Diagn Interv Radiol 2025; DOI: 10.4274/dir.2024.242702

BREAST IMAGING

ORIGINAL ARTICLE

Evaluation of guided reporting: quality and reading time of automatically generated radiology report in breast magnetic resonance imaging using a dedicated software solution

Martin H. Maurer[1]
Daniel Lorenz[2]
Maximilian Clemens Otterbach[1]
Igor Toker[2]
Alexander Huppertz[2,3]

[1]Universitätsinstitut für Diagnostische und Interventionelle Radiologie, Klinikum Oldenburg AöR, Department of Diagnostic and Interventional Radiology, Oldenburg, Germany

[2]Neo Q Quality in Imaging GmbH, Berlin, Germany

[3]University of Potsdam, Center of Sports Medicine, University Outpatient Clinic, Potsdam, Germany

PURPOSE

Unstructured, free-text dictation (FT), the current standard in breast magnetic resonance imaging (MRI) reporting, is considered time-consuming and prone to error. The purpose of this study is to assess the usability and performance of a novel, software-based guided reporting (GR) strategy in breast MRI.

METHODS

Eighty examinations previously evaluated for a clinical indication (e.g., mass and focus/non-mass enhancement) with FT were reevaluated by three specialized radiologists using GR. Each radiologist had a different number of cases (R1, n = 24; R2, n = 20; R3, n = 36). Usability was assessed by subjective feedback, and quality was assessed by comparing the completeness of automatically generated GR reports with that of their FT counterparts. Errors in GR were categorized and analyzed for debugging with a final software version. Combined reading and reporting times and learning curves were analyzed.

RESULTS

Usability was rated high by all readers. No non-sense, omission/commission, or translational errors were detected with the GR method. Spelling and grammar errors were observed in 3/80 patient reports (3.8%) with GR (exclusively in the discussion section) and in 36/80 patient reports (45%) with FT. Between FT and GR, 41 patient reports revealed no content differences, 33 revealed minor differences, and 6 revealed major differences that resulted in changes in treatment. The errors in all patient reports with major content differences were categorized as content omission errors caused by improper software operation (n = 2) or by missing content in software v. 0.8 displayable with v. 1.7 (n = 4). The mean combined reading and reporting time was 576 s (standard deviation: 327 s; min: 155 s; max: 1,517 s). The mean times for each reader were 485, 557, and 754 s, and the respective learning curves evaluated by regression models revealed statistically significant slopes (P = 0.002; P = 0.0002; P < 0.0001). Overall times were shorter compared with external references that used FT. The mean combined reading and reporting time of MRI examinations using FT was 1,043 s and decreased by 44.8% with GR.

CONCLUSION

GR allows for complete reporting with minimized error rates and reduced combined reading and reporting times. The streamlining of the process (evidenced by lower reading times) for the readers in this study proves that GR can be learned quickly. Reducing reporting errors leads to fewer therapeutic faults and lawsuits against radiologists. It is known that delays in radiology reporting hinder early treatment and lead to poorer patient outcomes.

CLINICAL SIGNIFICANCE

While the number of scans and images per examination is continuously rising, staff shortages create a bottleneck in radiology departments. The IT-based GR method can be a major boon, improving radiologist efficiency, report quality, and the quality of simultaneously generated data.

KEYWORDS

Breast magnetic resonance imaging, clinical informatics, quality, radiology report, radiology workflow, software-based reporting, structured reporting, workflows, human interactions

Corresponding author: Martin H. Maurer

E-mail: mhm.maurer@gmx.net

Received 16 February 2024; revision requested 07 March 2024; accepted 05 June 2024.

Epub: 02.09.2024

Publication date: 30.12.2024

DOI: 10.4274/dir.2024.242702

You may cite this article as: Maurer MH, Lorenz D, Otterbach MC, Toker I, Huppertz A. Evaluation of guided reporting: quality and reading time of automatically generated radiology report in breast magnetic resonance imaging using a dedicated software solution. *Diagn Interv Radiol.* 2025;31(1):19-28.

Radiology reporting has not changed significantly since the beginning of the 20th century. By general consensus, radiologists should communicate unambiguously with referring physicians according to the six C's of communication: clarity, correctness, concision, completeness, consistency, and a high level of confidence.[1] Few linguistic guidelines exist for the structure of radiological findings.[2,3] Most radiology reports are crafted by free-text dictation (FT), whereby individual styles of speech may not meet the expectations of a referring physician.[4,5] In addition, a high percentage of reports contain errors,[6] with the amount of radiology errors recently assessed as 3%–5%.[7] Errors can occur before, during, or after the reporting period, may cause direct, indirect, permanent, or temporary harm to patients,[8] and can be classified into 12 subgroups.[9] Contrary to the low incidence of "poor communication-related error" reported by Kim and Mansfield[9] 2014, Brady[7] concludes that poorly written or incoherent reports are a significant source of potential harm to patients.

Additionally, the time it takes radiologists to generate reports[10,11] creates a bottleneck. Waiting for imaging and its accompanying report affects length of hospital stay and the quality of patient care. In a Canadian evaluation, each additional hour spent waiting for magnetic resonance imaging (MRI) increased acute length of hospital stay by 1.2 h.[12] Modern MRI scanners produce increasingly detailed images in ever shorter times, often making image acquisition faster than analysis and reporting.

Main points

- The software-based guided reporting (GR) strategy is a novel technology for structured reporting in medicine.
- GR allows for complete radiology reporting with minimized error rates and reduced reading times.
- The shortening of reporting times is key for successful implementation of GR into the clinical workflow.
- Differences in content between GR and unstructured, free-text dictation are not caused by limitations of the software itself but by insufficient, user-dependent operation of the software. Thus, the introduction of this disruptive technology requires intensive training and adjustment to the process.
- The great potential of structured datasets will open doors for the future of radiology with respect to big data analysis, automatic, real-time International Statistical Classification of Diseases and Related Health Problems 10 coding, and the efficient integration of artificial intelligence.

Modern radiology reporting systems are expected to reduce error rates, increase comprehensibility, and shorten report generation times. Templates and structured reporting (SR) methods do not meet all these requirements and have been applied in clinical routine only to a limited extent.[13-15] Guided reporting (GR) is a new strategy for radiology reporting. To the best of our knowledge, this study is the first to investigate and evaluate the performance of GR in breast MRI. Our purpose is to overcome the known disadvantages of SR, which include greater time investment and effort, inconsistencies in interpretation, and poor user experiences (UX).[16]

Methods

Study design

The Ethics Committee of Carl von Ossietzky University in Oldenburg, Germany approved the protocol for this retrospective study (project no: 2023-217, date: 20.11.2023). This is an experimental retrospective study conducted to assess the usability, report quality, and reading time of GR software (RadioReport®) in breast MRI (Figure 1). The study used a pre-market version (v. 0.8) of the MR mammography module to validate the usability of the software and identify further areas for improvement needed for the finalization of a market version. Eighty MRI examinations covering the full spectrum of clinical breast MRI findings [mass (n = 57), focus or non-focus enhancement (NME) (n = 8), and exclusion of a mass (n = 15)] were selected from the department of radiology at University Hospital Oldenburg (Oldenburg, Germany). Broad consent for retrospective evaluation is routinely obtained at the study hospital. These cases had been reported more than 8 weeks prior using FT and were reevaluated by the same radiologist.

The GR evaluation in this study was completed by three radiologists specialized in breast diagnostics with 16, 13, and 11 years of experience in breast MRI. The three radiologists completed a 90-min online training session on how to use the software. Application support was available by phone. Readers were instructed to begin the image reading and reporting processes simultaneously. Reader 1 had 24 cases, Reader 2 had 20 cases, and Reader 3 had 36 cases. The number of orthographic errors in the FT and GR reports was evaluated.

Guided reporting software

The pre-market v. 0.8 of the GR software was used as a stand-alone web tool. The software automatically generates radiological reports based on decoded complex radiological decision trees. A comprehensive query system (Figure 2) guides the user through the reporting process, and a dedicated module for breast MRI is available. The program includes mandatory fields, predefined paths for information input, and plausibility checks, thus guaranteeing completeness in

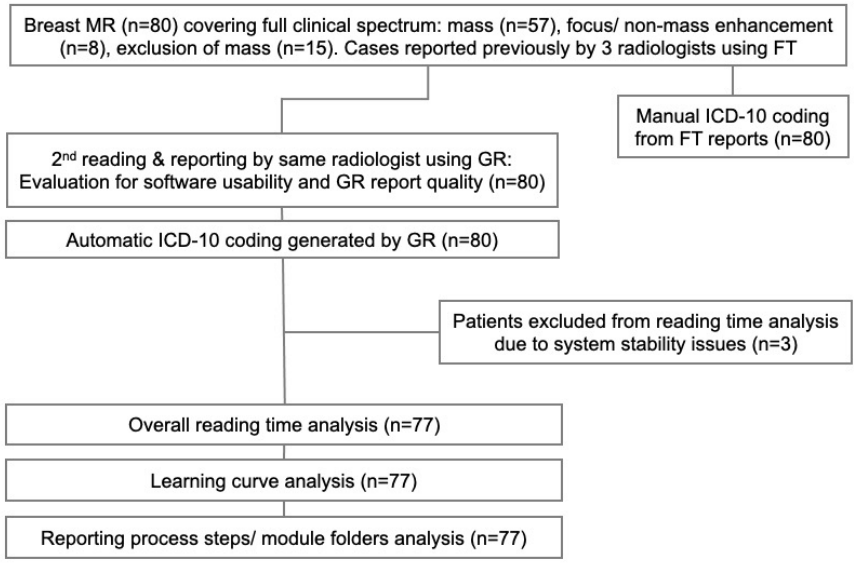
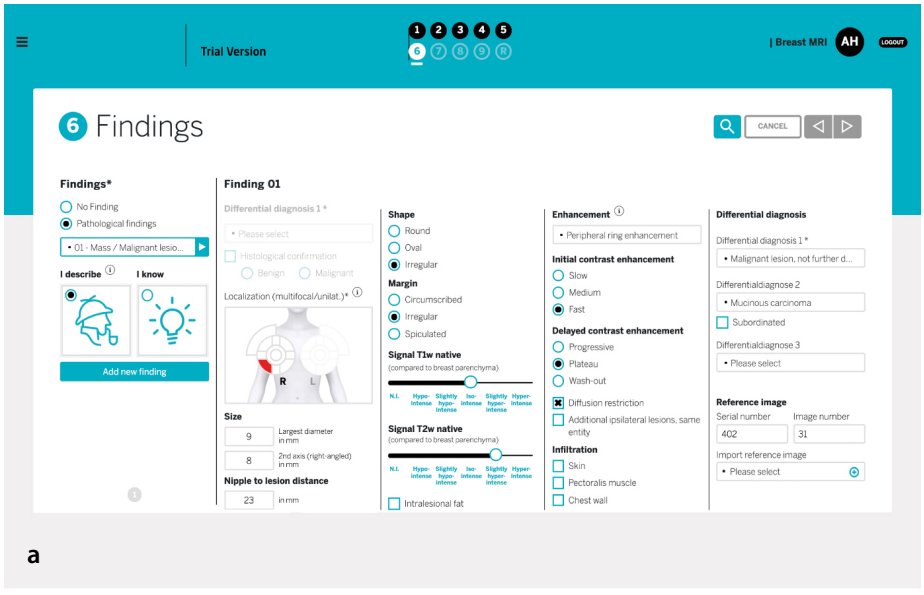


Figure 1. Study flowchart. FT, free-text dictation; GR, guided reporting; ICD-10, International Statistical Classification of Diseases and Related Health Problems.

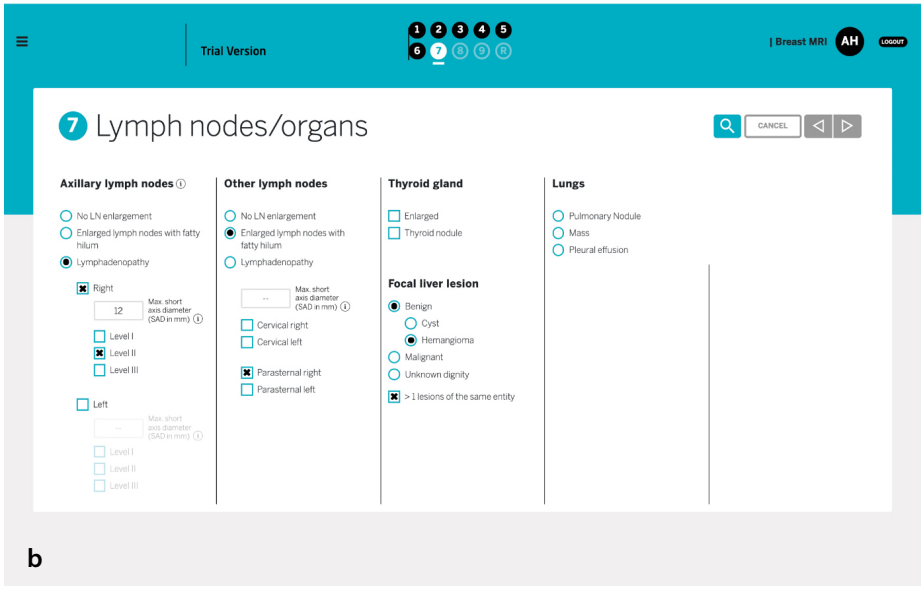
reporting. The radiologist is guided through 10 folders (patient/indication, anamnesis, technique, anatomy, background parenchymal enhancement, findings, lymph nodes/organs, bones, summary, and report). A key difference between GR and SR is the former's anatomy-based approach: every body region and modality is covered in a single module [e.g., thorax computed tomography (CT), abdomen MRI, breast MRI]. Each module covers all relevant pathologies for its specific anatomical region (e.g., tumors, inflammation, post-surgery pathologies), and a combination of different pathologies can be easily reported (rather than having to combine different templates for a single body region).[17]

Instead of having to offer and continuously update hundreds of templates, almost all MRI, CT, and breast imaging indications can be reported using 23 modules.

Each folder (e.g., "axillary lymphadenopathy" in the lymph nodes/organs folder; Figure 2b) opens with a mouse click and displays further items on the next level of depth (e.g., right/left). Each sub-item consists of several subordinated levels (e.g., level I/II/III). Detailed descriptions can be added (e.g., max. short-axis diameter in mm). Moreover, the software features information files. By selecting different items and sub-items, the software concomitantly generates semantic sentences from predefined and approved text phrases stored in its database.



a



b

Figure 2. Screenshots of guided reporting software graphic interface. **(a)** Folder 6, featuring the reporting of a finding (a mass in the right lower outer quadrant). The complex query system, with selections on different levels and entry for embedded reference images is depicted. **(b)** Folder 7, featuring the reporting of right axillary lymphadenopathy. The query system, with selections on four levels (score, site, size, and level) is depicted.

Additional information can be entered manually in the Discussion section (using speech recognition or keyboard entry) to enable a hybrid approach,[18] wherein GR with fully machine-readable data and individual information entries are combined to enhance flexibility and acceptance. The commercial and European conformity (CE)-certified version (v. 1.7) of the software includes an add-on of bullet points in the Impression section (i.e., to highlight the exclusion of suspected pathologies) (Figure 3).

With a final click, a fully automated text structured in standard compliant sections (e.g., patient data, indication, anamnesis, technique, findings, impression, and recommendations; Figures 3, 4) is generated. Diagnoses are automatically coded according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10).[19]

For the second phase of the study, v. 1.7 of the software was used. Here, technical and medical feedback was implemented, and the complexity of the breast MRI module was increased from 33,756 lines of code in v. 0.8 to 57,213 in v. 1.7 (front- and backend). Five hundred-thirteen logic lines were used in v. 1.7.

Evaluation

Usability and guided reporting quality

Usability was evaluated based on feedback from every reader along three parameters: intuitive operation, the practicability of simultaneous image analysis and reporting, and user confidence in the automatically generated report compared with FT (assessed on a 4-point scale: very good, good, medium, and poor). Readers were asked if they experienced a change in confidence while operating the software and if they would specifically recommend GR for certain subgroups.

Report quality was evaluated by comparing the completeness of the automatically generated GR product with its FT counterpart, as well as noting any differences between the two strategies. This evaluation was performed by consensus by two senior radiologists who organized the reports into three scores. A report received a score of A if there were no differences in content between FT and GR, a score of B if there were only minor differences in content that did not result in differences in treatment, and a score of C if there were major differences in content that resulted in differences in treat-

ment. Entries in the Discussion section were not included in this evaluation.

Content omission errors (COEs) in GR, v. 0.8, for patient reports receiving a score of B or C were further evaluated and categorized as follows: category 1 errors were caused by insufficient operation of the software; category 2 errors were caused by missing content in the v. 0.8 software displayable with v. 1.7; and category 3 errors were caused by missing content in the v. 0.8 software that was not displayable with v. 1.7. Errors in reporting were further categorized according to a modified classification system proposed by Hawkins et al.[6] as follows: (a) non-sense; (b) spelling/grammar; (c) omission/commission; and (d) translation (Table 1).

Reporting process and time

Reading time (s) was defined as the overall combined image reading time and report creation time, evaluated based on the implemented tracking option. Individual use of information files was tracked automatically. Overall reading times were compared between different readers as well as different lesions. The learning curve for reading time over the study period was evaluated for every reader using a multiple regression analysis. The reporting process was additionally analyzed on a segmental (folder) level and compared on an interindividual basis. FT was not evaluated in this portion of the study; however, benchmark data from large radiological information system (RIS) evaluations (n = 170,901 reports, including 5,622 MRI reports) were used as reference.[10] These benchmark data were extracted from two large teaching hospitals with 23 full-time radiologists in New Zealand. Outliers with exceptionally long reporting times (>60 min for MRI) were culled, removing 9.5% of the total. Automatically generated GR ICD-10 codes[19] were compared with manual FT coding by a senior radiologist (A.H.).

Statistical analysis

Differences in reading duration by number of cases per reader were assessed using a trend test. A regression analysis was used to assess trends, first by reader and then using a mixed model for repeated measures (MMRM). In the MMRM, the data of all three readers were used within the same model, with the factor, "reader", used to indicate repeated measures. For sensitivity analysis in case of abnormality, a non-parametric Jonckheere–Terpstra trend test was used. Analyses were performed using SAS v. 9.4 software (SAS Institute Inc., Cary, NC, USA).

Results

Usability was rated high by all readers. Intuitive operation was rated very good by two readers and good by one reader. The coworking of image analysis and reporting was rated very good by all readers, and confidence in the automatically generated report compared with its FT counterpart was rated good. All readers reported increasing confidence during the study.

With respect to content differences between GR and FT, 41 patient reports (51.3%) received an A score (no differences in content), 33 patient reports (41.3%) received a B score (minor differences in content not resulting in differences in treatment), and 6 patient reports (7.5%) received a C score (major differences in content resulting in differences in treatment) (Table 2).

Of the 39 patient reports scoring B or C, 26 (66.7%) involved category 1 errors (COEs in v. 0.8 caused by insufficient operation of the software), and 13 (33.3%) involved category 2 errors (COEs caused by missing content in v. 0.8 of the software but displayable with v. 1.7). No category 3 errors (COEs caused by missing content in the v. 0.8 software and not displayable with v. 1.7) were identified.

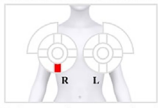
No non-sense, omission/commission, or translation errors were observed in the GR reports. Spelling/grammar errors were observed in 3/80 (3.8%) GR patient reports (exclusively in the free-text discussion section), compared with 36/80 (45%) in FT patient reports.

Figure 3. Breast magnetic resonance imaging (MRI) of a 37-year-old woman with suspected cancer in the right breast. **(a)** Subtracted, early dynamic phase from contrast-enhanced MRI. **(b)** Screenshot of folder 10: report preview (the highlighted blue text in the Impression section was entered with free-text dictation).

RadioReport® by NEO Q

Report Breast MRI. Pat-ID 080, born #
Patient data: The patient is 37 years old.
Indication: Lump/hardening in the right breast.
Anamnesis: Aunt (paternal) has a history of breast cancer.
Technique: Breast MRI with i.v. contrast agent (12.0 ml Gadobenic acid). Field strength: 3 T.

Findings:
Isomastia. No cutis thickening. Breast tissue: scattered fibroglandular tissue. Breast parenchyma has an inconspicuous appearance on T1w and T2w. Parenchyma with minimal contrast enhancement. Inconspicuous nipple on both sides.
Detection of a round mass right lower inter-quadrant.



Reference image: series 14/image 87. Diameter: 9 mm. Contrast enhancement: Peripheral ring. Initially fast contrast uptake. Subsequently detection of a plateau in the contrast dynamic. Recognition of restricted diffusion. Differential diagnosis: breast carcinoma, not further specified.

Detection of a non-mass enhancement. Localisation: right upper outer quadrant.

Segmental distribution. Differential diagnosis: breast carcinoma, not further specified.

No pathologic enlargement of axillary LN. No suspicious bone lesion.

Impression:
First breast MRI. No former MRI of the breast. Previous imaging: Former Mammography.

1. Detection of a round mass right lower inter-quadrant. Differential diagnosis: breast carcinoma, not further specified.
2. Detection of a non-mass enhancement. Localisation: right upper outer quadrant. Differential diagnosis: breast carcinoma, not further specified.
3. Not typical for a malignant phylloid tumor or granulomatous mastitis in asymptomatic patient.
4. No suspicious lesion in the left breast.
5. No suspicious lymph node.

Classification based on BI-RADS Version 4.0.
Amount of fibroglandular tissue: 2. BI-RADS (right side): 5. BI-RADS (left side): 1.
Recommendations:
Recommendation for ultrasound-guided biopsy of the right breast.

Discussion:

Question: The lump was detected in the right upper outer quadrant, scored BIRADS 0. Differential mastopathy or breast carcinoma?
Additional info: Distance between mass and NRE 10 mm. Direct contact between NRE and pectoralis muscle.

a

Report breast MRI, Pat-ID 080, Patient's date of birth: #

Question/indication
Palpable rigid lump in upper outer quadrant right breast, mammography BIRADS 0. Mastopathy, differential breast CA? Breast cancer aunt paternal side.

Technique: Breast MRI at 3 Tesla. Injection of 12 ml Multihance.

Findings
Initial examination.

Little-medium background activity.
Isomastia and symmetric nipples. Regular presentation of cutis and subcutis on both sides. Round lesion of 9 mm with ring enhancement in the right caudal breast at 6 o'clock. With a distance of approx. 10 mm normal mammary gland tissue of this extensive non-mass enhancement latero-cranially reaching far into the upper outer quadrant. The round lesion as well as the non-mass lesion mammary right show a rapid KM enhancement with plateau or an implied washout. Some small satellite-like lesions in addition, exemplified by 5 mm lesion series 14, Figure 87. Moderate diffusion restriction of these lesions. In depth, the lesion extends immediately to the pectoralis muscle.
The left breast shows no suspicious contrast enhancement or suspicious diffusion restriction.
No pathologically enlarged axillary or mammaria interna lymph nodes.
No evidence of larger round foci in the partially imaged lung.
Regular bone marrow signal of the rib thorax. Regarding the upper abdominal organs, as far as included, no major pathological changes.

Impression
1. Malignant round lesion as well as a large non-mass lesion of the right breast with pathological contrast dynamics although with only moderate quantitative diffusion restriction. Close demarcation to the skin short-stretched and more distinct to musculus pectoralis.
   Biopsy/biopsies are strongly recommended to confirm the diagnosis, primarily sonographic.
2. Not typical for a malignant phylloid tumor or granulomatous mastitis in asymptomatic patient.
4. No suspicious lesion in the left breast.
5. No suspicious lymph node.

Right: MRM-BIRADS 5
Left: MRM-BIRADS 1

b

Figure 4. Depiction of two reports of the same patient in **(a)** was created with guided reporting (GR) technology. **(b)** was created with free-text dictation (FT). Major differences include GR's form, which includes embedded key images and localizers, and the absence of spelling/grammar errors. The short and to-the-point phrases of GR are much easier to understand (especially compared with sentences 5 and 6 in the FT findings section: "Round lesion…outer quadrant", which correspond to sentences 6 and 7 in the GR Findings section: "Detection of …not further specified"). The standardized GR glossary corresponds exactly to the recommendation of the state-of-the art BI-RADS Atlas,[38] whereas FT uses terms that are not exactly defined (such as "lesion") and unclear classifications [such as "little-medium activity", which corresponds to "parenchyma with minimal contrast enhancement" (sentence 4 in the GR Findings section)]. Free-text dictation reveals additional differences. For example, in FT, the phrase, "…extends immediately to the pectoralis muscle," differs from its GR counterpart, where the choice for the pectoralis muscle is a facultative radio button with the phrase, "bilaterally circumscribed margin" *or* "muscle invasion." The radio button is a graphical control element that allows the user to choose only one of a predefined set of mutually exclusive options. Guided reporting fulfills BI-RADS recommendations exactly,[38] with the decision about "pectoralis muscle invasion" included in the section of associated features. The relative description, "suspicion of," is not possible. Other minor differences are the phrase, "plateau *or* an implied washout" in FT, as opposed to the GR facultative radio button for delayed contrast enhancement choice between "progressive," "plateau," or "wash-out." BI-RADS, Breast Imaging Reporting and Data System.

Table 1. Error categorization according to a modified segmentation from Hawkins et al.[6]

Error category	Error type	Definition	Intended phrase example	Transcribed phrase example
a	Non-sense	Passages, words, or phrases that make no sense or have no sensible meaning	No suspicious lesion in the breast	No suspension in the breast
b	Spelling/grammar (typographical and grammatical errors, homonyms, improper period usage)	Typographical errors, word misuse, duplicate periods, or lack of a period at the end of a sentence	There is a lesion in the left breast	There is a lesion in the light breast
c	c1: Omission (other)	Omitted words/phrases that do not result in a missense or non-sense error	There is no lesion in the left breast	There is no lesion left breast
	c2: Commission	A statement retained from a standardized template that contradicts dictated findings or impression	There is no lesion in the left breast	There is no lesion in the left breast. Mass in the left upper inner quadrant
d	Translational (other)	A translation error that does not result in a non-sense/missense error. The resulting sentence still has sensible meaning (as opposed to non-sense errors)	Breast parenchyma has an inconspicuous appearance	Breast parenchyma has an unobtrusive appearance

Table 2. Patient reports with a score of C: major differences in content resulting in differences in treatment

Patient; reader; case no.	MRM category; no. of lesions; lateralization	FT report content & MRM BI-RADS	GR report content & MRM BI-RADS	Difference FT vs. GR	Resulting difference in treatment	Error category	Change in GR b/t v. 0.8 and v. 1.7
34 y/o F; R3; case 5/36	Mass; 1; R	Large, centrally necrotic BC @ 12:00; diffuse, strongly enhancing parenchyma on the upper outer Q @ 9.00–10:00. Diffuse infiltration/ DCIS cannot be ruled out. BI-RADS (r/l): 6/2	Detection of mass in R upper interQ. Diagnosis: breast CA NOS. BIRADS (r/l): 6/2	Additional NME in R upper outer Q not described in GR (included in free-text "discussion").	Planning of more extended surgical resection due to additionally ipsilateral, NME.	2	Extended teaching of "loop function" (i.e., a software fxn to describe multiple lesions) during online training.
39 y/o F; R3; case 27/36	Mass; >3; R	MRM–BI-RADS 5 (or 6, if histological diagnosis available): malignant lesions, multicentric/ multifocal involvement of at least the two lower Qs. BI-RADS (l): 2	Detection of mass in R outer interQ, R lower outer Q, and R lower inner Q. Diagnosis: breast CA NOS. BI-RADS (r/l): 5/2	Explicit wording "multifocal" and "multicentric" missing in GR.	Multicentricity/ multi-focality have a (–) impact on prognosis, and more aggressive treatment options were used.[35]	2	Extended teaching of additional independent checkboxes, "multifoca" and "multicentric".
56 y/o F; R3; case 17/36	Exclusion of mass; 0; N/A	No malignant lesions on either side; no implant rupture of breast implants; no capsular fibrosis. BI-RADS (r/l): 1/1	Exclusion of a mass. BI-RADS (r/l): 1/1	Exclusion of implant rupture not explicitly expressed in GR (the automatically generated report would mention a rupture in case of a [+] finding).	Additional diagnostic examination(s) might be performed to check integrity of implants	1	In the chapter, "silicone implant", an additional radio button, "in-tact bilateral implants" added in v. 1.7
34 y/o F; R3; case 23/36	Mass; >4; bilateral	BRCA2 mutation. R: biopsy-confirmed, poorly differentiated breast CA w/ extension in all Qs; axillary lymph node metastases. L: three malignant areas. BI-RADS (r/l): 6/4	1. Detection of mass in R upper outer; R upper inner; R lower outer; and R lower inner Qs. Diagnosis: breast CA NOS; 2. Pathologic enlargement of R axillary lymph nodes; 3. Detection of mass in L lower interQ and L upper outer Q. Diagnosis: breast CA NOS. BI-RADS (r/l): 6/4	Information on BRCA mutation missing in GR.	The surgical treatment of patients with a genetic germline variant might be different from patients who are not carriers of BRCA.[36]	1	In the "indication" drop-down list, an additional criterion, "BRCA mutation/ genetically high risk", added in v. 1.7.
72 y/o F; R2; case 11/20	Mass; 2; R	Recurrence of moderately differentiated invasive breast CA 8 y after ipsilateral carcinoma. Two lesions: first lesion @ 10:00, max: 10 mm, irregular shape, margin spiculated, fast initial enhancement; second lesion @ 12:00; max: 4 mm; fast initial enhancement. BI-RADS (r/l): 6/1	1. Detection of mass in R upper outer Q. Diagnosis: invasive breast CA. 2. Detection of mass in R upper interQ. Diagnosis: invasive breast CA. BI-RADS (r/l): 6/1	Histological confirmation already available not explicitly covered in the wording of GR (though anticipated by categorization, BI-RADS 6 = known biopsy).	Next clinical step with FT would be surgery; next step with GR might be biopsy.	1	Checkbox "histological confirmation" added in v. 1.7.
77 y/o F; R3; case 35/36	NME; 1; L	Status after lobular breast CA on R with ablation, 2014; suspicion of mediastinal lymph node metastasis. L breast: newly developed, pronounced malignancy; suspicious contrast enhancement w/ suspicion of upper outer Q lobular CA. BI-RADS (r/l): 0/4	1. Detection of a NME L upper outer Q. Differential diagnosis: breast CA NOS. 2. Mediastinal lymph node metastasis. BI-RADS (r/l): 0/4	(Highly probable, from patient history) specific histology (= invasive lobular CA) of malignant lesion missing in GR.	Differences in clinical mgmt b/t invasive lobular and invasive ductal BC.[37]	1	Checkbox "histological confirmation" added/differential diagnoses of NME further extended to provide specific histology. In v. 1.7, different, more specific lesion types, including invasive lobular and invasive ductal breast CA, added to the drop-down list.

BRCA2, breast cancer gene 2; MRM, magnetic resonance mammography; BI-RADS, Breast Imaging Reporting and Data System; BC, breast cancer; CA, carcinoma; FT, free-text dictation; GR, guided reporting; Q, quadrant; NOS, not otherwise specified; DCIS, ductal carcinoma in situ; NME, non-mass enhancement.

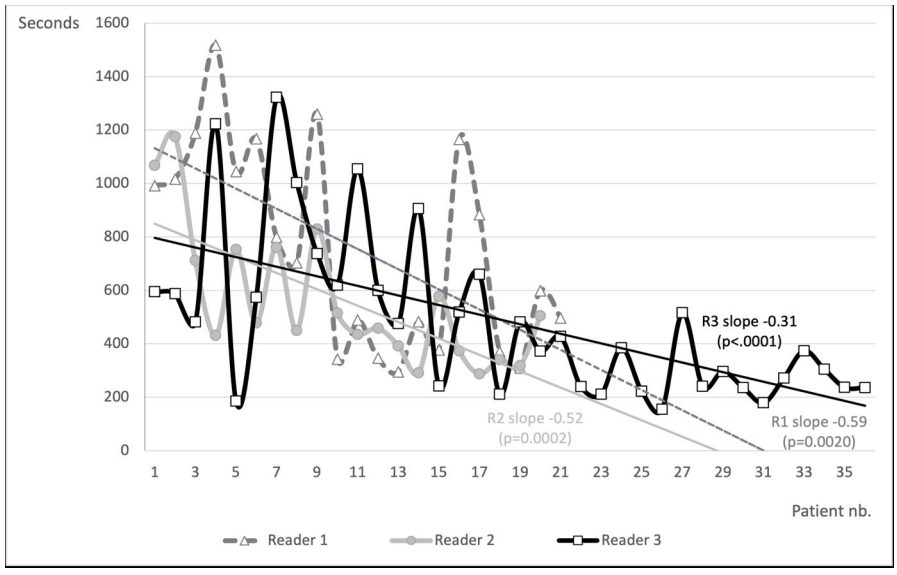


Figure 5. Reading duration for Readers 1, 2, and 3 with trend lines (number of cases per reader differed). The most significant learning curve occurred with Reader 1: with 21 reports, reading time decreased from 992–1,517 s for each of the first five reports to a minimum of 295 s by the 13th.

Three patients were excluded from the overall reading time analysis due to software issues (system instability resolved by remote maintenance). The mean reading time was 9 min 36 s (standard deviation: 5 min 27 s; min: 2 min 35 s; max: 25 min 17 s). The mean reading time per reader was 12 min 34 s (Reader 1), 9 min 17 s (Reader 2), and 8 min 5 s (Reader 3). The mean reading time for reports on patients without lesions was 11 min 16 s, 9 min 31 s for reports on patients with mass(es), and 7 min 31 s for reports on patients with focus/NME. The overall reading time was significantly shorter than FT radiology report reference values. The mean reading time for MRI examinations in general was 17 min 23 s (assessed from 1,629 h for 5,622 examinations).[10] Thus, compared with MRI examinations in general, reading time was shortened by 44.8%. More specifically, for breast MRI, a study-ascribable time (SAT) of 35 min was reported.[11] In contrast to the reading time analyzed in,[10] the SAT includes additional steps to reading and reporting, such as interpretation and clarification of request, prescription of the examination protocol, and communication with the referrer.[11]

The information fields in the software were unsurprisingly limited to two uses for one reader, as the users were highly specialized in breast imaging.

A learning curve with significant streamlining of the reading process (assessed by time to completion) during the training course was observed for all readers (Figure 5). Regression models show reading time learning curve slopes of –0.59 (P = 0.002) for Reader 1, –0.52 (P = 0.0002) for Reader 2, and –0.31 (P < 0.0001) for Reader 3, with an average slope of –0.36 (P < 0.0001). As a certain deviance from normality was found in the data, a sensitivity analysis was performed using a non-parametric Jonckheere–Terpstra trend test. This test supported the results of the parametric regression analyses (Reader 1: P = 0.0157; Reader 2: P = 0.0015; Reader 3: P < 0.0001). A detailed analysis of the steps in the reporting process revealed differences between the readers (Figure 6). Reader 2 invested the greatest time allotment (38.4%) (Figure 6b) to folders 1 and 2 (patient/indication and anamnesis) of the GR process (compared with 33.3% and 27.9%, for Readers 1 and 3, respectively) (Figure 6a, c), suggesting that Reader 2 analyzed the image dataset in detail before beginning the reporting process. The relative time allotment for folder 6 (findings, which is the crux of a MR mammography report) was longer for Reader 3 (28.8%) compared with Readers 1 and 2 (21.9% and 21.0%, respectively). For folder 10 (report), the relative time allotment was longer for Reader 2 (17.6%) compared with Readers 1 and 3 (9.0% and 9.3%, respectively). This folder is used to check the text of the automatically generated report and the free-text entry of the discussion section, meaning that Readers 1 and 3 spent objectively less time in the acceptance or modification of the report text generated by the software than Reader 2.

The ICD-10 coding of the automatically generated GR reports was identical to manual FT coding and included the following codes: C22.9 (malignant neoplasm of the liver, not specified as primary or secondary; n = 1), C50.11 (malignant neoplasm of the central portion of the breast, n = 2), C50.21 (malignant neoplasm of the upper, inner quadrant; n = 2), C50.31 (malignant neoplasm of the lower, inner quadrant; n = 1), C50.41 (malignant neoplasm of the upper, outer quadrant; n = 7), C50.51 (malignant neoplasm of the lower, outer quadrant; n = 4), C50.81 (malignant neoplasm of overlapping sites of the breast; n = 34), C50.9 (malignant neoplasm of unspecified site of the breast; n = 1), D13.4 (benign neoplasm of the liver; n = 1), D24 (benign neoplasm of the breast; n = 1), N60.1 (diffuse cystic mastopathy; n = 3), N62 (hypertrophy of the breast; n = 2), N63 (unspecified lump in the breast; n = 15), R59.0 (enlarged lymph nodes; n = 17), T85.4 (mechanical complication of breast prosthesis and implant; n = 4), Z80.3 (family history of malignant neoplasm of the breast; n = 7), and Z85.3 (personal history of malignant neoplasm of the breast; n = 20). In total, 122 ICD-10 codes were generated.

Discussion

Differences between guided and structured reporting and the clinical need for guided reporting technology

GR is a modern, information technology (IT)-based solution developed to improve the workflow and quality of radiology reports. A standalone version of the software was used in this study, as opposed to integration into existing IT that occurs in clinical routine (e.g., picture archiving and communication systems and RIS. The benefits of SR have been reported in several publications.[20,21] Traditional narrative reporting is associated with high variability and is prone to error. According to Hawkins et al.[6], 41% of radiology reports contain errors, and 33% contain non-grammatical errors that can lead to quality issues in patient care (e.g., errors in electronic patient records) and poor-quality routine data for research. SR allows for a significant improvement in the quality of written reports. Text is generated automatically based on information input and is edited in a concise and standardized language. SR may be easy to implement into clinical workflow for uncomplicated cases; however, patients with complex pathologies involving different diseases (e.g., breast carcinoma combined with a benign bone tumor in the humerus)

with a wide range of findings require greater flexibility. Thus, percentages of unstructured free-text reporting in clinical settings will remain comparatively high.[20,21]

GR is a further development of SR. Major differences include reporting by anatomical region, rather than by pathology, with a manageable number of 23 modules covering the whole spectrum of clinical MRI and CT examinations (plus X-ray mammography and breast ultrasound); the mapping of a complex cognitive decision tree developed by imaging specialists with a predefined point-by-point approach guiding radiologists through the reporting process, and the conscious exclusion of free text in the findings section and in the core components of the Impression section. The software provides error-free report texts: non-grammatical omission errors are reduced by the implementation of mandatory fields in the most relevant parts of the decision tree; plausibility checks reduce the risk of missense errors due to human failure; an intuitive and clear UX design focuses on risk reduction of commission errors; and intrinsic contradictions within the imaging report (e.g., between the findings and the impression sections) are prevented by waiving free-text entries. Digumarthy et al.[22] reported that 33 of 47876 radiology reports (876; 0.0007%) contained side discrepancies between the findings and impression sections. These discrepancies were revealed to be uncommon but quantifiable. Most discrepancies occurred in complex radiology reports involving multiple bilateral lesions with numerous citations of lateralization used to describe lesion distribution and location.[22]

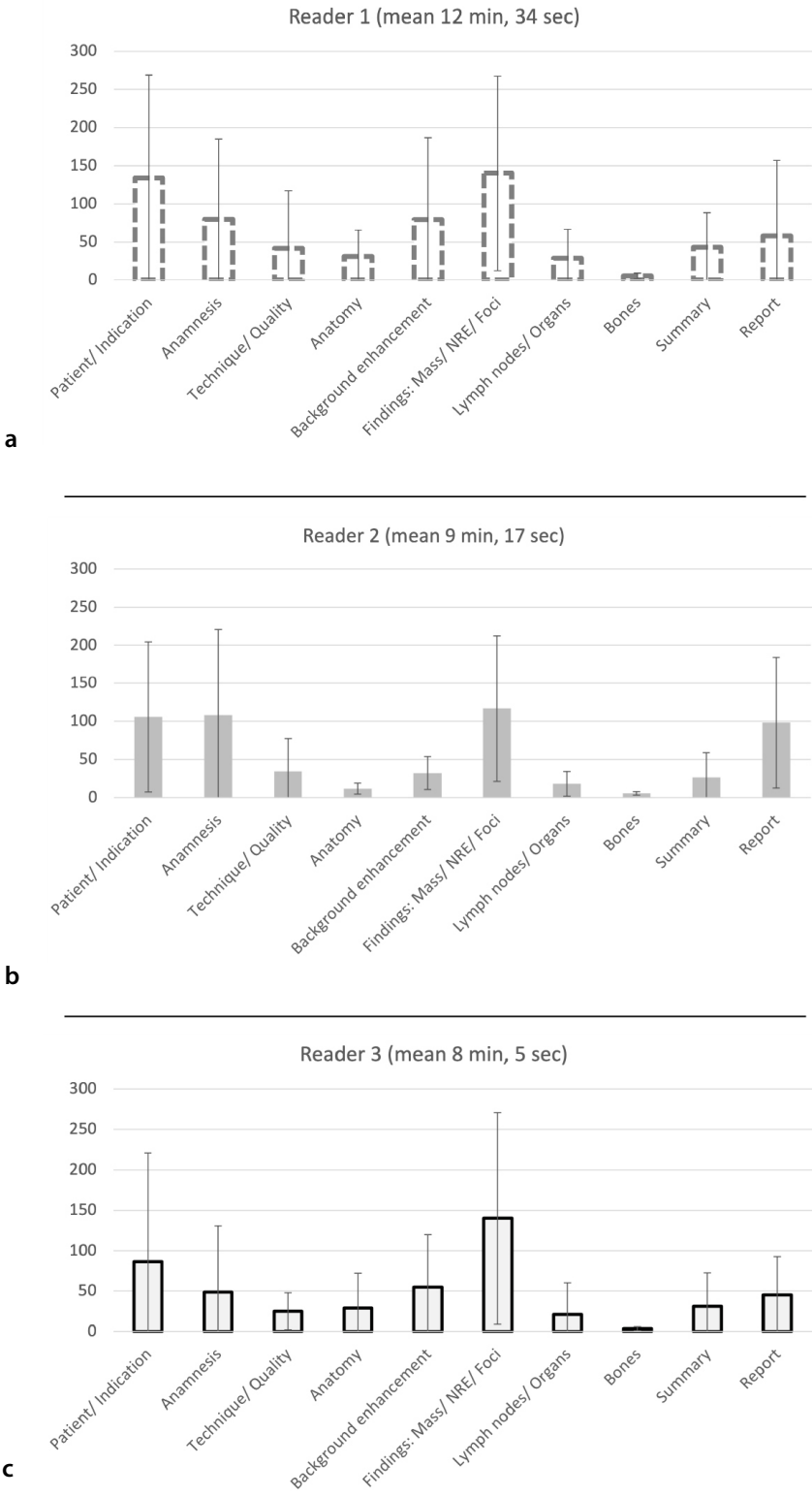


Figure 6. Evaluation times on a folder level; representations of means and standard deviations.

Usability and performance of guided reporting: learning how to use the technology

Our starting hypothesis that the usability of GR software is intuitive and enables radiologists to produce high-quality breast MRI reports was confirmed by the high ratings of all readers. Of the 80 reports, 51.3% revealed no content differences between FT and GR, 41.3 % contained minor differences in content between FT and GR not resulting in differences in treatment, and 7.5% contained major differences in content resulting in differences in treatment. All those reports containing major differences were COEs caused either by insufficient operation of the software or by an omission in the v. 0.8 software prototype displayable with the commercially available v. 1.7. Analysis revealed the need for extended training in 2/6 cases with the lowest score C (major differences in content, resulting in differences in treatment) and for additions to the decision tree in 4/6 cases. The importance of intensive training in the novel technology was clear from our study results. In conclusion, GR reports are complete, error-free, and without contradictions between the Findings and Impression sections. The error rate of 45% (mostly spelling and grammatical) in our evaluation of free-text narrative reporting by dictation can be lowered with the use of GR. Attached images and localizers within the GR system increase

comprehensibility for referring physicians. It had been shown previously that reports with embedded images save time, increase physician confidence in treatment decisions, and may influence patient management.[23,24]

Increased reporting speed

The major attraction for radiologists in the use of GR in their daily routines is the reduction in reporting time. Staff shortages are a current bane in radiology; the number of scans and images per examination are rapidly rising, while the number of specialist radiologists remains more or less stagnant. In the U.S., MRI and CT rates increased by 124% and 110%, respectively, between 2000 and 2016.[25] In contrast, the total number of radiologists increased by only 39% between 1995 and 2011.[26] Our study was able to demonstrate a shortening of the combined reading and reporting time with GR compared with reference values from a narrative reporting process. The mean reading time for general MRI examinations was 17 min 39 s[10] and 35 min for breast MRI specifically.[11] Radiologist reporting times are a key component of radiology department workload assessment. Although of pivotal importance for management decisions, reliable measurements remain challenging.[27] Different methods have been described. Site-specific median reporting times (MRTs) are considered more precise than the SATs[27] introduced by Pitman et al.[11] in 2018. As the MRT described by Cowan et al.[10,27] was very similar to our site-specific method, it was selected as a comparator. Reading time was decreased by 44.8% (general MRI). The learning curve for all readers promises great time and cost savings potential.

The necessity of leaving prosaic radiology reporting

Numerous studies have found that structured, disease-focused dictation templates improve the comprehensiveness and appropriateness of radiologist reports.[18,20,21,28,29] Structure in reporting can help reduce errors (particularly typographical errors), train residents in the evaluation of complex examinations, and assist residents in recalling critical findings.[29] Structure in reporting is recommended in many MRI indications.[30,31]

In accordance, the following recommendation was published by the European Society of Radiology in 2018: "moving from conventional prose reports to SR is endorsed as a positive development, and must be an international effort, with international design and adoption of SR templates that can be translated and adapted in local environments as needed. Industry involvement is key to success, based on international data standards and guidelines".[32] In addition, IT-based reporting tools, and especially novel GR technology, are the basis for additional export formats of structured data (i.e., direct ICD-10 coding), as used in our study. Patient-friendly reporting, created with a single click, is essential to providing patient-centered and value-based care in the radiology of tomorrow.[33]

The limitations of our study include the setting of FT as the gold standard, preventing intra-individual comparison between FT and GR regarding reporting time, as well as its relatively small sample size and single-institution, retrospective design. Future studies with double reading and reporting in a randomized order using the commercial, CE-certified version of GR software are needed to validate the reduced combined reading and reporting time and to quantify reduced radiology turnaround times. An international multicenter design with a wider range of reader experiences would provide further insight into workflow improvements in different regional and education settings. An evaluation of all readers reading all to assess intra- and inter-reader reliability and variability was not performed and should be conducted in future studies to confirm the robustness of the advantages of GR technology. In addition, testing for missense errors in GR reports, and how they compare with those in FT reports, would need direct comparative analysis, with the images as a gold standard. Missense errors as defined by Hawkins et al.[6] (including their subclassifications into E1: translation errors, E2: errors of omission, and E3: human error, any of which could change the meaning of a phrase/sentence) could not be detected in our evaluation because FT was used as the gold standard.

In conclusion, GR allows for complete radiology reporting with a minimized error rate and a reduced combined reading and reporting time. Introducing this disruptive reporting strategy into clinical workflow requires management adjustments and intensive training, as has been emphasized by reports showing minor differences between GR and FT that are not caused by limitations of the software itself but by user-dependent operation. Shortening of reporting time is key for successful implementation into the clinical workflow. The great potential of structured datasets will open the doors for the future of radiology with respect to big data analysis, automatic ICD coding, and efficient artificial intelligence development. SR, and GR even more so, has the potential to facilitate developments in machine learning for radiological applications.[34]

Acknowledgements

The authors would like to acknowledge Silvia Huppertz for her intensive and selfless support, not only in writing this manuscript but also throughout the last years. Silvia passed away in April 2024, and we miss her deeply. The authors would also like to thank Carsten Schwenke for his assistance in the statistical methods and evaluation used herein.

Conflict of interest disclosure

Authors Igor Toker, Daniel Lorenz, and Alexander Huppertz declared that they are full-time or part-time employees of Neo Q Quality in Imaging GmbH. The other authors declared no conflict of interest.

References

1. Armas RR. Letter: qualities of a good radiology report. *AJR Am J Roentgenol*. 1998;170(4):1110. [CrossRef]
2. ACR Practice parameter for communication of diagnostic imaging findings. Accessed May 9, 2024. [CrossRef]
3. Hartung MP, Bickle IC, Gaillard F, Kanne JP. How to create a great radiology report. *Radiographics*. 2020;40(6):1658-1670. [CrossRef]
4. Ignácio FCGR, de Souza LRMF, D'Ippolito G, Garcia MM. Radiology report: what is the opinion of the referring physician? *Radiol Bras*. 2018;51(5):308-312. [CrossRef]
5. Gunn AJ, Alabre CI, Bennett SE, et al. Structured feedback from referring physicians: a novel approach to quality improvement in radiology reporting. *AJR Am J Roentgenol*. 2013;201(4):853-857. [CrossRef]
6. Hawkins CM, Hall S, Zhang B, Towbin AJ. Creation and implementation of department-wide structured reports: an analysis of the impact on error rate in radiology reports. *J Digit Imaging*. 2014;27(5):581-587. [CrossRef]
7. Brady AP. Error and discrepancy in radiology: inevitable or avoidable?. *Insights Imaging*. 2017;8(1):171-182. [CrossRef]
8. Onder O, Yarasir Y, Azizova A, Durhan G, Onur MR, Ariyurek OM. Errors, discrepancies and underlying bias in radiology with case examples: a pictorial review. *Insights Imaging*. 2021;12(1):51. [CrossRef]
9. Kim YW, Mansfield LT. Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors. *AJR Am J Roentgenol*. 2014;202(3):465-470. [CrossRef]

10. Cowan IA, MacDonald SL, Floyd RA. Measuring and managing radiologist workload: measuring radiologist reporting times using data from a radiology information system. *J Med Imaging Radiat Oncol*. 2013;57(5):558-566. [CrossRef]
11. Pitman A, Cowan IA, Floyd RA, Munro PL. Measuring radiologist workload: progressing from RVUs to study ascribable times. *J Med Imaging Radiat Oncol*. 2018;62(5):605-618. [CrossRef]
12. Bartsch E, Shin S, Roberts S, et al. Imaging delays among medical inpatients in Toronto, Ontario: A cohort study. *PLoS One*. 2023;18(2):e0281327. [CrossRef]
13. Sistrom CL. Conceptual approach for the design of radiology reporting interfaces: the talking template. *J Digit Imag*. 2005;18(3):176-187. [CrossRef]
14. Jorg T, Kämpgen B, Feiler D, et al. Efficient structured reporting in radiology using an intelligent dialogue system based on speech recognition and natural language processing. *Insights Imaging*. 2023;14(1):47. [CrossRef]
15. European Society of Radiology (ESR). ESR paper on structured reporting in radiology-update 2023. *Insights Imaging*. 2023;14(1):199. [CrossRef]
16. Pesapane F, Tantrige P, De Marco P, et al. Advancements in standardizing radiological reports: a comprehensive review. *Medicina (Kaunas)*. 2023;59(9):1679. [CrossRef]
17. Nobel JM, van Geel K, Robben SGF. Structured reporting in radiology: a systematic review to explore its potential. *Eur Radiol*. 2022;32(4):2837-2854. [CrossRef]
18. Larson DB, Towbin AJ, Pryor RM, Donnelly LF. Improving consistency in radiology reporting through the use of department-wide standardized structured reporting. *Radiology*. 2013;267(1):240-250. [CrossRef]
19. ICD-10-CM Official Guidelines for Coding and Reporting. FY 2021 – UPDATED January 1, 2021. assessed May 9, 2024. [CrossRef]
20. Pinto Dos Santos D, Hempel JM, Mildenberger P, Klöckner R, Persigehl T. Structured Reporting in Clinical Routine. *Rofo*. 2019;191(1):33-39. [CrossRef]
21. Ganeshan D, Duong PT, Probyn L, et al. Structured reporting in radiology. *Acad Radiol*. 2018;25(1):66-73. [CrossRef]
22. Digumarthy SR, Vining R, Tabari A, et al. Process improvement for reducing side discrepancies in radiology reports. *Acta Radiol Open*. 2018;7(7-8):2058460118794727. [CrossRef]
23. Iyer VR, Hahn PF, Blaszkowsky LS, Thayer SP, Halpern EF, Harisinghani MG. Added value of selected images embedded into radiology reports to referring clinicians. *J Am Coll Radiol*. 2010;7(3):205-210. [CrossRef]
24. Patel BN, Lopez JM, Jiang BG, Roth CJ, Nelson RC. Image-Rich radiology reports: a value-based model to improve clinical workflow. *J Am Coll Radiol*. 2017;14(1):57-64. [CrossRef]
25. Radiology Business News, September 04, 2019. Imaging utilization continues to rise-should that be viewed as a negative? [CrossRef]
26. Rosenkrantz AB, Hughes DR, Duszak R Jr. The U.S. Radiologist workforce: an analysis of temporal and geographic variation by using large national datasets. *Radiology*. 2016;279(1):175-184. [CrossRef]
27. Cowan IA, Floyd RA. Measurement of radiologist reporting times: assessment of precision, comparison of three different measurement techniques and review of potential applications. *J Med Imaging Radiat Oncol*. 2023;67(7):734-741. [CrossRef]
28. Dickerson E, Davenport MS, Syed F, et al. Effect of template reporting of brain MR imagings for multiple sclerosis on report thoroughness and neurologist-rated quality: results of a prospective quality improvement project. *J Am Coll Radiol*. 2017;14(3):371-379. [CrossRef]
29. Johnson TF, Brinjikji W, Doolittle DA, Nagelschneider AA, Welch BT, Kotsenas AL. Structured head and neck CT angiography reporting reduces resident revision rates. *Curr Probl Diagn Radiol*. 2019;48(2):114-116. [CrossRef]
30. Beets-Tan RGH, Lambregts DMJ, Maas M, et al. MR imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol*. 2018;28(4):1465-1475. [CrossRef]
31. Maas M, Dijkhoff RAP, Beets-Tan R. Rectal cancer: assessing response to neoadjuvant therapy. *Magn Reson Imaging Clin N Am*. 2020;28:117-126. [CrossRef]
32. European Society of Radiology (ESR). ESR paper on structured reporting in radiology. *Insights Imaging*. 2018;9(1):1-7. [CrossRef]
33. Rockall AG, Justich C, Helbich T, Vilgrain V. Patient communication in radiology: moving up the agenda. *Eur J Radiol*. 2022;155:110464. [CrossRef]
34. Pinto Dos Santos D, Baeßler B. Big data, artificial intelligence, and structured reporting. *Eur Radiol Exp*. 2018;2(1):42. [CrossRef]
35. Neri A, Marrelli D, Megha T, et al. Clinical significance of multifocal and multicentric breast cancers and choice of surgical treatment: a retrospective study on a series of 1158 cases. *BMC Surg*. 2015;15(1):1. [CrossRef]
36. Tung NM, Boughey JC, Pierce LJ, et al. Management of hereditary breast cancer: American Society of Clinical Oncology, American Society for Radiation Oncology, and Society of Surgical Oncology Guideline. *J Clin Oncol*. 2020;38(18):2080-2106. [CrossRef]
37. Barroso-Sousa R, Metzger-Filho O. Differences between invasive lobular and invasive ductal carcinoma of the breast: results and therapeutic implications. *Ther Adv Med Oncol*. 2016;8(4):261-266. [CrossRef]
38. ACR BI-RADS Atlas. Breast Imaging Reporting and Data System. 5th ed. 2014. *American College of Radiology*. ISBN 9781559030168. [CrossRef]