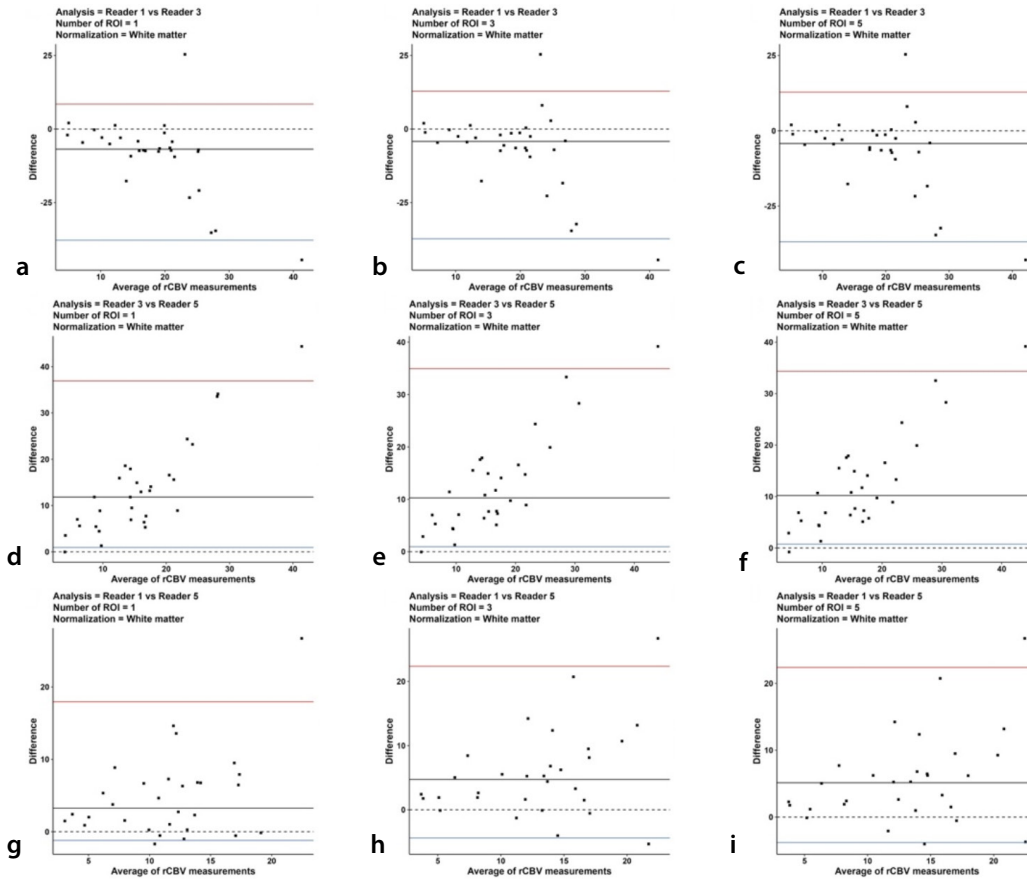
Supplementary Table 2. Intraclass correlation coefficients for the readers with different experience levels

Analysis	Location of ROIs for normalization	Number of ROIs	ICC	LB of 95% CI	UB of 95% CI
R1 vs. R3	Centrum semiovale	One	0.321	0.14	0.53
	Centrum semiovale	Three	0.313	0.15	0.51
	Centrum semiovale	Five	0.335	0.17	0.53
	White matter	One	0.296	0.15	0.49
	White matter	Three	0.297	0.15	0.48
	White matter	Five	0.305	0.16	0.49
R1 vs. R5	Centrum semiovale	One	0.321	0.14	0.53
	Centrum semiovale	Three	0.313	0.15	0.51
	Centrum semiovale	Five	0.335	0.17	0.53
	White matter	One	0.296	0.15	0.49
	White matter	Three	0.297	0.15	0.48
	White matter	Five	0.305	0.16	0.49
R3 vs. R5	Centrum semiovale	One	0.321	0.14	0.53
	Centrum semiovale	Three	0.313	0.15	0.51
	Centrum semiovale	Five	0.335	0.17	0.53
	White matter	One	0.296	0.15	0.49
	White matter	Three	0.297	0.15	0.48
	White matter	Five	0.305	0.16	0.49

ROI, region of interest; ICC, intraclass correlation coefficient; LB, lower bound; UB, upper bound; CI, confidence interval.



Supplementary Figure 1. Bland–Altman plots for readers with similar experience levels (a–c for neuroradiologists, R1 vs. R2; d–f for general radiologists, R3 vs. R4; g–i for residents, R5 vs. R6). Relative cerebral blood volume (rCBV) values are based on white matter normalization. The analysis is non-parametric and relies on the median. Solid black, red, and blue lines represent the medians of difference, the upper level of agreement bound (97.5th percentile), and the lower level of agreement bound (2.5th percentile), respectively. The dashed line stands for no difference. ROI, region of interest.



Supplementary Figure 2. Bland–Altman plots for readers with different experience levels (**a–c** for neuroradiologist vs. general radiologist; **d–f** for general radiologist vs. resident; **g–i** for neuroradiologist vs. resident). Relative cerebral blood volume (rCBV) values are based on white matter normalization. The analysis is non-parametric and relies on the median. Solid black, red, and blue lines represent the medians of difference, the upper level of agreement bound (97.5th percentile), and the lower level of agreement bound (2.5th percentile), respectively. The dashed line stands for no difference. ROI, region of interest.



Focal fatty sparing areas of the pediatric steatotic liver: pseudolesions on hepatobiliary phase magnetic resonance images

Gözde Özer
 H. Nursun Özcan
 Berna Oğuz
 Mithat Haliloğlu

Hacettepe University Faculty of Medicine,
Department of Radiology, Ankara, Türkiye

PURPOSE

Focal fatty sparing in liver can be detected as hyperintense pseudolesions on hepatobiliary phase magnetic resonance imaging (MRI). Distinguishing these pseudolesions from liver lesions may make diagnosis challenging. The aim of this study was to evaluate the imaging features of fatty sparing areas on liver MRI in pediatric patients who have been administered gadoxetate disodium.

METHODS

A total of 63 patients between January 2018 and June 2023 underwent gadoxetate disodium-enhanced liver MRI, and 9 (14%) patients with a focal fatty sparing were included in the study. The fat spared areas were evaluated qualitatively and quantitatively including signal intensity measurements and fat fraction calculations.

RESULTS

The liver MRI examinations of 9 patients (5 boys, 4 girls; aged 8–18 years, median age: 14.4) using gadoxetate disodium were evaluated. Based on in-phase and opposed-phase sequences, 13 areas of focal fatty sparing were identified. The mean fat fraction of the liver and fat spared areas were 26.2% (range, 15–47) and 9% (range, 2–17), respectively. All fat spared areas were hyperintense in the hepatobiliary phase images. The mean relative enhancement ratios of the liver and fat spared areas were 0.78 (range, 0.35–1.6) and 1.11 (range, 0.45–1.9), respectively.

CONCLUSION

Focal fatty sparing in liver in children was observed as hyperintense on hepatobiliary phase MRI, and it should not be identified as a focal liver lesion.

KEYWORDS

Liver, magnetic resonance imaging, gadoxetate disodium, hepatic steatosis, focal fatty sparing, children

Corresponding author: Gözde Özer

E-mail: gozdetufan@gmail.com

Received 19 August 2023; revision requested 20
September 2023; last revision received 22 September
2023; accepted 25 September 2023.



Epub: 30.11.2023

Publication date: 05.03.2024

DOI: 10.4274/dir.2023.232447

Non-alcoholic fatty liver disease has become common among children and includes a broad range of clinicopathologic features ranging from simple steatosis (fat without inflammation and/or fibrosis) and steatohepatitis/non-alcoholic steatohepatitis to cirrhosis.¹ Fatty liver in children can have various imaging manifestations, including diffuse and homogeneous, geographic, focal, and multifocal fat accumulation.² Focal areas of steatosis and fatty sparing in the liver can be detected as mass-like pseudolesions on ultrasonography or computed tomography (CT); in addition, these pseudolesions may show increased fluorodeoxyglucose (FDG) uptake on positron emission tomography (PET)/CT.³ Distinguishing these pseudolesions from metastases, particularly in pediatric patients with cancer, is crucial for preventing misdiagnosis. Liver magnetic resonance imaging (MRI) could be utilized as a problem-solving tool to assess focal liver lesions detected in a steatotic liver in both children and adults. Recently, the use of hepatobiliary contrast agents in children has become more common, and although these pseudolesions can be easily recognized with dual-echo imaging, hepatobiliary phase imaging may cause confusion because of metabolic alterations of liver parenchyma.

You may cite this article as: Özer G, Özcan HN, Oğuz B, Haliloğlu M. Focal fatty sparing areas of the pediatric steatotic liver: pseudolesions on hepatobiliary phase magnetic resonance images. *Diagn Interv Radiol.* 2024;30(2):135-138.

Focal fatty infiltration and fatty sparing in liver are well-known phenomenon in adults; however, in children, because of the low incidence of hepatic steatosis, these pseudolesions may make diagnosis challenging. The purpose of this study was to evaluate signal intensity (SI) features of fat spared areas on liver MRI in pediatric patients who have been administered gadoxetate disodium.

Methods

This retrospective study was approved by the Hacettepe University Non-Interventional Clinical Research Ethics Committee; informed patient consent was waived because the study was based on retrospective data analysis (GO 21/1162). The archive of the pediatric radiology unit was retrospectively reviewed for liver MRI examinations performed in our institution between January 2018 and June 2023. A total of 63 patients with indications of focal liver lesion, primary liver tumor, metastasis, and chronic liver disease underwent liver MRI with gadoxetate disodium administration. Patients with chronic parenchymal liver disease were excluded, and 9 patients with fat spared areas were included in the study. The MRI examinations were evaluated by two pediatric radiologists (H.N.O. and G.O.) with 11 and 2 years of experience, respectively, through consensus, using a picture archiving and communication system (PACS; GE Medical Systems, Milwaukee, WI, USA). The following clinical and radiological features were recorded: primary diagnosis, patients' age at the time of MRI, and fat spared areas and true lesions in the liver parenchyma on MRI. The fat spared areas were evaluated qualitatively and quantitatively. For the quantitative assessment of the fat fraction, a region of interest with an average size of 0.5 cm² was placed at the fat

spared areas and steatotic liver parenchyma on in-phase and opposed-phase images for SI measurements. The fat fraction was calculated according to the following formula: fat fraction = in-phase SI – opposed-phase SI / 2 × in-phase SI. The delta fat fraction was defined as the difference between the liver parenchyma and fat spared area. In addition, an SI measurement was also performed on precontrast fat-suppressed T1-weighted and hepatobiliary phase images at the liver parenchyma and fat spared areas. The relative enhancement ratio in the hepatobiliary phase images was calculated in both the liver parenchyma and fat spared areas using the following formula: (hepatobiliary phase SI) – (precontrast SI) / (precontrast SI).

The MRI examinations were performed using 1.5T MRI system (GE Signa HDx Healthcare, Milwaukee, WI, USA) units with an eight-channel phased-array body coil. The imaging protocol of the liver included breath-hold coronal TRUE-FISP [repetition time (TR), 4.3 ms; time to echo (TE), 2.1 ms; flip angle (FA), 60; matrix, 416 × 512; slice thickness, 4.5 mm], axial T2-weighted half-Fourier acquisition single-shot turbo spin-echo (TR, 1350 ms; TE, 92 ms; FA, 160; matrix, 256 × 256; slice thickness, 6 mm), axial in-phase and opposed-phase chemical shift imaging (TR, 160 ms; TE, in-phase: 4.9 ms, opposed-phase: 2.4 ms; FA, 70; matrix, 256 × 192; slice thickness, 6 mm), breath-hold T2-weighted fast spin-echo with fat suppression (TR, 3050 ms; TE, 125 ms; FA, 150; matrix, 256 × 256, slice thickness, 6 mm), and three-dimensional T1-weighted gradient-recalled echo fat-suppressed sequences (TR, 5 ms; TE, 2.4 ms; FA, 10; matrix, 320 × 240; slice thickness, 3 mm) before and after the injection of the contrast agent. A bolus injection of gadoxetate disodium (Primovist, Bayer HealthCare, Berlin, Germany) was administered at a rate of 1 mL/s. The total contrast dose was 0.1 mL/kg of body weight. Diffusion-weighted imaging

was used to acquire single-shot echo-planar images (under free-breathing) with *b* values of 50, 400, and 800 s/mm². The images were acquired in accordance with delayed hepatobiliary phase imaging at 20 min for gadoxetate disodium. Before the MRI examination, an informed consent form was obtained from the patients' parents regarding the use of gadoxetate disodium. Gadoxetate disodium is a widely used contrast agent in children and has been reported as safe in the literature.⁴

Results

Liver MRI examinations of 9 patients (5 boys, 4 girls; aged 8–18 years, median age: 14.4) using gadoxetate disodium were evaluated. The demographic characteristics are summarized in Table 1. None of the patients included in the study had liver cirrhosis.

A total of 13 focal fat spared areas were detected on in-phase and opposed-phase images (Table 2). On the opposed-phase images, the fat spared areas had high SI. The mean fat fraction of the liver and fat spared areas were 26.2% (range, 15–47) and 9% (range, 2–17), respectively. The median delta fat fraction was 15% (range, 12–34). The fat spared areas were hyperintense in 7 (78%) patients and isointense in 2 (22%) patients on fat-suppressed precontrast T1-weighted images (Figure 1). The mean SI of liver and fat spared areas on precontrast fat-suppressed T1-weighted images were 405 (range, 129–891) and 481 (range, 130–1277). The mean SI of liver and fat spared areas on hepatobiliary phase images were 736 (range, 175–1693) and 1112 (range, 203–2834). In the fat-suppressed T2-weighted images, the fat spared areas were hypointense in 5 (55%) patients and isointense in 4 (45%) patients. There was no signal alteration in any of the patients on the diffusion-weighted images. All the detected focal fat spared areas were hyperintense in the hepatobiliary phase images (Figure 1).

Main points

- Fatty liver disease has become more common in children in recent years.
- Focal fatty sparing can be detected as mass-like lesions on ultrasonography or computed tomography (CT) and may even show increased fluorodeoxyglucose uptake in positron emission tomography/CT.
- Liver magnetic resonance imaging with hepatobiliary contrast agents can be used as a problem-solving imaging modality in the evaluation of steatotic liver in children.
- Focal fat spared areas in the liver parenchyma may appear as increased signal intensity in the hepatobiliary phase, presumably because of the preserved parenchymal function.

Table 1. Demographic characteristics of the patients			
Patient no	Age (years)	Sex	Primary diagnoses
1	13	M	Obesity, geographic liver lesion on abdominal ultrasound
2	11	F	Hodgkin lymphoma
3	18	M	Glycogen storage disease
4	16	F	Diabetes mellitus, PCOS, geographic liver lesion on abdominal ultrasound
5	8	F	Hypertriglyceridemia, liver lesion on abdominal ultrasound
6	17	M	Testicular yolk sac tumor
7	17	M	Hodgkin lymphoma
8	13	F	Glycogen storage disease
9	17	M	Hepatic adenoma
M, male; F, female; PCOS, polycystic ovary syndrome.			

Table 2. Imaging findings of the patients

Patient no	Segments of FSAs	T2W fat-suppressed	T1W fat-suppressed	Arterial phase	Portal phase	Delayed phase	Hepatobiliary phase
1	Segment 7 and 8	Hypointense	Hyperintense	Hyperintense	Hyperintense	Hyperintense	Hyperintense
2	Segment 4 and 5	Isointense	Hyperintense	Hyperintense	Hyperintense	Hyperintense	Hyperintense
3	Segment 2 and 4	Hypointense	Hyperintense	Hyperintense	Hyperintense	Hyperintense	Hyperintense
4	Segment 3 and 4	Hypointense	Hyperintense	Hyperintense	Hyperintense	Hyperintense	Hyperintense
5	Segment 4	Isointense	Hyperintense	Hyperintense	Hyperintense	Hyperintense	Hyperintense
6	Segment 1 and 4	Hypointense	Hyperintense	Hyperintense	Hyperintense	Hyperintense	Hyperintense
7	Segment 4	Isointense	Isointense	Isointense	Hyperintense	Hyperintense	Hyperintense
8	Segment 3	Isointense	Isointense	Isointense	Hyperintense	Hyperintense	Hyperintense
9	Segment 4	Hypointense	Hyperintense	Hyperintense	Hyperintense	Hyperintense	Hyperintense

FSAs, fat spared areas.

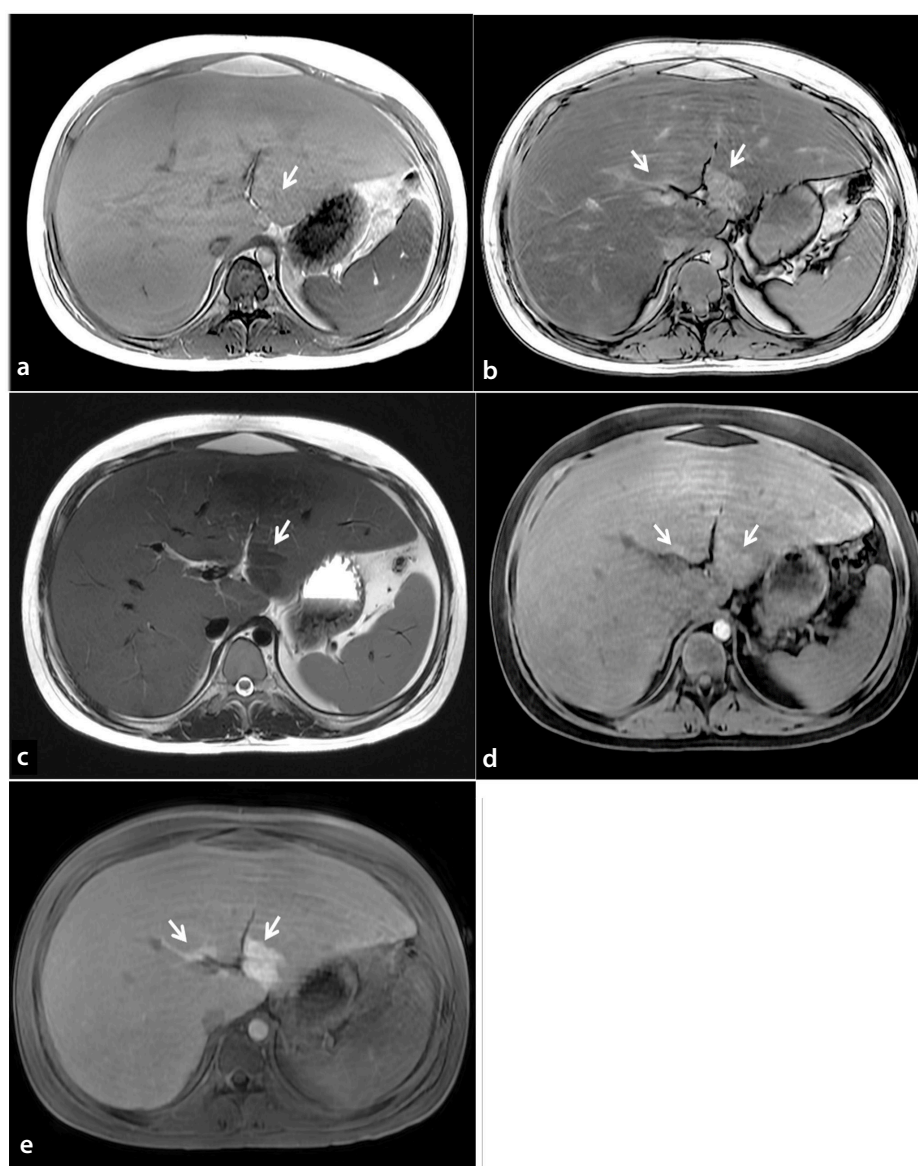


Figure 1. An 18-year-old boy with type 1 glycogen storage disease underwent liver magnetic resonance imaging using a hepatobiliary contrast agent. (a, b) In-phase (a) and opposed-phase (b) images indicating liver steatosis with decreased signal intensity on the opposed-phase image and fat spared areas in segments 2 and 4 (arrows). (c, d) Axial T2-weighted image indicating hypointensity (arrow), and precontrast fat-suppressed T1-weighted image displaying hyperintensity (arrow) at the focal fat spared area. (e) Hyperintensity at the focal fat spared area (arrows) on the hepatobiliary phase image 20 min after gadoxetate disodium injection.

The mean relative enhancement ratios of the liver and fat spared areas were 0.78 (range, 0.35–1.6) and 1.11 (range, 0.45–1.9), respectively. Fat spared areas were present at segment 1 ($n = 1$), segment 2 ($n = 1$), segment 3 ($n = 2$), segment 4 ($n = 7$), segment 5 ($n = 1$), segment 7 ($n = 1$), and segment 8 ($n = 1$). Five patients had a fat spared area in more than one liver segment.

Focal nodular hyperplasia was detected in 5 patients, and all of these lesions exhibited gadoxetate disodium retention in the hepatobiliary phase. One patient had histopathologically confirmed inflammatory hepatocellular adenoma that displayed wash-out on hepatobiliary phase images.

Discussion

This study produced two major results. First, focal fatty sparing in the pediatric steatotic liver demonstrates increased SI on hepatobiliary phase images. Second, we observed that most of these areas have increased SI on precontrast fat-suppressed T1-weighted images. On in-phase and opposed-phase images, fatty sparing has high SI on the opposed-phase images as a result of the suppressed signal of the other parts of the steatotic liver.

In our study, most of the fat spared areas were in segment 4. Some segments of the liver, such as the gallbladder fossa, medial segment of the left lobe adjacent to the portal vein, and subcapsular areas are more prone to focal fatty sparing.⁵ This phenomenon is caused by a third inflow, which is a venous inflow to the liver in addition to the typical dual blood supply (portal vein and hepatic artery). The most common anatomic variations that cause a third inflow are an aberrant right gastric vein, epigastric and para-umbilical veins (Sappey's and Burow's veins), and cholecystic veins.^{6,7} Focal fat spared ar-

eas are detected as focal hypoechoic areas on ultrasonography and hyperdense areas on CT, and these findings may be confused with solid liver lesions. MRI is considered the most reliable non-invasive diagnostic tool for evaluating hepatic steatosis.⁸ The dual-echo method can easily detect focal fat deposition or fatty sparing.

Hepatic steatosis leads to parenchymal inflammation and fibrosis and may cause decreased parenchymal function.^{9,10} Gadoxetate disodium-enhanced liver MRI can be used to evaluate the functional capacity of the liver parenchyma, and decreased enhancement on hepatobiliary phase images might be a sign of hepatocyte dysfunction caused by liver fibrosis and inflammation.^{11,12} Impaired hepatic function can be observed as decreased enhancement on hepatobiliary phase images.¹⁰ Therefore, focal fatty sparing can be observed as hyperintense pseudolesions on hepatobiliary phase images, presumably because of the preserved hepatocyte function. Ünal et al.¹³ reported similar findings in adult patients and suggested that fat spared areas demonstrating hyperintensity on hepatobiliary phase images may include hyperfunctioning hepatocytes compared with other parts of the liver. Fat spared areas of the liver may appear as focal areas of increased FDG uptake in FDG PET/CT.¹⁴⁻¹⁶ In addition, focal fat spared areas have been reported to mimic neuroendocrine tumor metastases in ⁶⁸Ga-Dotatate PET/CT.¹⁷ These reports may support the hypothesis of preserved or maybe even increased hepatocyte function in these areas.

In our study, 5 patients had focal nodular hyperplasia, which is an uncommon lesion in the pediatric population¹⁸ and has hyperintensity on hepatobiliary phase images; however, early arterial phase enhancement, persistent enhancement on delayed phases, and isointense to hyperintense signals on T2-weighted images might allow the differentiation of these lesions from fat spared areas.^{19,20} In addition, some subtypes of hepatocellular adenomas are observed as hyperintense on hepatobiliary phase images. In our study, 1 patient had an inflammatory hepatocellular adenoma that exhibited wash-out of the contrast media at 20 min. Inflammatory adenoma is the most common subtype related to oral contraceptives and obesity and may show contrast retention on hepatobiliary phase images.²¹ The fact that fatty liver is more common in these patients may make diagnosis challenging in liver MRI. Fatty sparing can be distinguished by the

strong enhancement of adenomas in the arterial phase.¹⁹

This study has several limitations. Our study group was small, and no histopathological correlation was identified in any patients. Biopsy was considered unnecessary because of the typical MRI findings and the benign nature of focal fatty sparing. The hyperintensity on hepatobiliary phase images may be secondary to precontrast T1 hyperintensity or hyperfunctional hepatocytes in fatty sparing areas.

In conclusion, liver steatosis may have various imaging manifestations in pediatric patients. Focal fat spared areas in children have been observed as hyperintense on hepatobiliary phase MRI, and they should not be identified as a focal liver lesion.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

- Schwimmer JB, Deutsch R, Kahen T, et al. Prevalence of fatty liver in children and adolescents. *Pediatrics*. 2006;118(4):1388-1393. [\[CrossRef\]](#)
- Özcan HN, Oğuz B, Haliloğlu M, Orhan D, Karçaaltıncaba M. Imaging patterns of fatty liver in pediatric patients. *Diagn Interv Radiol*. 2015;21(4):355-360. [\[CrossRef\]](#)
- Rydzak C, Chauhan A, Gupta N, Chuang HH, Rohren EM, Bhosale PR. Fat-containing hypermetabolic masses on FDG PET/CT: a spectrum of benign and malignant conditions. *AJR Am J Roentgenol*. 2016;207(5):1095-1104. [\[CrossRef\]](#)
- Ayyala RS, Anupindi SA, Gee MS, Trout AT, Callahan MJ. Intravenous gadolinium-based hepatocyte-specific contrast agents (HSCAs) for contrast-enhanced liver magnetic resonance imaging in pediatric patients: what the radiologist should know. *Pediatr Radiol*. 2019;49(10):1256-1268. [\[CrossRef\]](#)
- Tom WW, Yeh BM, Cheng JC, Qayyum A, Joe B, Coakley FV. Hepatic pseudotumor due to nodular fatty sparing: the diagnostic role of opposed-phase MRI. *AJR Am J Roentgenol*. 2004;183(3):721-724. [\[CrossRef\]](#)
- Yoshimitsu K, Honda H, Kuroiwa T, et al. Unusual hemodynamics and pseudolesions of the noncirrhotic liver at CT. *Radiographics*. 2001;21:81-96. [\[CrossRef\]](#)
- Hamer OW, Aguirre DA, Casola G, et al. Fatty liver: imaging patterns and pitfalls. *Radiographics*. 2006;26(6):1637-1653. [\[CrossRef\]](#)
- Mennesson N, Dumortier J, Hervieu V, et al. Liver steatosis quantification using magnetic resonance imaging: a prospective

- comparative study with liver biopsy. *J Comput Assist Tomogr*. 2009;33(5):672-677. [\[CrossRef\]](#)
- Poetter-Lang S, Bastati N, Messner A, et al. Quantification of liver function using gadoxetic acid-enhanced MRI. *Abdom Radiol (NY)*. 2020;45(11):3532-3544. [\[CrossRef\]](#)
- Wu Z, Matsui O, Kitao A, et al. Usefulness of Gd-EOB-DTPA-enhanced MR imaging in the evaluation of simple steatosis and nonalcoholic steatohepatitis. *J Magn Reson Imaging*. 2013;37(5):1137-1143. [\[CrossRef\]](#)
- Hojreh A, Lischka J, Tamandl D, et al. Relative enhancement in gadoxetate disodium-enhanced liver MRI as an imaging biomarker in the diagnosis of non-alcoholic fatty liver disease in pediatric obesity. *Nutrients*. 2023;15(3):558. [\[CrossRef\]](#)
- Verloh N, Probst U, Utpatel K, et al. Influence of hepatic fibrosis and inflammation: correlation between histopathological changes and Gd-EOB-DTPA-enhanced MR imaging. *PLoS One*. 2019;14(5):e0215752. [\[CrossRef\]](#)
- Ünal E, İdilman İS, Karaosmanoğlu AD, et al. Hyperintensity at fat spared area in steatotic liver on the hepatobiliary phase MRI. *Diagn Interv Radiol*. 2019;25(6):416-420. [\[CrossRef\]](#)
- Nguyen BD, Heller MT, Roarke MC. Nodular fat-sparing hepatic parenchyma as 11C-choline-avid finding on PET/CT. *Clin Nucl Med*. 2020;45(3):228-229. [\[CrossRef\]](#)
- Purandare NC, Rangarajan V, Rajnish A, Shah S, Arora A, Pathak S. Focal fat spared area in the liver masquerading as hepatic metastasis on F-18 FDG PET imaging. *Clin Nucl Med*. 2008;33(11):802-805. [\[CrossRef\]](#)
- Harisankar CN. Focal fat sparing of the liver: a nonmalignant cause of focal FDG uptake on FDG PET/CT. *Clin Nucl Med*. 2014;39(7):359-361. [\[CrossRef\]](#)
- Hod N, Levin D, Anconina R, et al. 68Ga-DOTATATE PET/CT in focal fatty sparing of the liver. *Clin Nucl Med*. 2019;44(10):815-817. [\[CrossRef\]](#)
- Smith EA, Salisbury S, Martin R, Towbin AJ. Incidence and etiology of new liver lesions in pediatric patients previously treated for malignancy. *AJR Am J Roentgenol*. 2012;199(1):186-191. [\[CrossRef\]](#)
- Vasireddi AK, Leo ME, Squires JH. Magnetic resonance imaging of pediatric liver tumors. *Pediatr Radiol*. 2022;52(2):177-188. [\[CrossRef\]](#)
- Özcan HN, Karçaaltıncaba M, Seber T, et al. Hepatocyte-specific contrast-enhanced MRI findings of focal nodular hyperplasia-like nodules in the liver following chemotherapy in pediatric cancer patients. *Diagn Interv Radiol*. 2020;26(4):370-376. [\[CrossRef\]](#)
- Schooler GR, Hull NC, Lee EY. Hepatobiliary MRI contrast agents: pattern recognition approach to pediatric focal hepatic lesions. *AJR Am J Roentgenol*. 2020;214(5):976-986. [\[CrossRef\]](#)