



# Inter-reader reliability of functional liver imaging score derived from gadoxetic acid-enhanced MRI: a meta-analysis

Nam Hee Kim<sup>1</sup> · Ji Hun Kang<sup>2</sup>

Received: 9 September 2022 / Revised: 18 September 2022 / Accepted: 16 December 2022 / Published online: 28 December 2022  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

**Purpose** This study aimed to systematically determine the inter-reader reliability of the functional liver imaging score (FLIS) and explore the factors affecting it.

**Methods** Original articles reporting the inter-reader reliability of FLIS derived from gadoxetic acid-enhanced magnetic resonance imaging (MRI) were systematically searched in the MEDLINE and EMBASE databases from January 2013 to June 2022. Data synthesis was performed to calculate the meta-analytic pooled estimates of the FLIS and its three subcategories, including enhancement quality score (EnQS), excretion quality score (ExQS), and portal vein sign quality score (PVsQS) using the DerSimonian-Laird random-effects model. To explore any cause of study heterogeneity, we conducted a meta-regression analysis.

**Results** Six studies with data from 1419 patients were included. The meta-analytic pooled inter-reader reliability of FLIS was 0.93 (95% confidence interval [CI], 0.88–0.98). That of the three FLIS subcategories were 0.93 (95% CI, 0.85–1.00), 0.95 (95% CI, 0.91–1.00), and 0.90 (95% CI, 0.81–0.99) for EnQS, ExQS, and PVsQS, respectively. The pooled FLIS data was moderately heterogenous, but heterogeneity was not associated with the study methodology, MRI-related factors, and reader experience.

**Conclusion** The FLIS and its three subcategories showed almost perfect inter-reader reliability. Therefore, FLIS may be a reliable imaging parameter that reflects liver function and outcomes in patients with chronic liver disease. Further studies should be conducted to confirm any factors affecting the inter-reader reliability of FLIS.

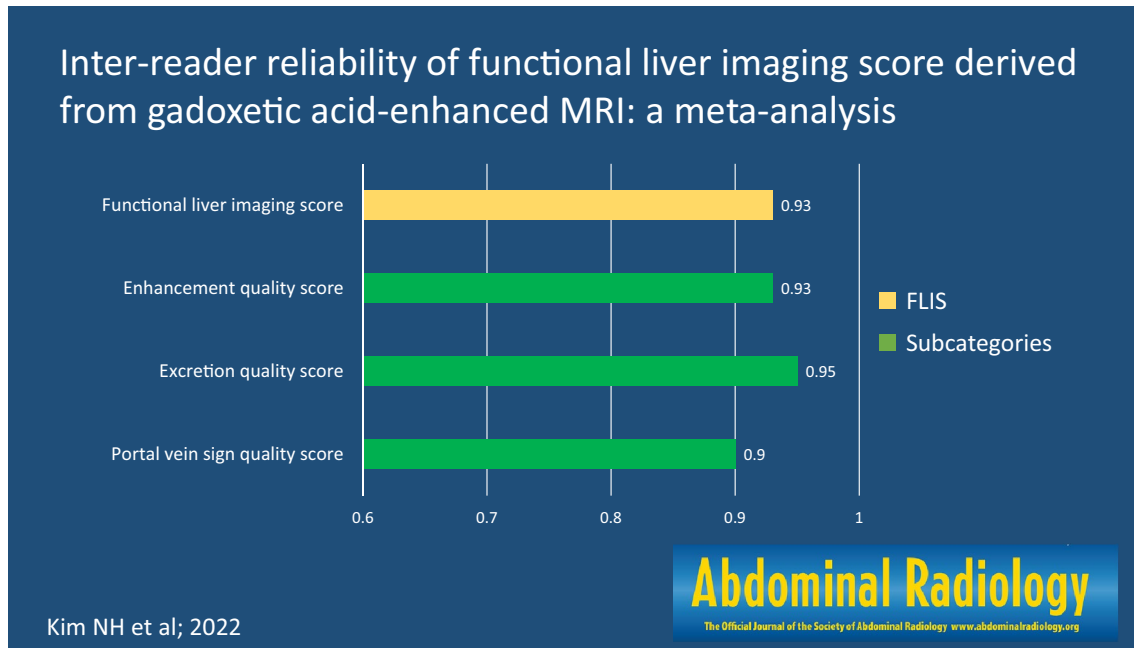
---

✉ Ji Hun Kang  
jihunkang@hanyang.ac.kr

<sup>1</sup> Division of Gastroenterology, Department of Internal Medicine, Sungkyunkwan University School of Medicine, Kangbuk Samsung Hospital, Seoul, Korea

<sup>2</sup> Department of Radiology, Hanyang University College of Medicine, Hanyang University Guri Hospital, 153 Gyeongchun-Ro, Guri-Si 11923, Gyeonggi-Do, Korea

## Graphical abstract



**Keywords** Liver · Gadolinium ethoxybenzyl DTPA · Magnetic resonance imaging · Reproducibility of results · Meta-analysis

### Abbreviations

MRI	Magnetic resonance imaging
FLIS	Functional liver imaging score
EnQS	Enhancement quality score
ExQS	Excretion quality score
PVsQS	Portal vein sign quality score
ICC	Intraclass correlation coefficient
$\kappa$	Kappa value
CI	Confidence interval

### Introduction

Gadoxetic acid-enhanced magnetic resonance imaging (MRI) has been widely used in patients with chronic liver disease and liver cirrhosis [1]. It is taken up by organic anion transporters into normal hepatocytes during the transitional and hepatobiliary phases and has unique characteristics that enable the detection of focal hepatic lesions and the assessment of hepatic function and chronic liver disease severity [2, 3]. Previous studies have introduced several methods to evaluate hepatobiliary phase uptake of gadoxetic acid as a noninvasive surrogate parameter for hepatic function. These studies focus on parameters such as relative liver enhancement, hepatic uptake index, contrast uptake index,

liver-to-spleen contrast index, and T1 values [2, 4]. Although these methods demonstrate results quantitatively, they are time-consuming and depend on the vendor, magnetic field strength, and imaging sequence, making clinical application difficult.

Bastati et al. developed the functional liver imaging score (FLIS), a scoring system to evaluate liver function based on qualitative MRI features [5]. FLIS is the sum of three simple visual features evaluated the hepatobiliary phase of gadoxetic acid-enhanced MRI: enhancement quality score (EnQS), excretion quality score (ExQS), and portal vein sign quality score (PVsQS) [5]. This semi-quantitative scoring system makes it easier to evaluate hepatic function than other quantitative methods because it doesn't need to measure signal intensity, calculate complex equations, or use dedicated software [6, 7]. FLIS was associated with probability of graft survival in liver transplant recipients, and also associated with first hepatic decompensation and mortality in advanced chronic liver disease patients [5, 8]. In addition, a newly suggested algorithm based on combination of FLIS and splenic diameter measured using MRI could stratify the risk of mortality in patients with advanced chronic liver disease [9]. Taken together, these results indicate that FLIS is a promising imaging biomarker.

High reproducibility is essential for reliable imaging biomarkers [10]. Previous studies have reported almost perfect inter-reader reliability of FLIS [5, 8, 9, 11–14]. However, the inter-reader reliability of FLIS was an ancillary finding in each study, and it has not been systematically determined. Nonetheless, FLIS is not expected to be affected by other factors, such as reader and MRI-related factors. Therefore, the purpose of this study was to systematically determine the inter-reader reliability of the FLIS and explore possible factors that affect it.

## Materials and methods

This study was conducted and reported following the guidelines for Meta-analysis of Observational Studies in Epidemiology [15] and Preferred Reporting Items for Systematic Reviews and Meta-Analyses [16, 17].

### Literature search

Original research articles reporting the inter-reader reliability of FLIS derived from gadoxetic acid-enhanced MRI were systematically searched in the MEDLINE and EMBASE databases. The representative terms used for the sensitive literature search were “functional liver imaging score,” “gadoxetic acid,” and “MRI,” and the detailed search query is presented in Supplementary Table 1. The literature search was limited to original studies on human subjects that were published in English. The search period began with studies published in January 2013 and was updated until June 2022. The bibliographies of the identified studies were reviewed to include additional eligible studies.

### Eligibility criteria

Studies were included if the following criteria were met: (a) Population: patients who underwent gadoxetic acid-enhanced MRI for the evaluation of the hepatobiliary system or liver graft; (b) Index test: gadoxetic acid-enhanced MRI that included 20-min delayed hepatobiliary phase imaging; (c) Comparator: no requirements; (d) Outcome: inter-reader reliability of FLIS; and (e) Study design: any type of study including observational studies and clinical trials. Studies were excluded if they met the following criteria: (a) review articles, conference abstracts, letters, and editorials; (b) studies in which patient cohorts and data overlapped; (c) studies unrelated to the field of interest of

this study; and (d) studies that did not provide sufficient data to determine inter-reader reliability. The titles and abstracts of potentially eligible studies were reviewed based on eligibility criteria before conducting full-text reviews of the remaining studies.

### Data extraction

The following data were extracted from the final studies included using a predefined form: (a) study characteristics: author, year of publication, study design, study type, subject enrollment method, and country in which the study was performed; (b) demographic and clinical data: number of patients, patient age, and underlying hepatobiliary disease; (c) MRI data: vendor, type of scanner, magnet field strength (1.5-T or 3.0-T), and type of contrast agent; (d) image interpretation data: number of readers, reader experience, and clarity of blindness to reference standard during the review; and (e) study outcomes: inter-reader reliability of FLIS and its three subcategories including EnQS, ExQS, and PVsQS. The intraclass correlation coefficient (ICC) or kappa value ( $\kappa$ ) with standard error was extracted to calculate the meta-analytic estimation of inter-reader reliability. Two reviewers independently performed data extraction, and cases of disagreement were resolved at a consensus meeting.

### Quality evaluation

The quality of the eligible studies was evaluated according to the Guidelines for Reporting Reliability and Agreement studies [18]. Risk of bias in the following seven domains was evaluated: (a) index test; (b) study subjects; (c) readers; (d) reading process; (e) clarity of blinding during the review; (f) statistical analysis; and (g) the actual number of subjects. Details of the questionnaires for each domain are described in Supplementary Table 2. Each category was rated as high-quality when the study detailed measures to limit potential bias. Two reviewers independently performed the study quality evaluation, and disagreements were resolved at a consensus meeting.

### Data synthesis and statistical analysis

R version 4.2.1, with the *meta* and *metafor* packages, was used to perform analyses. ICC or  $\kappa$  with standard error was summarized for the FLIS and its three subcategories (EnQS, ExQS, and PVsQS) from each study to calculate meta-analytic pooled estimates. When an original study did not report a standard error, it was estimated using the 95% confidence interval (CI). If a study only reported the inter-reader reliability of FLIS subcategories

without reporting that of FLIS itself, the median of the reported subcategories was considered FLIS. Meta-analytic pooled estimates with 95% CI were calculated using the DerSimonian-Laird random-effects model with or without the Knapp and Hartung adjustment [19]. The meta-analytic pooled estimates were categorized based on Landis and Koch as follows: < 0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, almost perfect reliability [20]. Heterogeneity was evaluated using the Cochran Q-test and  $I^2$  statistics as follows: < 25%, low heterogeneity; 25–75%, moderate heterogeneity; and > 75%, high heterogeneity [21]. Publication bias was assessed using funnel plots and rank tests.

Meta-regression analyses were performed to explore the causes of study heterogeneity according to the following covariates: (a) subject enrollment (consecutive vs. selective), (b) country of study (western vs. eastern), (c) MRI magnet field strength (3.0-T vs. 1.5-T included), (d) MRI vendor (single vs. multiple), (e) MRI scanner (single vs. multiple), (f) number of readers (two readers vs. more than two readers), (g) difference in reader experience (all experienced readers in abdominal imaging vs. multiple readers with trainees), (h) average reader experience ( $\geq 9.6$  years of experience in abdominal imaging vs. < 9.6 years, according to the mean 9.6-year reader experience of the included studies), and (i) homogeneity in reader experience (i.e., a difference in reader experience of less than 3 years vs. other).

## Results

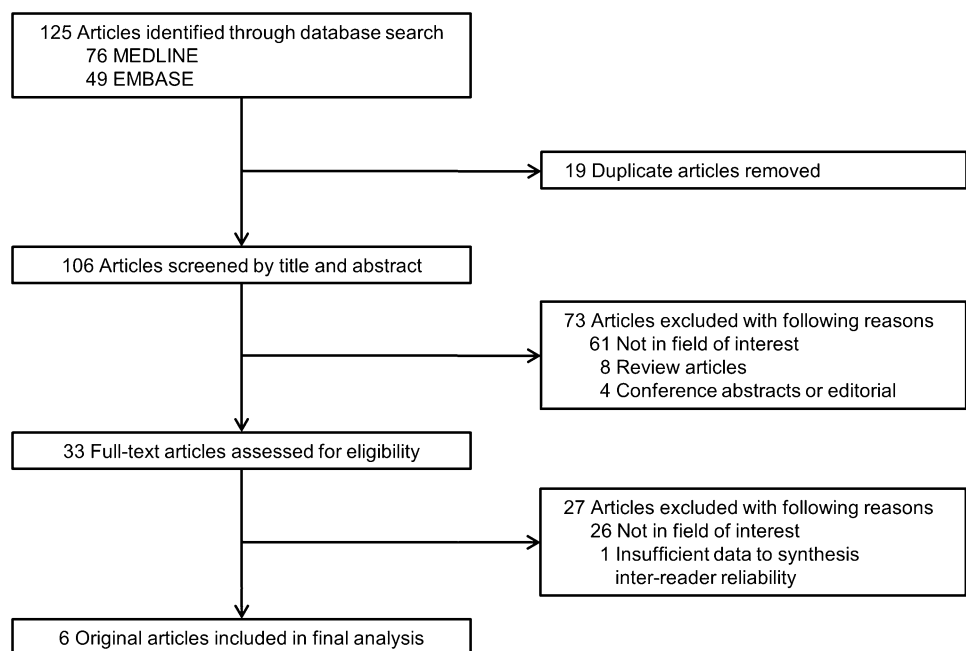
### Literature search

Initially, 125 studies were identified through a systematic literature search. After removing 19 duplicate articles, 73 were excluded upon reviewing their titles and abstracts during the screening. Subsequently, 27 articles were further excluded after full-text review. Finally, six original studies with a total of 1419 patient data were included in this study [5, 8, 11–14]. The detailed study selection process is shown in Fig. 1.

### Characteristics of the included studies

The detailed characteristics of the included studies are summarized in Table 1. All the included studies were retrospective cohort studies [5, 8, 11–14]. Four studies enrolled subjects consecutively [5, 8, 11, 13], and two studies selectively enrolled subjects who underwent surgery or biopsy [12, 14]. Three studies were performed in Western countries [5, 8, 11] and three in Eastern countries [12–14]. Four studies used a single MRI machine [5, 8, 11, 12], and two studies used two different machines [13, 14]. All included studies used a standard dose of gadoteric acid (0.025 mmol/kg; Primovist/Eovist, Bayer) as the contrast agent [5, 8, 11–14]. Four studies used two readers [5, 12–14], while the remaining two used more than two readers [8, 11]. All included studies supplied information about the details of each reader's experience [5, 8, 11–14]. The experience level of each reader

**Fig. 1** Flow diagram of the study selection process



**Table 1** Characteristics of the included studies

Article	Study design and enrollment type	Subject enrollment	Duration	Country	No. of patients (male)	Age, y <sup>a</sup>	Hepato-biliary disease	Reference standard	MRI magnet, Tesla	MRI vendor	MRI machine	Contrast media	No. of readers (Years of experience)	Blinding	Analyzed features
[5]	Retrospective cohort	Consecutive	2007–2015	Austria	128 (80)	54.6 (46–64)	Liver graft	Histology	3.0	Siemens	Magnetom Trio	Gadoxetic acid	2 (6, 20)	Yes	FLIS, EnQS, ExQS, PVsQS
[8]	Retrospective cohort	Consecutive	2011–2015	Austria	265 (164)	53 (14)	CLD	CCRS	3.0	Siemens	Magnetom Trio	Gadoxetic acid	3 (8, 20, trainee)	Yes	FLIS
[13]	Retrospective cohort	Consecutive	2015–2018	Korea	134 (88)	58.8 (29–80)	LC or CLD	CCRS	1.5 or 3.0	Siemens	Magnetom Avanto or Magnetom Skyra	Gadoxetic acid	2 (8, trainee)	Yes	FLIS, EnQS, ExQS, PVsQS
[11]	Retrospective cohort	Consecutive	2019–2022	Türkiye	131 (70)	53.7 (14.6)	LC or CLD	CCRS	1.5	Siemens	Magnetom Symphony	Gadoxetic acid	3 (10, trainee)	Yes	FLIS
[12]	Retrospective cohort	Selective	2011–2017	Korea	259 (162)	68.0 (8.3)	EHD cancer	Surgery or biopsy	3.0	Philips	Achieva	Gadoxetic acid	2 (11, 12)	Yes	FLIS, EnQS, ExQS, PVsQS
[14]	Retrospective cohort	Selective	2017–2019	China	502 (422)	51 (25–76)	HCC	Surgery	1.5 or 3.0	Siemens or GE	Magnetom Avanto or Discovery MR 750	Gadoxetic acid	2 (8, 20)	Yes	FLIS

Articles are listed alphabetically according to the year of publication and names of the first authors

<sup>a</sup>Data are mean values with standard deviations or ranges in parentheses, or medians with interquartile ranges in parentheses

*FLIS* functional liver imaging score; *EnQS* enhancement quality score; *ExQS* excretion quality score; *PVsQS* portal vein sign quality score; *CLD* chronic liver disease; *CCRS* composite clinical reference standard; *LC* liver cirrhosis; *EHD* extrahepatic duct; *HCC* hepatocellular carcinoma

**Table 2** Summary of the meta-analysis

Analyzed factors	No. of studies	Estimate (95% CI)	Higgins $I^2$ , %
Functional liver imaging score	6	0.93 (0.88–0.98)	73.2
Enhancement quality score	3	0.93 (0.85–1.00)	85.2
Excretion quality score	3	0.95 (0.91–1.00)	63.0
Portal vein sign quality score	3	0.90 (0.81–0.99)	78.5

Results were obtained using the DerSimonian-Laird random-effects model with or without the Knapp and Hartung adjustment  
*CI* confidence interval

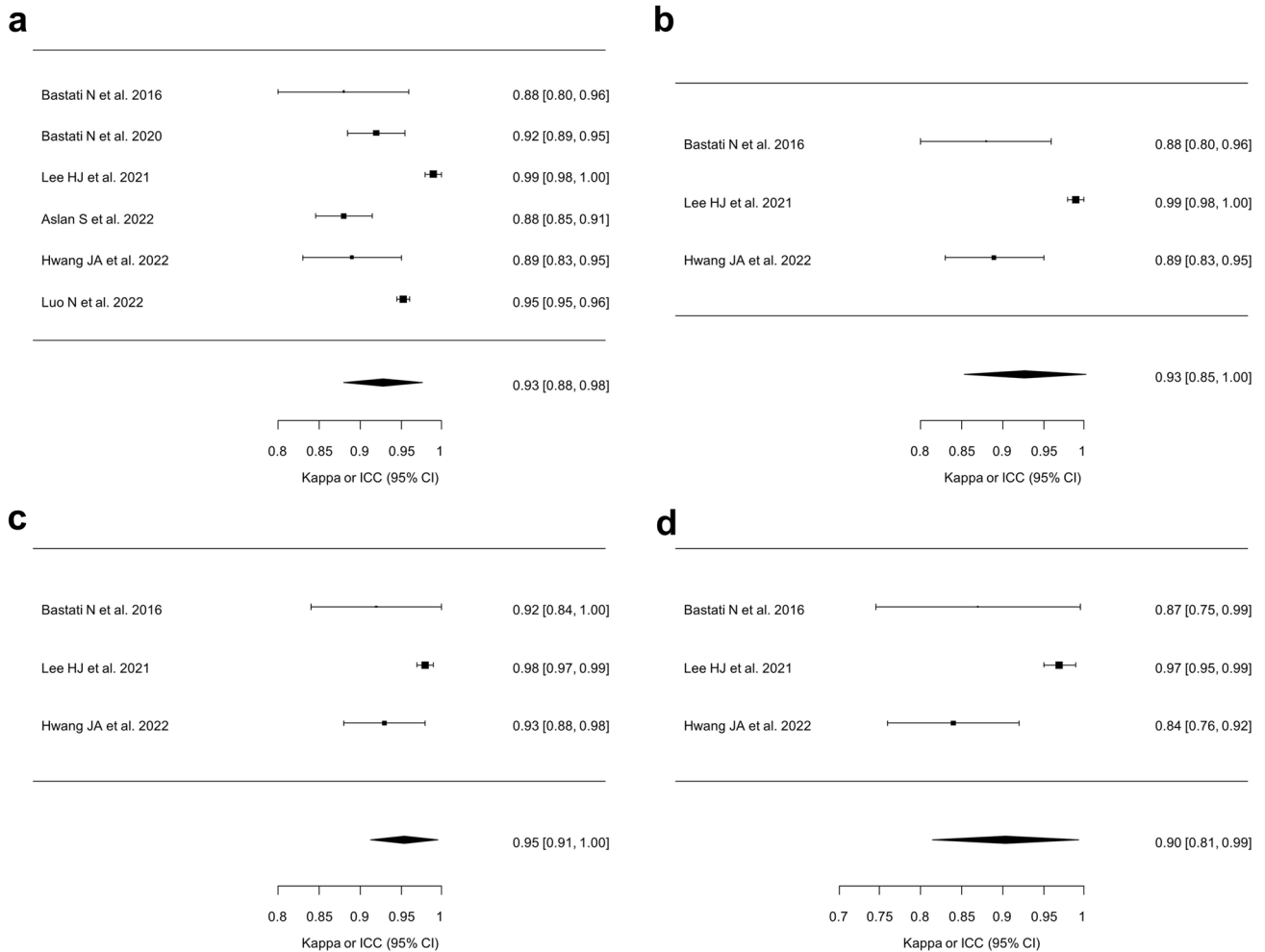
varied, ranging from trainee to 20 years of experience in abdominal imaging, with an average of 9.6 years. The readers in all included studies were blinded to the clinical information [5, 8, 11–14].

**Study quality**

All included studies demonstrated a quality score of five or more for the seven domains evaluated (Supplementary Table 2).

**Meta-analytic pooled inter-reader reliability of functional liver imaging score**

The meta-analytic pooled estimates of the inter-reader reliability of FLIS derived from gadoteric acid-enhanced MRI are summarized in Table 2 and Fig. 2. The meta-analytic pooled inter-reader reliability of FLIS was 0.93 (95% CI, 0.88–0.98), showing almost perfect inter-reader reliability. In addition, the meta-analytic pooled inter-reader reliability of the three FLIS subcategories was as follows; 0.93 (95% CI, 0.85–1.00) for EnQS, 0.95 (95% CI, 0.91–1.00)



**Fig. 2** Meta-analytic pooled inter-reader reliability. **a** Functional liver imaging score, **b** enhancement quality score, **c** excretion quality score, and **d** portal vein sign quality score

for ExQS, and 0.90 (95% CI, 0.81–0.99) for PVsQS, also showing almost perfect inter-reader reliability.

### Meta-regression analysis

The meta-analytic pooled inter-reader reliability of FLIS showed moderate study heterogeneity, which did not reach high study heterogeneity ( $I^2 = 73.2$ ). According to the meta-regression analysis, the subject enrollment method, MRI-related factors (vendor, type of scanner, and magnetic field strength), number of readers, difference in reader experience, average reader experience, and homogeneity of reader experience were not significantly associated with study heterogeneity (See Table 3).

There was no significant publication bias regarding the inter-reader reliability of FLIS and its three subcategories ( $p > 0.44$ , Supplementary Fig. 1).

### Discussion

This study demonstrated that FLIS and its three subcategories, enhancement quality score, excretion quality score, and portal vein sign quality score, derived from gadoxetic acid-enhanced MRI, had almost perfect inter-reader reliability, showing a meta-analytic pooled estimate of 0.90–0.95. Meta-analytic pooled inter-reader reliability of FLIS showed

moderate study heterogeneity, but study methodology, MRI-related factors, and reader experience were not significantly associated with study heterogeneity.

In modern practice, MRI is widely used for diagnosis and follow-up in patients with chronic liver disease and cirrhosis. Under these circumstances, Bastati et al. introduced FLIS, a simple parameter derived from gadoxetic acid-enhanced MRI [5]. FLIS is directly associated with liver function and can predict the risk of liver-related complications or death [5, 8, 9]. These results suggest that FLIS is a promising imaging biomarker, and high reproducibility is essential for imaging biomarkers [10]. FLIS demonstrated almost perfect inter-reader reliability in this study, highlighting its reproducibility and robustness. The high reliability may have been associated with the simplicity and intuitiveness of FLIS as a scoring system. FLIS is a semi-quantitative parameter that does not require signal intensity measurements, complex equations, or specific software.

Bias among readers can cause changes in measurements because their subjectivity influences the test results [22]. It can result from differences in training, experience, and frames of reference between readers. However, in this meta-analysis, the difference in reader experience was not a significant factor in the inter-reader reliability of the FLIS. All covariates associated with reader experience, namely, differences in reader experience (all experienced readers vs. multiple readers with trainees), average reader experience,

**Table 3** Meta-regression analysis of the functional liver imaging score

Covariates	Subgroup	No. of studies	Estimate (95% CI)	<i>p</i> value
Subject enrollment	Consecutive	4	0.92 (0.87–0.98)	0.94
	Selective	2	0.93 (0.87–0.99)	
Country of study	Western	3	0.90 (0.87–0.93)	0.06
	Eastern	3	0.95 (0.90–1.00)	
MRI field-strength	3.0-T	3	0.90 (0.88–0.94)	0.25
	1.5-T included	3	0.94 (0.88–1.00)	
MRI vendor	Single	5	0.92 (0.87–0.96)	0.50
	Multiple	1	0.95 (0.95–0.96)	
MRI scanner	Single	4	0.90 (0.86–0.93)	0.10
	Multiple	2	0.97 (0.74–1.00)	
Number of readers	Two readers	4	0.94 (0.89–0.99)	0.30
	≥ 3 readers	2	0.90 (0.86–0.94)	
Difference in reader experience	All experienced readers	3	0.92 (0.87–0.97)	0.62
	Multiple readers with trainee	3	0.93 (0.87–1.00)	
Average reader experience	≥ 9.6 years in abdominal imaging <sup>a</sup>	4	0.92 (0.89–0.96)	0.62
	< 9.6 years	2	0.94 (0.83–1.00)	
Homogeneity in reader experience	Homogenous <sup>b</sup>	1	0.89 (0.83–0.95)	0.48
	Heterogenous	5	0.93 (0.89–0.97)	

Results were obtained using meta-regression analysis with a random-effects model *CI* confidence interval

<sup>a</sup>Average reader experience in all included studies

<sup>b</sup>Difference in reader experience of less than three years

and homogeneity in reader experience (homogenous vs. heterogeneous), showed almost perfect inter-reader reliability. These results are consistent with previous studies in which there was no significant difference in the inter-reader reliability between board-certified radiologists and trainees [8, 11, 13]. Considering the results of previous studies and this meta-analysis, FLIS is a reliable and reproducible grading system that can be used independently of the reader's experience.

FLIS was developed as an alternative to other complex and quantitative methods for evaluating the hepatobiliary phase uptake of gadoxetic acid [2, 4]. Thus, FLIS is designed not to be affected by MRI-related factors and is a simple visual assessment of the relative signal intensity of the liver parenchyma and portal vein and the presence of biliary secretion of contrast agents. This meta-analysis also showed that MRI-related factors, including vendor, scanner type, and magnetic field strength, did not affect the interpretation of the FLIS, resulting in high inter-reader reliability.

This study had some limitations. First, we could not include an original study that did not supply the standard variance of inter-reader reliability [9]. The standard variance and the ICC or  $\kappa$  from each study were needed to calculate meta-analytic pooled estimates of inter-reader reliability. Second, the meta-analytic pooled inter-reader reliability of FLIS showed moderate heterogeneity, and we could not identify the cause despite the robust meta-regression analysis. Nonetheless, because the inter-reader reliability of the FLIS from each original article before synthesis showed almost perfect reliability, moderate heterogeneity may not be a significant problem. However, further studies should be conducted to confirm any factors affecting the inter-reader reliability of FLIS. Third, some included studies reported the inter-reader reliability of FLIS subcategories only, without reporting FLIS itself. Therefore, we considered the median inter-reader reliability of the reported subcategories to be that of the FLIS.

In conclusion, the meta-analytic pooled estimate of the inter-reader reliability of FLIS and its three subcategories showed almost perfect reliability. Therefore, FLIS may be a reliable imaging parameter that reflects liver function and outcomes in patients with chronic liver disease. Further studies should be performed to confirm any factors affecting the inter-reader reliability of FLIS.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00261-022-03785-x>.

**Funding** This work was supported by a research fund from Hanyang University (HY-20220000001418).

## Declarations

**Conflict of interest** All authors declare that they have no conflict of interest.

**Ethical approval** Institutional review board approval and written informed consent were not required because this was a meta-analysis based on previously published studies.

## References

1. Van Beers BE, Pastor CM, Hussain HK (2012) Primovist, Eovist: what to expect? *J Hepatol* 57:421–429. doi: <https://doi.org/10.1016/j.jhep.2012.01.031>.
2. Ba-Ssalamah A, Bastati N, Wibmer A, et al. (2017) Hepatic gadoxetic acid uptake as a measure of diffuse liver disease: Where are we? *J Magn Reson Imaging* 45:646–659. doi: <https://doi.org/10.1002/jmri.25518>.
3. Cogley JR, Miller FH (2014) MR imaging of benign focal liver lesions. *Radiol Clin North Am* 52:657–682. doi: <https://doi.org/10.1016/j.rcl.2014.02.005>.
4. Poetter-Lang S, Bastati N, Messner A, et al. (2020) Quantification of liver function using gadoxetic acid-enhanced MRI. *Abdom Radiol (NY)* 45:3532–3544. doi: <https://doi.org/10.1007/s00261-020-02779-x>.
5. Bastati N, Wibmer A, Tamandl D, et al. (2016) Assessment of Orthotopic Liver Transplant Graft Survival on Gadoxetic Acid-Enhanced Magnetic Resonance Imaging Using Qualitative and Quantitative Parameters. *Invest Radiol* 51:728–734. doi: <https://doi.org/10.1097/rli.0000000000000286>.
6. Yoon JH, Lee JM, Kang HJ, et al. (2019) Quantitative Assessment of Liver Function by Using Gadoxetic Acid-enhanced MRI: Hepatocyte Uptake Ratio. *Radiology* 290:125–133. doi: <https://doi.org/10.1148/radiol.2018180753>.
7. Yoon JH, Lee JM, Kim E, et al. (2017) Quantitative Liver Function Analysis: Volumetric T1 Mapping with Fast Multi-section B(1) Inhomogeneity Correction in Hepatocyte-specific Contrast-enhanced Liver MR Imaging. *Radiology* 282:408–417. doi: <https://doi.org/10.1148/radiol.2016152800>.
8. Bastati N, Beer L, Mandorfer M, et al. (2020) Does the Functional Liver Imaging Score Derived from Gadoxetic Acid-enhanced MRI Predict Outcomes in Chronic Liver Disease? *Radiology* 294:98–107. doi: <https://doi.org/10.1148/radiol.2019190734>.
9. Bastati N, Beer L, Ba-Ssalamah A, et al. (2022) Gadoxetic Acid-enhanced MRI-derived Functional Liver Imaging Score (FLIS) and Spleen Diameter Predict Outcomes in ACLD. *J Hepatol*. doi: <https://doi.org/10.1016/j.jhep.2022.04.032>.
10. Barnhart HX, Barboriak DP (2009) Applications of the repeatability of quantitative imaging biomarkers: a review of statistical analysis of repeat data sets. *Transl Oncol* 2:231–235. doi: <https://doi.org/10.1593/tlo.09268>.
11. Aslan S, Eryuruk U, Tasdemir MN, Cakir IM (2022) Determining the efficacy of functional liver imaging score (FLIS) obtained from gadoxetic acid-enhanced MRI in patients with chronic liver disease and liver cirrhosis: the relationship between Albumin-Bilirubin (ALBI) grade and FLIS. *Abdom Radiol (NY)*. doi: <https://doi.org/10.1007/s00261-022-03557-7>.
12. Hwang JA, Min JH, Kim SH, et al. (2022) Total Bilirubin Level as a Predictor of Suboptimal Image Quality of the Hepatobiliary Phase of Gadoxetic Acid-Enhanced MRI in Patients with Extrahepatic Bile Duct Cancer. *Korean J Radiol* 23:389–401. doi: <https://doi.org/10.3348/kjr.2021.0407>.
13. Lee HJ, Hong SB, Lee NK, et al. (2021) Validation of functional liver imaging scores (FLIS) derived from gadoxetic acid-enhanced MRI in patients with chronic liver disease and liver cirrhosis: the relationship between Child-Pugh score and FLIS. *Eur Radiol* 31:8606–8614. doi: <https://doi.org/10.1007/s00330-021-07955-1>.



14. Luo N, Huang X, Ji Y, et al. (2022) A functional liver imaging score for preoperative prediction of liver failure after hepatocellular carcinoma resection. *Eur Radiol*. doi: <https://doi.org/10.1007/s00330-022-08656-z>.
15. Stroup DF, Berlin JA, Morton SC, et al. (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Jama* 283:2008-2012. doi: <https://doi.org/10.1001/jama.283.15.2008>.
16. Liberati A, Altman DG, Tetzlaff J, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Bmj* 339:b2700. doi: <https://doi.org/10.1136/bmj.b2700>.
17. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj* 339:b2535. doi: <https://doi.org/10.1136/bmj.b2535>.
18. Kottner J, Audigé L, Brorson S, et al. (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 64:96-106. doi: <https://doi.org/10.1016/j.jclinepi.2010.03.002>.
19. IntHout J, Ioannidis JP, Borm GF (2014) The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol* 14:25. doi: <https://doi.org/10.1186/1471-2288-14-25>.
20. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159-174.
21. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *Bmj* 327:557-560. doi: <https://doi.org/10.1136/bmj.327.7414.557>.
22. Bartlett JW, Frost C (2008) Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol* 31:466-475. doi: <https://doi.org/10.1002/uog.5256>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.