Original Article | Experiment, Engineering, and Physics

https://doi.org/10.3348/kjr.2017.18.6.888
pISSN 1229-6929 · eISSN 2005-8330
Korean J Radiol 2017;18(6):888-897

Korean Journal of Radiology
KJR

Selection and Reporting of Statistical Methods to Assess Reliability of a Diagnostic Test: Conformity to Recommended Methods in a Peer-Reviewed Journal

Ji Eun Park, MD, PhD[1]*, Kyunghwa Han, PhD[2]*, Yu Sub Sung, PhD[1], Mi Sun Chung, MD, PhD[3], Hyun Jung Koo, MD[1], Hee Mang Yoon, MD, PhD[1], Young Jun Choi, MD, PhD[1], Seung Soo Lee, MD, PhD[1], Kyung Won Kim, MD, PhD[1], Youngbin Shin, MS[1], Suah An, BS[1], Hyo-Min Cho, PhD[4], Seong Ho Park, MD, PhD[1]

[1]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul 05505, Korea; [2]Department of Radiology, Research Institute of Radiological Science, Yonsei University College of Medicine, Seoul 03722, Korea; [3]Department of Radiology, Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul 06973, Korea; [4]Korea Research Institute of Standards and Science, Daejeon 34113, Korea

Objective: To evaluate the frequency and adequacy of statistical analyses in a general radiology journal when reporting a reliability analysis for a diagnostic test.

Materials and Methods: Sixty-three studies of diagnostic test accuracy (DTA) and 36 studies reporting reliability analyses published in the *Korean Journal of Radiology* between 2012 and 2016 were analyzed. Studies were judged using the methodological guidelines of the Radiological Society of North America-Quantitative Imaging Biomarkers Alliance (RSNA-QIBA), and COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative. DTA studies were evaluated by nine editorial board members of the journal. Reliability studies were evaluated by study reviewers experienced with reliability analysis.

Results: Thirty-one (49.2%) of the 63 DTA studies did not include a reliability analysis when deemed necessary. Among the 36 reliability studies, proper statistical methods were used in all (5/5) studies dealing with dichotomous/nominal data, 46.7% (7/15) of studies dealing with ordinal data, and 95.2% (20/21) of studies dealing with continuous data. Statistical methods were described in sufficient detail regarding weighted kappa in 28.6% (2/7) of studies and regarding the model and assumptions of intraclass correlation coefficient in 35.3% (6/17) and 29.4% (5/17) of studies, respectively. Reliability parameters were used as if they were agreement parameters in 23.1% (3/13) of studies. Reproducibility and repeatability were used incorrectly in 20% (3/15) of studies.

Conclusion: Greater attention to the importance of reporting reliability, thorough description of the related statistical methods, efforts not to neglect agreement parameters, and better use of relevant terminology is necessary.

Keywords: *Reliability; Reproducibility; Repeatability; Agreement; Statistical method; Statistical analysis; Software program; Repeatability coefficient*

Received July 28, 2017; accepted after revision July 28, 2017.
This study was supported by a grant from the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (HI17C1862).

*These authors contributed equally to this work.

Corresponding author: Seong Ho Park, MD, PhD, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.
• Tel: (822) 3010-5984 • Fax: (822) 476-4719 • E-mail: parksh.radiology@gmail.com

Korean Journal of Radiology
KJR

INTRODUCTION

In addition to its accuracy, reliability (used in this article as an umbrella term to cover various concepts such as reproducibility, repeatability, and agreement except when used in a fixed expression of "reliability parameter," which will be further explained later in the Materials and Methods section) is an important performance metric of a diagnostic test (1, 2). The problem of omitting a proper analysis of reliability in diagnostic research studies has previously been recognized (1, 2). However, this issue was still cited as one of the top 10 statistical errors seen in the submissions to one prominent journal in the field of medical imaging in the recent past (3). The lack of familiarity of the investigators and peer reviewers with the statistical tools designed for this purpose was among the main reasons for the suboptimal reporting reliability analysis in diagnostic research studies (1). Regarding this, to help guide the proper use of the statistical tools for reliability analysis, the Radiological Society of North America-Quantitative Imaging Biomarkers Alliance (RSNA-QIBA) (https://www.rsna.org/QIBA), and COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative (http://www.cosmin.nl) have recently provided methodological guides (4-6). Furthermore, it appears that investigators, and perhaps also journals themselves, might be less attentive to reporting the reliability analysis when compared with the accuracy analysis. For example, although the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (7) exist, these do not seem to be well-known or referred to as often as the STAndards for Reporting of Diagnostic accuracy (STARD) (8). According to a study by a general radiology journal, the *Korean Journal of Radiology*, many more studies reporting diagnostic accuracy were published compared with those reporting reliability in the same period (9). Furthermore, in contrast with multiple secondary research studies analyzing the reporting quality of diagnostic test accuracy (DTA) (9-14), similar secondary research studies of reliability analyses are scarce.

In this regard, we performed this study to evaluate the frequency of reporting a reliability analysis in DTA studies. In addition, we aimed to assess how appropriately the statistical methods for reliability analysis were selected and reported in published studies using the methodological guides provided by the RSNA-QIBA and COSMIN initiative as the adjudication tool with studies from a general radiology journal as a sample.

MATERIALS AND METHODS

Article Search Strategy and Study Selection

We conducted a search to identify all potentially relevant original research papers from the articles published in a single peer-reviewed journal, the *Korean Journal of Radiology*, during the 5-year period between January 1, 2012 and December 31, 2016 using the PubMed Medline database. The search terms to find DTA studies were "sensitivity" OR "specificity" OR "accuracy" OR "performance" OR "receiver operating" OR "ROC." The search terms to find studies that analyzed reliability included "reliability" OR "repeatability" OR "reproducibility" OR "agreement" OR "precision" OR "biomarker." Retrieved articles were screened for eligibility. Regarding the DTA studies, one reviewer experienced in DTA studies selected eligible articles according to criteria established elsewhere (9) with additional confirmation by another DTA expert in cases of ambiguity. Of the initial 124 candidate articles, 63 articles (15-77) were finally included. Regarding the studies that analyzed reliability, eligible articles were chosen by consensus after review by two of four independent reviewers experienced in the relevant methodology. When the two reviewers disagreed or in cases of ambiguity, a third reviewer experienced in related methodology was invited as an adjudicator. We excluded studies that investigated the agreement between continuous or ordinal outcomes/test results and fixed reference standard results (78-80). These studies could be viewed as extensions of DTA analysis of non-binary data, which require different statistical analyses (81), than the standard analysis used for reliability, although some published studies seem to have failed to distinguish between them. Of the initial 71 article candidates, 36 articles (15, 19, 42, 45, 53, 57, 58, 64, 66, 67, 69, 82-106) were finally included.

Data Extraction for DTA Studies

Diagnostic test accuracy studies were evaluated regarding whether they also analyzed the reliability of the investigated tests/methods and, when reliability was not assessed, whether the reliability analysis was deemed necessary per se. We considered reliability analysis unnecessary if the tests/methods investigated in a DTA study were only a minor component of the study or if their reliability was already well established. The extraction of this information was performed by nine independent editorial board members of the journal (names are listed

in the acknowledgment section). Each reviewer was assigned to the articles in his/her area of expertise (two to ten articles per reviewer). When there is doubt, a second reviewer additionally reviewed the article to make a consensus decision with the original reviewer.

Data Extraction for Reliability Studies

Before data extraction, we first established the recommended statistical methods for the analysis of the reliability of a test/method (Table 1) according to the methodological guides provided by the RSNA-QIBA and COSMIN initiative (4, 6, 107, 108). We then used the table as the reference when evaluating if the articles conformed to the recommended statistical methods. Each article was evaluated by two of four independent reviewers experienced in the statistical methodology. Disagreements between two reviewers were adjudicated by two additional reviewers (a biostatistician) both of whom were also experienced in the statistical methodology. The reviewers extracted the data using a predetermined standardized set of questionnaires, which were intended to address the following issues. First, if authors used the proper statistical methods according to the suggestions that we established for this study (Table 1). Second, if authors provided a detailed description of the statistical methods. Third, for studies assessing the reliability of a continuous outcome, if authors distinguished the difference between the "reliability parameter" and "agreement parameter" (Table 1) and used them appropriately with respect to the study purpose and conclusion. Fourth, when the terms "reproducibility" and "repeatability" were used, if authors used the correct definitions.

The "reliability parameter" is a term that has a specific meaning as defined elsewhere (4, 108), unlike reliability which is used as a general umbrella term. Reliability parameters, such as the intraclass correlation coefficient (ICC) or concordance correlation coefficient, explain how well the subjects in a study set can be distinguished from each other (108), but they do not show the exact measurement uncertainties. Small measurement uncertainties (as opposed to large measurement uncertainties) would allow for a clear distinction between the subjects, yielding a large reliability parameter score. However, a clear distinction between subjects can also be obtained even with large measurement uncertainties if there are large differences between subjects (statistically referred to as a large between-subject variance). Therefore, although reliability parameters are useful in making a relative comparison between different tests/methods regarding their levels of reliability, i.e., a higher score means greater reliability (109), they are not helpful if one wants to know what specific range of measurement differences should be considered true changes instead of mere measurement uncertainties in a longitudinal follow-up. On the other hand, "agreement parameters" assess exactly how close the results for repeated measurements are (108). Therefore, agreement parameters can be used both for the relative comparison of reliability and assessment of absolute measurement uncertainties. Agreement parameters are needed when investigating a test/method for potential use in a longitudinal follow-up setting. Repeatability, as defined by RSNA-QIBA, concerns repeated measurements

Table 1. Recommended Statistical Methods for Analysis of Reliability

Dichotomous or Nominal Data (e.g., Benign vs. Malignant)	Ordinal Data (e.g., Grades I, II, III, and IV)	Continuous Data (e.g., Tumor Volume in mL)
Kappa	Weighted kappa	Reliability parameters:
Proportion of agreement	ICC	ICC
		CCC
		Agreement parameters:
		Within-subject standard deviation
		Repeatability coefficient and reproducibility coefficient
		Coefficient of variation
		Bland-Altman limits of agreement

ICC has three different models including one-way random, two-way random, and two-way mixed models, and can use either consistency or absolute agreement assumptions. As ICC value for same set of data may change according to model and assumption used, it is desirable to describe model and assumption, for example, as shown in study by Yoo et al. (86). ICC calculated using one-way random model is appropriate for assessing repeatability (112). CCC or ICC calculated using two-way model, random or mixed according to data and setting (6), are appropriate for analyzing reproducibility. Intraobserver reliability could be regarded as similar to repeatability depending on study setting, whereas interobserver reliability should be regarded as reproducibility. CCC = concordance correlation coefficient, ICC = intraclass correlation coefficient

Korean Journal of Radiology
KJR

of the same or similar experimental units under identical or near-identical conditions, using the same measurement procedure, same operators, same measuring system, same operating conditions, and same physical location over a short period (5, 6). On the other hand, reproducibility applies to rerunning a measurement in slightly different settings, for example, different locations, operators, scanners, etc. (5, 6).

Statistical Analysis

We obtained the following study outcomes in a descriptive manner using proportions, i.e., the percentage of articles out of all eligible articles, for each of the following outcome categories:

- Reporting of reliability along with accuracy
- Use of the recommended statistical methods. We considered that a study satisfied this item if the study used at least one method listed in Table 1 and did not require any further details (for example, explanations of weighting methods for weighted kappa or descriptions of the ICC model and assumption were not considered). The results were obtained for each of three different data types (dichotomous/nominal, ordinal, and continuous data).
- Reporting of weighting method when weighted kappa was used.
- Reporting of model and assumption when ICC was used.
- Appropriate use/interpretation of reliability parameters
- Correct use of the terms reproducibility and repeatability

RESULTS

Reporting of Reliability along with Accuracy

Of the 63 DTA studies (15-77), 32 studies (50.8%) included an analysis of reliability (n = 22) or did not include reliability analysis when the analysis was not necessary (n = 10). Thirty-one articles (49.2%) did not include a reliability analysis in cases where the analysis was deemed necessary.

Selection and Reporting of Statistical Methods to Assess Reliability

The results obtained from the 36 eligible studies (15, 19, 42, 45, 53, 57, 58, 64, 66, 67, 69, 82-106) are summarized in Table 2.

Of the five studies that reported an analysis of dichotomous/nominal data, four studies used kappa, and one study used both kappa and proportion of agreement.

Of the 15 studies that reported an analysis of ordinal data, six studies used weighted kappa, and one study used both weighted kappa and ICC, whereas eight studies used kappa without clarifying if they calculated weighted kappa.

Of the 21 studies that reported an analysis of continuous data, one study used Pearson's correlation coefficient instead of the recommended methods. The 20 other studies used the recommended methods, including reliability parameters alone (n = 13, 65%), agreement parameters alone (n = 2, 10%), and both reliability and agreement parameters (n = 5, 25%). Of the 17 studies that used ICC, 11 studies (64.7%) did not report the ICC model, and 12 studies (70.6%) did not explain the assumptions made for the ICC.

Of the 13 studies that used reliability parameters alone, ten studies properly used and interpreted the analysis for the study purpose and conclusion, whereas three studies (23.1%) inappropriately considered the reliability parameters as if they were agreement parameters.

Among the 15 studies that used reproducibility or repeatability, three studies did not use them accurately, with two studies incorrectly using reproducibility instead of repeatability and one study incorrectly using repeatability

Table 2. Selection and Reporting of Statistical Methods to Assess Reliability

Items	No. of Eligible Articles (Denominator)	Yes (%)	No or Uncertain (%)
Use of recommended statistical methods			
Analysis of dichotomous/nominal data	5	5 (100.0)	0 (0.0)
Analysis of ordinal data	15	7 (46.7)	8 (53.3)
Analysis of continuous data	21	20 (95.2)	1 (4.8)
Reporting of weighting method for weighted kappa	7	2 (28.6)	5 (71.4)
Reporting of model for ICC	17	6 (35.3)	11 (64.7)
Reporting of assumption for ICC	17	5 (29.4)	12 (70.6)
Appropriate use/interpretation of reliability parameters	13	10 (76.9)	3 (23.1)
Correct meaning of reproducibility and repeatability	15	12 (80.0)	3 (20.0)

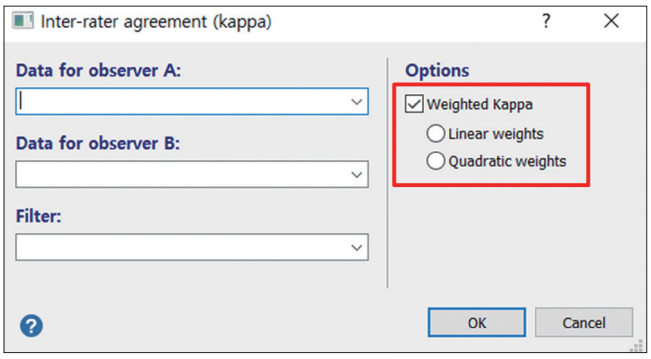
Data are numbers of articles with proportion of eligible articles for each item described as percentage in parentheses.

instead of reproducibility.

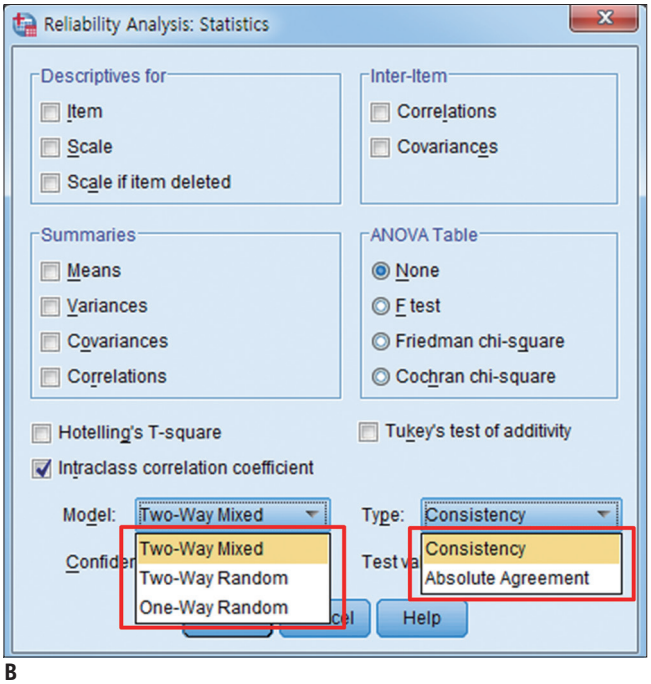
DISCUSSION

In our study, approximately half of the DTA studies did not include a reliability analysis when it was deemed necessary. Most of the reliability studies seem to have selected the proper statistical methods for the analysis. However, description of the further details of the statistical methods, including the weighting method for weighted kappa and specific model and assumption for ICC, were generally poor. This study is limited in that we analyzed a single peer-reviewed journal and did not have specific data from other journals. However, according to the current authors' experience, other radiology journals seem to have similar trends. Another notable observation was that studies more frequently used reliability parameters than agreement parameters for analyzing the reliability of continuous data, and a small but notable (23.1%) fraction of studies imprecisely interpreted the reliability parameters. Lastly, the distinction between repeatability and reproducibility was not perfect. These weaknesses found in the published papers would indicate the areas to require improvements in the future.

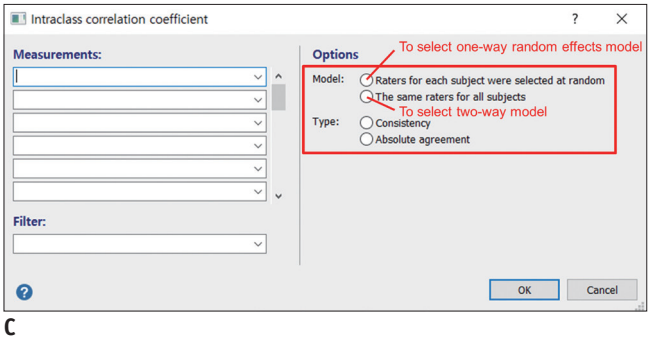
The importance of reporting reliability along with accuracy needs to be further emphasized because these two parameters are necessary complementary parameters of technical performance and clinical utility for an imaging biomarker (110). It is reassuring that the published studies overall selected the proper methods for reliability analysis. For those investigators who are not familiar with the statistical methods, the table of suggested methods we made for this study (Table 1) could be a useful reference as it succinctly summarizes the well-thought-out methodological guides by the RSNA-QIBA and COSMIN initiative (4, 6, 107, 108). Regarding the suboptimal reporting of the details of the statistical methods, in fact, some user-friendly software programs for statistical analysis, which authors frequently quote as having been used for statistical analysis, often include the details as optional parameters and report them in their output (Fig. 1). Paying closer attention to these features would facilitate reporting them more clearly and would also help investigators to select the most appropriate statistical analysis. The use of agreement parameters, when applicable, should also be more encouraged. It was reported that agreement parameters were often neglected in medical research



A



B



C

Fig. 1. Display of detailed options associated with statistical tests used for reliability analysis in some user-friendly software programs.
A. Selection of weighting method to calculate weighted kappa with MedCalc Version 17.6 (MedCalc Software BVBA; https://www.medcalc.org). **B.** Selection of model and assumption to calculate ICC with IBM SPSS Statistics for Windows Version 21 (IBM Corp.). **C.** Selection of model and assumption to calculate ICC with MedCalc Version 17.6 (MedCalc Software BVBA). This software program does not distinguish between random and fixed effects models. ICC = intraclass correlation coefficient

Korean Journal of Radiology
KJR

studies (108), as was also seen in our study. Among these parameters, the repeatability coefficient (RC) is particularly important as it is the smallest detectable change based on the intrinsic technical uncertainties of a quantitative measurement method and its importance is highlighted by the RSNA-QIBA (6, 108). One of the reasons why the agreement parameters are underutilized compared with reliability parameters may be the lack of readily available user-friendly software programs, except for the Bland-Altman analysis. In this regard, we have developed a web calculator to compute RC and its 95% confidence interval for two or more repeat measurements of a continuous parameter (available at http://datasharing.aim-aicro.com/reliability) according to the methods proposed elsewhere (6, 111). A software tool like this would help promote the use of agreement parameters such as RC in analyzing the reliability of quantitative imaging parameters.

Limitations of this study include the fact that the eligible articles were selected from a single journal and, therefore, there could be an issue regarding generalizability. Nevertheless, the journal, the *Korean Journal of Radiology*, is a representative general journal in the radiology/medical imaging field ranked 53rd out of 126 journals in the field according to the 2016 Journal Citation Reports by Clarivate Analytics. Given its rank and the coverage of topics, the *Korean Journal of Radiology* may be a suitable litmus test for journals in general in the radiology/medical imaging field. Second, as we focused on the quality of the reporting of the statistical analysis, our results do not necessarily reflect the overall reporting quality or quality of the research.

In conclusion, the quality of reporting the reliability analysis of a diagnostic test can be improved through greater attention to the importance of reporting the reliability of a test, more thorough description of the related statistical methods, efforts not to neglect agreement parameters, and a clearer distinction of reproducibility and repeatability. Some of the tips discussed in this article, including the software tool to calculate the RC, may be helpful.

Acknowledgments

We appreciate following editorial board members of the *Korean Journal of Radiology* for their help with the literature analysis:

Jung Hwan Baek, MD, PhD (University of Ulsan, Korea), Joon Young Choi, MD, PhD (Sungkyunkwan University, Korea), Boo-Kyung Han, MD, PhD (Sungkyunkwan University, Korea), Chang Hee Lee, MD, PhD (Korea University, Korea), Hyun-Ju Lee, MD, PhD (Seoul National University, Korea), Jeong Min Lee, MD, PhD (Seoul National University, Korea), Won-Jin Moon, MD, PhD (Konkuk University, Korea), Deuk Jae Sung, MD, PhD (Korea University, Korea), Young Cheol Yoon, MD, PhD (Sungkyunkwan University, Korea)

REFERENCES

1. Bankier AA, Levine D, Halpern EF, Kressel HY. Consensus interpretation in imaging research: is there a better way? *Radiology* 2010;257:14-17
2. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645-651
3. Levine D, Bankier AA, Halpern EF. Submissions to radiology: our top 10 list of statistical errors. *Radiology* 2009;253:288-290
4. Hernaez R. Reliability and agreement studies: a guide for clinical investigators. *Gut* 2015;64:1018-1027
5. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Stat Methods Med Res* 2015;24:9-26
6. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res* 2015;24:27-67
7. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96-106
8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology* 2015;277:826-832
9. Choi YJ, Chung MS, Koo HJ, Park JE, Yoon HM, Park SH. Does the reporting quality of diagnostic test accuracy studies, as defined by STARD 2015, affect citation? *Korean J Radiol* 2016;17:706-714
10. Gallo L, Hua N, Mercuri M, Silveira A, Worster A; Best Evidence in Emergency Medicine (BEEM; beem.ca). Adherence to standards for reporting diagnostic accuracy in emergency medicine research. *Acad Emerg Med* 2017 Jun 16 [Epub]. https://doi.org/10.1111/acem.13233
11. Grob AT, van der Vaart LR, Withagen MI, van der Vaart CH. The quality of reporting of diagnostic accuracy studies in pelvic floor transperineal three-dimensional ultrasound: a systematic review. *Ultrasound Obstet Gynecol* 2016 Dec 21 [Epub]. https://doi.org/10.1002/uog.17390
12. Hong PJ, Korevaar DA, McGrath TA, Ziai H, Frank R, Alabousi M, et al. Reporting of imaging diagnostic accuracy studies

with focus on MRI subgroup: Adherence to STARD 2015. *J Magn Reson Imaging* 2017 Jun 22 [Epub]. https://doi.org/10.1002/jmri.25797
13. Korevaar DA, Cohen JF, Hooft L, Bossuyt PM. Literature survey of high-impact journals revealed reporting weaknesses in abstracts of diagnostic accuracy studies. *J Clin Epidemiol* 2015;68:708-715
14. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, Smidt N, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology* 2015;274:781-789
15. Xu XQ, Hu H, Su GY, Liu H, Shi HB, Wu FY. Diffusion weighted imaging for differentiating benign from malignant orbital tumors: diagnostic performance of the apparent diffusion coefficient based on region of interest selection method. *Korean J Radiol* 2016;17:650-656
16. Park HJ, Lee SY, Rho MH, Chung EC, Kim MS, Kwon HJ, et al. Single-shot echo-planar diffusion-weighted MR imaging at 3T and 1.5T for differentiation of benign vertebral fracture edema and tumor infiltration. *Korean J Radiol* 2016;17:590-597
17. Liu J, Ji Y, Ai H, Ning B, Zhao J, Zhang Y, et al. Liver shear-wave velocity and serum fibrosis markers to diagnose hepatic fibrosis in patients with chronic viral hepatitis B. *Korean J Radiol* 2016;17:396-404
18. Lim HJ, Chung MJ, Shin KE, Hwang HS, Lee KS. The impact of iterative reconstruction in low-dose computed tomography on the evaluation of diffuse interstitial lung disease. *Korean J Radiol* 2016;17:950-960
19. Lee JE, Cho JS, Shin KS, Kim SS, You SK, Park JW, et al. Diffuse infiltrative splenic lymphoma: diagnostic efficacy of arterial-phase CT. *Korean J Radiol* 2016;17:734-741
20. Lee H, Lee WW, Park SY, Kim SE. F-18 sodium fluoride positron emission tomography/computed tomography for detection of thyroid cancer bone metastasis compared with bone scintigraphy. *Korean J Radiol* 2016;17:281-288
21. Lee EH, Kim KW, Kim YJ, Shin DR, Park YM, Lim HS, et al. Performance of screening mammography: a report of the alliance for breast cancer screening in Korea. *Korean J Radiol* 2016;17:489-496
22. Joo SM, Kim YP, Yum TJ, Eun NL, Lee D, Lee KH. Optimized performance of flightplan during chemoembolization for hepatocellular carcinoma: importance of the proportion of segmented tumor area. *Korean J Radiol* 2016;17:771-778
23. Hong HS, Cho HS, Woo JY, Lee Y, Yang I, Hwang JY, et al. Intra-appendiceal air at CT: is it a useful or a confusing sign for the diagnosis of acute appendicitis? *Korean J Radiol* 2016;17:39-46
24. Ha EJ, Moon WJ, Na DG, Lee YH, Choi N, Kim SJ, et al. A multicenter prospective validation study for the Korean thyroid imaging reporting and data system in patients with thyroid nodules. *Korean J Radiol* 2016;17:811-821
25. Ding Y, Zeng M, Rao S, Chen C, Fu C, Zhou J. Comparison of biexponential and monoexponential model of diffusion-weighted imaging for distinguishing between common renal cell carcinoma and fat poor angiomyolipoma. *Korean J Radiol* 2016;17:853-863
26. Choi TW, Lee JM, Kim JH, Yu MH, Han JK, Choi BI. Comparison of multidetector CT and gadobutrol-enhanced MR imaging for evaluation of small, solid pancreatic lesions. *Korean J Radiol* 2016;17:509-521
27. Alan B, Göya C, Tunç S, Teke M, Hattapoğlu S. Assessment of placental stiffness using acoustic radiation force impulse elastography in pregnant women with fetal anomalies. Korean *J Radiol* 2016;17:218-223
28. Ahn JH, Yu JS, Cho ES, Chung JJ, Kim JH, Kim KW. Diffusion-weighted MRI of malignant versus benign portal vein thrombosis. *Korean J Radiol* 2016;17:533-540
29. Ryoo I, Suh S, Lee YH, Seo HS, Seol HY. Comparison of ultrasonographic findings of biopsy-proven tuberculous lymphadenitis and kikuchi disease. *Korean J Radiol* 2015;16:767-775
30. Park B, Kim HK, Choi YS, Kim J, Zo JI, Choi JY, et al. Prediction of pathologic grade and prognosis in mucoepidermoid carcinoma of the lung using ^{18}F-FDG PET/CT. *Korean J Radiol* 2015;16:929-935
31. Niu XK, Bhetuwal A, Yang HF. CT-guided core needle biopsy of pleural lesions: evaluating diagnostic yield and associated complications. *Korean J Radiol* 2015;16:206-212
32. Lu Q, Huang BJ, Wang WP, Li CX, Xue LY. Qualitative and quantitative analysis with contrast-enhanced ultrasonography: diagnosis value in hypoechoic renal angiomyolipoma. *Korean J Radiol* 2015;16:334-341
33. Lee S, Lee YH, Chung TS, Jeong EK, Kim S, Yoo YH, et al. Accuracy of diffusion tensor imaging for diagnosing cervical spondylotic myelopathy in patients showing spinal cord compression. *Korean J Radiol* 2015;16:1303-1312
34. Lee EK, Choi SH, Yun TJ, Kang KM, Kim TM, Lee SH, et al. Prediction of response to concurrent chemoradiotherapy with temozolomide in glioblastoma: application of immediate post-operative dynamic susceptibility contrast and diffusion-weighted MR imaging. *Korean J Radiol* 2015;16:1341-1348
35. Kim SA, Chang JM, Cho N, Yi A, Moon WK. Characterization of breast lesions: comparison of digital breast tomosynthesis and ultrasonography. *Korean J Radiol* 2015;16:229-238
36. Kim J, Kim YH, Lee KH, Lee YJ, Park JH. Diagnostic performance of CT angiography in patients visiting emergency department with overt gastrointestinal bleeding. *Korean J Radiol* 2015;16:541-549
37. Kim DW, Jung SL, Kim J, Ryu JH, Sung JY, Lim HK. Comparison between ultrasonography and computed tomography for detecting the pyramidal lobe of the thyroid gland: a prospective multicenter study. *Korean J Radiol* 2015;16:402-409
38. Kang KA, Kim YK, Kim E, Jeong WK, Choi D, Lee WJ, et al. T2-weighted liver MRI using the multivane technique at 3T: comparison with conventional T2-weighted MRI. *Korean J Radiol* 2015;16:1038-1046
39. Jung SI, Park HS, Yim Y, Jeon HJ, Yu MH, Kim YJ, et al. Added value of using a CT coronal reformation to diagnose

adnexal torsion. *Korean J Radiol* 2015;16:835-845
40. Chun KY, Choi YS, Lee SH, Kim JS, Young KW, Jeong MS, et al. Deltoid ligament and tibiofibular syndesmosis injury in chronic lateral ankle instability: magnetic resonance iImaging evaluation at 3T and comparison with arthroscopy. *Korean J Radiol* 2015;16:1096-1103
41. Yu H, Cui JL, Cui SJ, Sun YC, Cui FZ. Differentiating benign from malignant bone tumors using fluid-fluid level features on magnetic resonance imaging. *Korean J Radiol* 2014;15:757-763
42. Yoon SH, Goo JM, Jung J, Hong H, Park EA, Lee CH, et al. Computer-aided classification of visual ventilation patterns in patients with chronic obstructive pulmonary disease at two-phase xenon-enhanced CT. *Korean J Radiol* 2014;15:386-396
43. Yi J, Lee EH, Kwak JJ, Cha JG, Jung SH. Retrieval rate and accuracy of ultrasound-guided 14-G semi-automated core needle biopsy of breast microcalcifications. *Korean J Radiol* 2014;15:12-19
44. Woo S, Kim SY, Cho JY, Kim SH. Shear wave elastography for detection of prostate cancer: a preliminary study. *Korean J Radiol* 2014;15:346-355
45. Ucar M, Guryildirim M, Tokgoz N, Kilic K, Borcek A, Oner Y, et al. Evaluation of aqueductal patency in patients with hydrocephalus: three-dimensional high-sampling-efficiency technique (SPACE) versus two-dimensional turbo spin echo at 3 Tesla. *Korean J Radiol* 2014;15:827-835
46. Luczyńska E, Heinze-Paluchowska S, Dyczek S, Blecharz P, Rys J, Reinfuss M. Contrast-enhanced spectral mammography: comparison with conventional mammography and histopathology in 152 women. *Korean J Radiol* 2014;15:689-696
47. Lee JH, Yoon YC, Jee S, Kwon JW, Cha JG, Yoo JC. Comparison of three-dimensional isotropic and two-dimensional conventional indirect MR arthrography for the diagnosis of rotator cuff tears. *Korean J Radiol* 2014;15:771-780
48. Lee JH, Byun JH, Kim JH, Lee SS, Kim HJ, Lee MG. Solid pancreatic tumors with unilocular cyst-like appearance on CT: differentiation from unilocular cystic tumors using CT. *Korean J Radiol* 2014;15:704-711
49. Lee JE, Lee JM, Lee KB, Yoon JH, Shin CI, Han JK, et al. Noninvasive assessment of hepatic fibrosis in patients with chronic hepatitis B viral infection using magnetic resonance elastography. *Korean J Radiol* 2014;15:210-217
50. Kim YP, Kannengiesser S, Paek MY, Kim S, Chung TS, Yoo YH, et al. Differentiation between focal malignant marrow-replacing lesions and benign red marrow deposition of the spine with T2*-corrected fat-signal fraction map using a three-echo volume interpolated breath-hold gradient echo Dixon sequence. *Korean J Radiol* 2014;15:781-791
51. Jung SI, Park HS, Kim YJ, Jeon HJ. Multidetector computed tomography for the assessment of adnexal mass: is unenhanced CT scan necessary? *Korean J Radiol* 2014;15:72-79
52. Iannicelli E, Di Renzo S, Ferri M, Pilozzi E, Di Girolamo M, Sapori A, et al. Accuracy of high-resolution MRI with lumen distention in rectal cancer staging and circumferential margin involvement prediction. *Korean J Radiol* 2014;15:37-44
53. Bang SH, Lee JY, Woo H, Joo I, Lee ES, Han JK, et al. Differentiating between adenomyomatosis and gallbladder cancer: revisiting a comparative study of high-resolution ultrasound, multidetector CT, and MR imaging. *Korean J Radiol* 2014;15:226-234
54. Ahn HS, Kim SM, Jang M, Yun BL, Kim B, Ko ES, et al. A new full-field digital mammography system with and without the use of an advanced post-processing algorithm: comparison of image quality and diagnostic performance. *Korean J Radiol* 2014;15:305-312
55. Yoon JH, Lee JM, Woo HS, Yu MH, Joo I, Lee ES, et al. Staging of hepatic fibrosis: comparison of magnetic resonance elastography and shear wave elastography in the same individuals. *Korean J Radiol* 2013;14:202-212
56. Wu EH, Chen YL, Wu YM, Huang YT, Wong HF, Ng SH. CT-guided core needle biopsy of deep suprahyoid head and neck lesions. *Korean J Radiol* 2013;14:299-306
57. Song YS, Choi SH, Park CK, Yi KS, Lee WJ, Yun TJ, et al. True progression versus pseudoprogression in the treatment of glioblastomas: a comparison study of normalized cerebral blood volume and apparent diffusion coefficient by histogram analysis. *Korean J Radiol* 2013;14:662-672
58. Rief M, Stenzel F, Kranz A, Schlattmann P, Dewey M. Time efficiency and diagnostic accuracy of new automated myocardial perfusion analysis software in 320-row CT cardiac imaging. *Korean J Radiol* 2013;14:21-29
59. Liu X, Peng W, Zhou L, Wang H. Biexponential apparent diffusion coefficients values in the prostate: comparison among normal tissue, prostate cancer, benign prostatic hyperplasia and prostatitis. *Korean J Radiol* 2013;14:222-232
60. Lim HJ, Chung MJ, Lee G, Yie M, Shin KE, Moon JW, et al. Interpretation of digital chest radiographs: comparison of light emitting diode versus cold cathode fluorescent lamp backlit monitors. *Korean J Radiol* 2013;14:968-976
61. Lee SJ, Lee WW, Kim SE. Bone positron emission tomography with or without CT is more accurate than bone scan for detection of bone metastasis. *Korean J Radiol* 2013;14:510-519
62. Lee KH, Lee JM, Park JH, Kim JH, Park HS, Yu MH, et al. MR imaging in patients with suspected liver metastases: value of liver-specific contrast agent gadoxetic acid. *Korean J Radiol* 2013;14:894-904
63. Lee DH, Lee JM, Klotz E, Kim SJ, Kim KW, Han JK, et al. Detection of recurrent hepatocellular carcinoma in cirrhotic liver after transcatheter arterial chemoembolization: value of quantitative color mapping of the arterial enhancement fraction of the liver. *Korean J Radiol* 2013;14:51-60
64. Koo JH, Kim CK, Choi D, Park BK, Kwon GY, Kim B. Diffusion-weighted magnetic resonance imaging for the evaluation of prostate cancer: optimal B value at 3T. *Korean J Radiol*

2013;14:61-69
65. Ko ES, Han BK, Kim SM, Ko EY, Jang M, Lyou CY, et al. Comparison of new and established full-field digital mammography systems in diagnostic performance. *Korean J Radiol* 2013;14:164-170
66. Kim SH, Kang BJ, Choi BG, Choi JJ, Lee JH, Song BJ, et al. Radiologists' performance for detecting lesions and the interobserver variability of automated whole breast ultrasound. *Korean J Radiol* 2013;14:154-163
67. Kim MY, Cho N, Yi A, Koo HR, Yun BL, Moon WK. Sonoelastography in distinguishing benign from malignant complex breast mass and making the decision to biopsy. *Korean J Radiol* 2013;14:559-567
68. Kim JI, Kim YH, Lee KH, Kim SY, Lee YJ, Park YS, et al. Type-specific diagnosis and evaluation of longitudinal tumor extent of borrmann type IV gastric cancer: CT versus gastroscopy. *Korean J Radiol* 2013;14:597-606
69. Kim JE, Lee JY, Bae KS, Han JK, Choi BI. Acoustic radiation force impulse elastography for focal hepatic tumors: usefulness for differentiating hemangiomas from malignant tumors. *Korean J Radiol* 2013;14:743-753
70. Jeh SK, Kim SH, Kang BJ. Comparison of the diagnostic performance of response evaluation criteria in solid tumor 1.0 with response evaluation criteria in solid tumor 1.1 on MRI in advanced breast cancer response evaluation to neoadjuvant chemotherapy. *Korean J Radiol* 2013;14:13-20
71. Choi HS, Kim AH, Ahn SS, Shin NY, Kim J, Lee SK. Glioma grading capability: comparisons among parameters from dynamic contrast-enhanced MRI and ADC value on DWI. *Korean J Radiol* 2013;14:487-492
72. Yoo JY, Chung MJ, Choi B, Jung HN, Koo JH, Bae YA, et al. Digital tomosynthesis for PNS evaluation: comparisons of patient exposure and image quality with plain radiography. *Korean J Radiol* 2012;13:136-143
73. Wu CH, Huang CC, Wang LJ, Wong YC, Wang CJ, Lo WC, et al. Value of CT in the discrimination of fatal from non-fatal stercoral colitis. *Korean J Radiol* 2012;13:283-289
74. Sohn CH, Lee HP, Park JB, Chang HW, Kim E, Kim E, et al. Imaging findings of brain death on 3-Tesla MRI. *Korean J Radiol* 2012;13:541-549
75. Lee KH, Goo JM, Park CM, Lee HJ, Jin KN. Computer-aided detection of malignant lung nodules on chest radiographs: effect on observers' performance. *Korean J Radiol* 2012;13:564-571
76. Kang KM, Choi SI, Chun EJ, Kim JA, Youn TJ, Choi DJ. Coronary vasospastic angina: assessment by multidetector CT coronary angiography. *Korean J Radiol* 2012;13:27-33
77. Chung SY, Park SH, Lee SS, Lee JH, Kim AY, Park SK, et al. Comparison between CT colonography and double-contrast barium enema for colonic evaluation in patients with renal insufficiency. *Korean J Radiol* 2012;13:290-299
78. Ko ES, Han H, Han BK, Kim SM, Kim RB, Lee GW, et al. Prognostic significance of a complete response on breast MRI in patients who received neoadjuvant chemotherapy according to the molecular subtype. *Korean J Radiol* 2015;16:986-995
79. Shin CI, Kim HC, Song YS, Cho HR, Lee KB, Lee W, et al. Rat model of hindlimb ischemia induced via embolization with polyvinyl alcohol and N-butyl cyanoacrylate. *Korean J Radiol* 2013;14:923-930
80. Suh YJ, Kim YJ, Hong YJ, Lee HJ, Hur J, Im DJ, et al. Measurement of opening and closing angles of aortic valve prostheses in vivo using Dual-Source Computed Tomography: comparison with those of manufacturers' in 10 different types. *Korean J Radiol* 2015;16:1012-1023
81. Yokoo T, Shiehmorteza M, Hamilton G, Wolfson T, Schroeder ME, Middleton MS, et al. Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. *Radiology* 2011;258:749-759
82. Yoo YH, Yoon CS, Eun NL, Hwang MJ, Yoo H, Peters RD, et al. Interobserver and test-retest reproducibility of T1ρ and T2 measurements of lumbar intervertebral discs by 3T magnetic resonance imaging. *Korean J Radiol* 2016;17:903-911
83. Yoo H, Lee JM, Yoon JH, Lee DH, Chang W, Han JK. Prospective comparison of liver stiffness measurements between two point shear wave elastography methods: virtual touch quantification and elastography point quantification. *Korean J Radiol* 2016;17:750-757
84. Song JS, Hwang SB, Chung GH, Jin GY. Intra-individual, inter-vendor comparison of diffusion-weighted MR imaging of upper abdominal organs at 3.0 Tesla with an emphasis on the value of normalization with the spleen. *Korean J Radiol* 2016;17:209-217
85. Yoon SJ, Yoon YC, Bae SY, Wang JH. Bone tunnel diameter measured with CT after anterior cruciate ligament reconstruction using double-bundle auto-hamstring tendons: clinical implications. *Korean J Radiol* 2015;16:1313-1318
86. Yoo YH, Kim HS, Lee YH, Yoon CS, Paek MY, Yoo H, et al. Comparison of multi-echo Dixon methods with volume interpolated breath-hold gradient echo magnetic resonance imaging in fat-signal fraction quantification of paravertebral muscle. *Korean J Radiol* 2015;16:1086-1095
87. Yi JS, Cha JG, Han JK, Kim HJ. Imaging of herniated discs of the cervical spine: inter-modality differences between 64-slice multidetector CT and 1.5-T MRI. *Korean J Radiol* 2015;16:881-888
88. Park HS, Han JK, Lee JM, Kim YI, Woo S, Yoon JH, et al. Dynamic contrast-enhanced MRI using a macromolecular MR contrast agent (P792): evaluation of antivascular drug effect in a rabbit VX2 liver tumor model. *Korean J Radiol* 2015;16:1029-1037
89. Park EA, Lee W, Kim HK, Chung JW. Effect of papillary muscles and trabeculae on left ventricular measurement using cardiovascular magnetic resonance imaging in patients with hypertrophic cardiomyopathy. *Korean J Radiol* 2015;16:4-12
90. Lee GY, Lee JW, Choi SW, Lim HJ, Sun HY, Kang Y, et al. MRI inter-reader and intra-reader reliabilities for assessing injury morphology and posterior ligamentous complex integrity of the spine according to the thoracolumbar injury

classification system and severity score. *Korean J Radiol* 2015;16:889-898
91. Lee E, Choi JA. Associations between alpha angle and herniation pit on MRI revisited in 185 asymptomatic hip joints. *Korean J Radiol* 2015;16:1319-1325
92. Kim S, Lee JW, Chai JW, Yoo HJ, Kang Y, Seo J, et al. A new MRI grading system for cervical foraminal stenosis based on axial T2-weighted images. *Korean J Radiol* 2015;16:1294-1302
93. Kim JR, Lee YS, Yu J. Assessment of bone age in prepubertal healthy Korean children: comparison among the Korean standard bone age chart, Greulich-Pyle method, and Tanner-Whitehouse method. *Korean J Radiol* 2015;16:201-205
94. Ding J, Xing W, Wu D, Chen J, Pan L, Sun J, et al. Evaluation of renal oxygenation level changes after water loading using susceptibility-weighted imaging and T2* mapping. *Korean J Radiol* 2015;16:827-834
95. Choi YJ, Baek JH, Hong MJ, Lee JH. Inter-observer variation in ultrasound measurement of the volume and diameter of thyroid nodules. *Korean J Radiol* 2015;16:560-565
96. Yun BL, Cho N, Li M, Jang MH, Park SY, Kang HC, et al. Intratumoral heterogeneity of breast cancer xenograft models: texture analysis of diffusion-weighted MR imaging. *Korean J Radiol* 2014;15:591-604
97. Lim HK, Hong SH, Yoo HJ, Choi JY, Kim SH, Choi JA, et al. Visual MRI grading system to evaluate atrophy of the supraspinatus muscle. *Korean J Radiol* 2014;15:501-507
98. Lee CB, Choi SJ, Ahn JH, Ryu DS, Park MS, Jung SM, et al. Ectopic insertion of the pectoralis minor tendon: inter-reader agreement and findings in the rotator interval on MRI. *Korean J Radiol* 2014;15:764-770
99. Ko SY, Kim EK, Kim MJ, Moon HJ. Mammographic density estimation with automated volumetric breast density measurement. *Korean J Radiol* 2014;15:313-321
100. Cho YD, Kim KM, Lee WJ, Sohn CH, Kang HS, Kim JE, et al. Time-of-flight magnetic resonance angiography for follow-up of coil embolization with enterprise stent for intracranial aneurysm: usefulness of source images. *Korean J Radiol* 2014;15:161-168
101. Yoo SY, Kim Y, Cho HH, Choi MJ, Shim SS, Lee JK, et al. Dual-energy CT in the assessment of mediastinal lymph nodes: comparative study of virtual non-contrast and true non-contrast images. *Korean J Radiol* 2013;14:532-539
102. Shin SM, Kim WS, Cheon JE, Kim HS, Lee W, Jung AY, et al. Bronchopulmonary dysplasia: new high resolution computed tomography scoring system and correlation between the high resolution computed tomography score and clinical severity. *Korean J Radiol* 2013;14:350-360
103. Kang Y, Choi JA, Chung JH, Hong SH, Kang HS. Accuracy of preoperative MRI with microscopy coil in evaluation of primary tumor thickness of malignant melanoma of the skin with histopathologic correlation. *Korean J Radiol* 2013;14:287-293
104. Zhao F, Deng M, Yuan J, Teng GJ, Ahuja AT, Wang YX. Experimental evaluation of accelerated T1rho relaxation quantification in human liver using limited spin-lock times. *Korean J Radiol* 2012;13:736-742
105. Song SE, Seo BK, Yie A, Ku BK, Kim HY, Cho KR, et al. Which phantom is better for assessing the image quality in full-field digital mammography?: American College of Radiology Accreditation phantom versus digital mammography accreditation phantom. *Korean J Radiol* 2012;13:776-783
106. Seok JH, Choi HS, Jung SL, Ahn KJ, Kim MJ, Shin YS, et al. Artificial luminal narrowing on contrast-enhanced magnetic resonance angiograms on an occasion of stent-assisted coiling of intracranial aneurysm: in vitro comparison using two different stents with variable imaging parameters. *Korean J Radiol* 2012;13:550-556
107. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539-549
108. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59:1033-1039
109. Kim JS, Jang HY, Park SH, Kim KJ, Han K, Yang SK, et al. MR enterography assessment of bowel inflammation severity in crohn disease using the MR index of activity score: modifying roles of DWI and effects of contrast phases. *AJR Am J Roentgenol* 2017;208:1022-1029
110. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology standards for quantitative imaging biomarkers. *Radiology* 2015;277:813-825
111. Barnhart HX, Barboriak DP. Applications of the repeatability of quantitative imaging biomarkers: a review of statistical analysis of repeat data sets. *Transl Oncol* 2009;2:231-235
112. Seo N, Park SH, Kim KJ, Kang BK, Lee Y, Yang SK, et al. MR enterography for the evaluation of small-bowel inflammation in crohn disease by using diffusion-weighted imaging without intravenous contrast material: a prospective noninferiority study. *Radiology* 2016;278:762-772