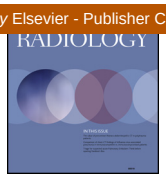
European Journal of Radiology 85 (2016) 199–204

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad



Double reading of current chest CT examinations: Clinical importance of changes to radiology reports

Peter M. Lauritzen[a,d,*], Knut Stavem[b,c,d], Jack Gunnar Andersen[e], Mali Victoria Stokke[f], Anne Lise Tennstrand[f], Gisle Bjerke[b], Petter Hurlen[a,d], Gunnar Sandbæk[e,d], Fredrik A. Dahl[c], Pål Gulbrandsen[d,c]

[a] Department of Diagnostic Imaging, Akershus University Hospital, Postboks 1000, NO-1478 Lørenskog, Norway
[b] Department of Pulmonary Medicine, Akershus University Hospital, Postboks 1000, NO-1478 Lørenskog, Norway
[c] Health Services Research Unit, Akershus University Hospital, Postboks 1000, NO-1478 Lørenskog, Norway
[d] Institute of Clinical Medicine, University of Oslo, Postboks 1171 Blindern, NO-0318 Oslo, Norway
[e] Department of Radiology and Nuclear Medicine, Oslo University Hospital, Postboks 4956 Nydalen, NO-0424 Oslo, Norway
[f] Department of Diagnostic Imaging, Vestre Viken Hospital, C/O Drammen Sykehus, NO-3004 Drammen, Norway

ARTICLE INFO

Article history:
Received 25 July 2015
Received in revised form 22 October 2015
Accepted 4 November 2015

Keywords:
Health care quality assurance
Health care peer review
Diagnostic errors
Computed tomography
Chest
Continuing professional development

ABSTRACT

Objectives: Misinterpretation of radiological examinations is an important contributing factor to diagnostic errors. Double reading reduces interpretation errors and increases sensitivity. Consultant radiologists in Norwegian hospitals submit 39% of computed tomography (CT) reports for quality assurance by double reading. Our objective was to estimate the proportion of radiology reports that were changed during double reading and to assess the potential clinical impact of these changes.

Materials and methods: In this retrospective cross-sectional study we acquired preliminary and final reports from 1023 consecutive double read chest CT examinations conducted at five public hospitals. The preliminary and final reports were compared for changes in content. Three experienced pulmonologists independently rated the clinical importance of these changes. The severity of the radiological findings in clinically important changes was classified as increased, unchanged, or decreased.

Results: Changes were classified as clinically important in 91 (9%) of 1023 reports. Of these: 3 were critical (demanding immediate action), 15 were major (implying a change in treatment) and 73 were intermediate (affecting subsequent investigations). More clinically important changes were made to urgent examinations and less to female first readers. Chest radiologist made more clinically important changes than other second readers. The severity of the radiological findings was increased in 73 (80%) of the clinically important changes.

Conclusion: A 9% rate of clinically important changes made during double reading may justify quality assurance of radiological interpretation. Using expert second readers and targeting a selection of urgent cases prospectively may increase the yield of discrepant cases and reduce harm to patients.

1. Introduction

Misinterpretation of radiological studies may have serious clinical consequences. An autopsy study of patients dying in hospital showed that radiological misinterpretation caused 8%, and contributed to another 33% of diagnostic errors in patients with relevant imaging [1]. Double reading is a practice in which two readers interpret an imaging examination, that reduces errors and increases sensitivity [2]. There are large variations in the reported effect of double reading in different settings, and the cost effectiveness is not well established [3–5]. Double reading can be conducted in several ways. Applied prospectively, it is used for quality assurance of current radiology reports, and it is routine in the training of residents when consultants check their preliminary reports [3,6,7]. Some screening programs apply independent double reading, in which readers are blinded to the interpretations of their colleague [8].

* Corresponding author.
E-mail addresses: peter.m.lauritzen@gmail.com (P.M. Lauritzen), knut.stavem@medisin.uio.no (K. Stavem), uxjaga@ous-hf.no (J.G. Andersen), mali.victoria.stokke@vestreviken.no (M.V. Stokke), al-tenn@online.no (A.L. Tennstrand), gisle.bjerke@ahus.no (G. Bjerke), petter@hurlen.no (P. Hurlen), gsandb@ous-hf.no (G. Sandbæk), fredrik.a.dahl@ahus.no (F.A. Dahl), pal.gulbrandsen@medisin.uio.no (P. Gulbrandsen).

http://dx.doi.org/10.1016/j.ejrad.2015.11.012

Recently, in the United Kingdom, the Royal college of Radiologists recommended that all radiology departments implement "peer feedback" with a systematic review of 5% of reports by December 2018, and that this effort be coupled with regular "Learning from Discrepancies meetings" [9,10]. Similarly, in the United States, continuous peer review of 5% of cases has been required for on-going credentialing by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) for some time [11]. In order to meet this standard with minimal impact on workflow, peer review programs such as RADPEER use retrospective double reading (review) of previous examinations when they are compared to current ones being interpreted [12]. It is the reviewing radiologist who selects the examinations, and the goals are quality improvement through shared learning from discrepancies and benchmarking of performance, rather than quality assurance of the individual report.

In the Scandinavian countries prospective double reading of current examinations is used for quality assurance, even to examinations read by consultants [6,13,14]. When reading an examination a consultant may choose to finalize the report directly or select it for double reading. The basis for this selection is the consultant's judgement of whether this quality assurance is warranted or not. Double reading may be requested explicitly or the examination may be routed to a queue for second reading when the interpreting radiologist chooses not to sign the report [6]. Consultant radiologists in Norwegian hospitals submit 39% of computed tomography (CT) examinations for double reading, and for all modalities combined this practice consumes 20–25% of their working hours [14]. The main goal is quality assurance of the report before it is finalized, and benchmarking of radiologist performance is uncommon [14].

The objective of this study was to estimate the proportion of radiology reports that were changed during double reading of current chest CT examinations, and to assess the potential clinical impact of these changes. We also aimed to explore whether characteristics of examinations or radiologists were associated with a higher proportion of clinically important changes.

2. Material and methods

2.1. Study design

In this retrospective cross-sectional multicentre study, preliminary and final radiology reports from 1023 consecutive double read chest CT examinations were collected and compared for changes (Fig. 1). Data was collected from the Radiology Information System and Electronic Patient Records at five public hospitals where double reading was carried out as routine quality assurance. All patients were from the department of internal medicine and were 18 years or older. The number of examinations collected from each hospital was in relative proportion to the number of consultant full time equivalents in the radiology department. All included examinations were conducted between 17 October 2011 and 29 March 2013, and had been double read by two consultant radiologists. The examinations had been selected for double reading by the interpreting radiologists. They submit examinations for this quality assurance according to their own judgement, as there are no established selection criteria. Accordingly, the reasons for submitting and the number of examinations submitted vary between radiologists. The clinical importance of changes made to the radiology reports following double reading was rated by experienced pulmonologists. Approval for the study and waiver of informed consent was obtained from the Regional Ethics Committee and the Data Protection Officer.

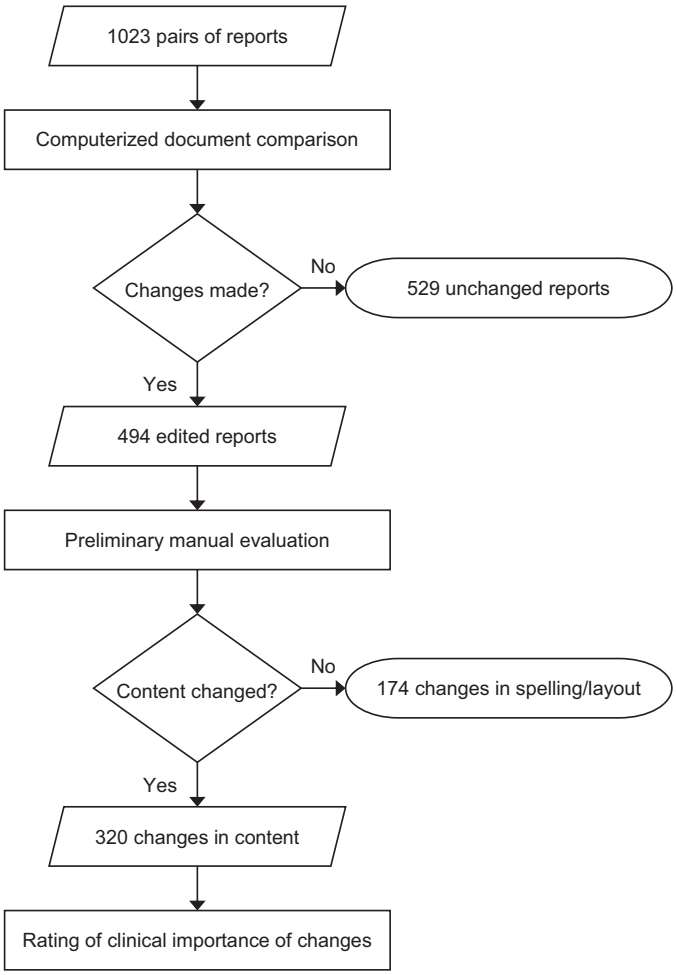


Fig. 1. Selection of radiology reports for clinical rating.

2.2. Patient and examination data

We collected data on patient gender and age, inpatient/outpatient status, urgency of the examination (routine or urgent, defined as requested within 24 h), the identities of the first and second reader, and the time of examination, time of preliminary and final reports (during working hours: 7 am–4 pm, or out of working hours).

2.3. Text comparison

The pairs of preliminary and final reports were compared using "Diff Doc Professional" (Softinterface Inc., Los Angeles, CA), a document comparison software, which labeled deletions, additions, and changes in the reports by color-coding. All radiology reports with changes in content beyond simple corrections of misspelling and layout were submitted for clinical rating (Fig.1).

2.4. Clinical rating

Two pulmonologists independently rated the clinical importance of changes in the reports on an ordinal five-point scale. We designed the clinical rating scale with the intention to be dichotomized in the statistical analysis as shown in Fig. 2. All changes with potential implications for patient management were considered clinically important.

In cases of discrepant ratings the clinical importance was classified according to the mean of the two ratings given. To resolve

Rating	Label	Explanation	Classification
1	Minimal change	Appearing not to affect treatment, investigation or prognosis. E.g. specifying known/unchanged conditions (such as known pleural thickening) or adequately located medical equipment (such as tubes or central venous catheters).	Not clinically important
2	Minor change	Appearing not to dictate any change in treatment, investigation or prognosis. E.g. moving information from text to conclusion or elaboration on information given in the preliminary report such as location of enlarged lymph nodes.	
3	Intermediate change	Appearing not to dictate a change in the treatment of the current condition, however necessitating a change in patient management such as added controls, change in further investigation or an altered prognosis. E.g. previously unknown lung noduli requiring follow up.	Clinically important
4	Major change	Likely to dictate a change in the treatment of the current condition or an altered primary diagnosis. E.g. altered assessment of treatment effect, missed tumor, effusion or pneumonia, misplaced essential medical equipment.	
5	Critical change	Implying that the patient is receiving erroneous or harmful treatment and that there is risk of death or permanent harm to health unless the treatment is corrected. E.g. missed central pulmonary embolus or aortic dissection.	

Fig. 2. Clinical importance rating scale.

non-integer mean ratings >2, an independent rating was obtained from a third pulmonologist, classifying clinical importance as the median of the three ratings. All three raters had >25 years of clinical experience. Their ratings were based on the preliminary and final radiology reports with color-coded changes, the referral, and the patients' age and gender.

The pulmonologists also made written comments about the assumed consequences of the changes they rated clinically important. Aided by these comments we classified all clinically important changes according to the clinical issues concerned. We also distinguished between increased, unchanged, or decreased severity of the radiological findings resulting from the changes. If the change contained added pathological findings or diagnostic suggestions leading to more comprehensive investigations or treatment, the severity was considered increased. The severity was considered decreased if initially reported pathological findings were removed or downgraded. Some changes could not easily be classified and were labeled: unchanged severity.

2.5. Radiologists

The involved consultant radiologists were classified based on length of experience as a consultant and subspecialty into five groups: inexperienced (<3 years), general radiologist (≥3 years, not working within a limited field of expertise), chest radiologist (≥3 years, working predominantly with chest imaging), oncology radiologist (≥3 years, working predominantly with oncology imaging), or other subspecialist (≥3 years, working within any other limited field of expertise).

2.6. Statistical analysis

The interrater agreement for the five-point scale was assessed using raw agreement and weighted kappa [15]. We used a weight of $1-[(i-j)/(k-1)]^2$, where "i" and "j" index the rows and columns of the ratings by the two raters and "k" is the maximum number of possible ratings. Differences in ratings between the two raters were tested with a related samples Wilcoxon signed rank test. Cohen's kappa was calculated for the dichotomized ratings.

Exploratory analysis of associations between clinical importance of changes and characteristics of patients, examinations, and readers was conducted with univariate logistic regression. A multivariate logistic regression model was constructed by entering variables whose univariate test had $p<0.25$ as candidate variables. Subsequently the candidate variable with the highest p value was removed. This process was repeated until only variables with $p<0.05$ remained.

We believed there could be associations between the readings of separate examinations conducted by the same radiologist. Clustering of clinically important changes in separate readings conducted by the same reader, were tested by constructing two random effects logistic regression models. The significant variables from the multivariate analysis were included as fixed effects coefficients in the models, and the random effects coefficients in the two models were the identity of the first and second readers, respectively.

Statistical analysis was conducted using Statistical Package for Social Sciences, version 22 (IBM corp, Somers, NY) and Stata, version 12.1 (StataCorp, College Station, TX). All p values are two sided. $p<0.05$ indicates statistical significance.

3. Results

3.1. Changes to reports

We included 1023 pairs of reports. Descriptive statistics regarding hospitals, patients, examinations, and radiologists are shown in Table 1. There were no changes in 529 reports (52%), and minor orthographical corrections or changes in lay out in 174 reports (17%). There were changes in content in 320 reports (31%), which were submitted for clinical rating (Fig. 1).

3.2. Clinical rating

The distribution of the ratings given by the two raters in the initial rating is shown in Table 2. Rater 1 gave lower ratings than rater 2 to 148 reports and higher ratings to 25 reports ($p>0.001$). On the five-point scale, the two raters were in agreement on 147 (46%) of 320 ratings, and achieved a weighted kappa of 0.42 (95% CI: 0.33–0.50). On the dichotomized scale, there was agreement on 232 (73%) of 320 ratings returning a kappa score of 0.34 (95% CI: 0.25–0.43).

The 173 discrepant ratings were resolved as follows: 38 reports were classified according to an integer mean rating; 70 reports with a mean rating of 1.5 were considered unequivocally "not clinically

Table 1
Descriptive statistics of examinations, patients, hospitals and radiologists. Number (%) unless stated otherwise.

Examinations (n = 1023)		
Urgent referral[a]	408	(40)
Conducted during ordinary working hours[b]	772	(75)
First reading during ordinary working hours[b,c]	700	(69)
Second reading during ordinary working hours[b,d]	852	(84)
Patients (n = 1023)		
Inpatients	700	(68)
Female gender	499	(49)
Age, mean (SD)	65.1	(15.4) years
Specialist experience of readers, mean (SD)		
First readers[e]	6.4	(8.5) years
Second readers[f]	12.0	(11.3) years
Hospitals (n = 5), median (range)		
No of beds, per hospital, dept. of internal medicine	104	(50–293)
Annual output, dept. of internal medicine[g]	9688	(5822–41,872)
No of annual CT exams, per hospital[h]	13,006	(5862–43,584)
Catchment population, per hospital	170,936	(77,836–471,661)
No of involved radiologists, per hospital	14	(5–34)
No of reports collected, per hospital	185	(43–405)
Proportion of double reading[i]	0.45	(0.15–0.59)
Subspecialty of radiologists (n = 88)		
Inexperienced consultant	22	(24)
General radiologist	26	(29)
Chest radiologist	4	(4)
Oncology radiologist	12	(13)
Other subspecialty	24	(26)
Role of radiologists (n = 91)		
First readings only	8	(9)
Second readings only	17	(19)
Both first and second readings	66	(73)
Gender of radiologists (n = 91)		
Female	34	(37)

[a] Urgent: requested within 24 h.
[b] Working hours: monday–friday 7 am–4 pm.
[c] $n = 1016$.
[d] $n = 1020$.
[e] $n = 1000$.
[f] $n = 1002$.
[g] Diagnosis Related Group (DRG)-weighted (no of admissions × DRG-index).
[h] Norwegian classification of Radiological Procedures (NCRP), 2012.
[i] Chest CT, referred from dept. of internal medicine, double read by consultants.

important" and not resolved further; 65 reports with non-integer mean ratings were submitted for a third rating. In the final classification, changes to 91 reports (8.9%, 95% CI: 7.2–10.8%) from 1023 double read examinations were clinically important. Changes to 73 reports (7.1%, 95% CI: 5.6–8.9%) were intermediate, 15 (1.5%, 95% CI: 0.8–2.4%) were major, and 3 (0.3%, 95% CI: 0.06–0.9%) were critical.

3.3. Clinical issues and clinical importance

The clinical issues concerned in changes classified as clinically important are presented in Table 3. Changes in the categories major and critical were predominantly concerning cardiovascular issues, cancer, or infection. Changes in the intermediate category were predominantly concerning cancer, pulmonary noduli or opacities, or enlarged lymph nodes or mediastinal mass. From the 91 clinically important changes, the severity was increased in 73 (80%), unchanged in 8 (9%) and decreased in 10 (11%).

3.4. Factors associated with clinical importance

Associations between clinical importance of changes and characteristics of patients, examinations and radiologists are shown in Table 4. The multivariate analysis showed that more clinically important changes were made to urgent examinations and less to female first readers. Subspecialty of the second reader was also associated with the rate of clinically important changes, with chest radiologists tending to make more clinically important changes than the other consultants.

Inpatient/outpatient status was associated with clinically important changes in the univariate model, but not in the multivariate model. The random effects logistic regression models did not show a significant clustering effect neither with regards to the identity of reader 1 ($p = 0.2$) nor reader 2 ($p = 0.2$)

4. Discussion

We have shown that double reading applied prospectively to radiologist-selected current chest CT examinations, may reveal clinically important interpretation discrepancies, with potential impact on patient management, in 9% of cases. Changes to 7% of reports were rated intermediate, affecting controls or subsequent investigations or altering prognosis. Although their results are unknown, the subsequent investigations are not inconsequential neither with regards to the patients nor to resource consumption. Changes to 2% were rated major or critical, implying changes in patient treatment. Although our data stem from a different approach both to double reading and rating of discrepancies our results are not significantly different from a previously reported 8.2% pooled total discrepancy rate and 2.8% major discrepancy rate for chest CT, suggesting that some quality assurance of radiological interpretation may be justified [16].

In our study, discrepancies were rated according to their potential clinical consequences, and the rating was performed by experienced pulmonologists. From a quality improvement perspective a clinical rating scale may be favorable since it is designed to identify which discrepancies are most likely cause harm [17]. Clinician raters should increase the validity of the ratings as they have superior clinical knowledge and are accustomed to making clinical decisions based partly on the content of radiology reports. In

Table 2
Rating of clinical importance; distribution of initial ratings ($n = 320$).

Individual rating			Rater 1					Total
			Not clinically important		Clinically important			(rater 2)
			Minimal	Minor	Intermediate	Major	Critical	
Rater 2	Not clinically important	Minimal	31	18	0	1	0	50
		Minor	52	89	5	1	0	147
	Clinically important	Intermediate	8	39	21	0	0	68
		Major	1	29	14	4	0	48
		Critical	0	4	0	1	2	7
Total (rater 1)			92	179	40	7	2	320

Table 3
Clinical issues and change in severity according to clinical importance of report changes. ($n = 91$), numbers of reports.

Classification of clinical importance		Intermediate ($n = 73$)			Major and Critical ($n = 18$)		
Change in severity		Increased	Unchanged	Decreased	Increased	Unchanged	Decreased
	n						
Cardiovascular	22	9	3	2	7		1
Cancer	21	13	1	1	6		
Infection	17	10	4	2	1		
Noduli[a] and opacities[b] requiring follow up	14	10		3	1		
Musculoskeletal/Trauma	5	4		1			
Airways/Bronchi/Emphysema	5	5					
Enlarged lymph nodes/Mediastinal mass	3	3					
Other[c]	4	2			2		
Sum	91	56	8	9	17	0	1

[a] According to the Fleischner Society Guidelines (recommended in radiology report and/or by one or both raters).
[b] Opacities not resolving or increasing on adequate therapy (follow up recommended in report and/or by one or both raters).
[c] Pneumothorax (1), pulmonary parenchymal necrosis (1), thickened oesophagus wall (1) and postoperative thoracic wall defect (1).

Table 4
Associations between clinically important report changes and characteristics of examinations, patients, and radiologists.

Variable		Logistic regression analysis					
		Univariate			Multivariate ($n = 1004$)		
	n	OR[a]	95% CI	*p*	OR[a]	95% CI	*p*
Examination							
Urgency[b]	1023	2.0	1.3–3.0	0.002	2.1	1.3–3.3	0.002
Examination time[c]	1023	1.3	0.8–2.1	0.2			
Time of first reading[c]	1016	1.1	0.7–1.8	0.6			
Time of second reading[c]	1020	0.8	0.5–1.6	0.6			
Patient							
Age (increase of 10 years)	1023	1.1	0.9–1.2	0.5			
Gender[d]	1023	1.5	1.0–2.4	0.06			
Admission status[e]	1023	2.1	1.2–3.7	0.007			
First reader, gender[d]	1023	0.7	0.5–1.1	0.2	0.6	0.4–1.0	0.04
First reader subspecialty	1014			0.6			
Inexperienced consultant[f]	368	1.0					
General radiologist	246	0.6	0.4–1.2				
Chest radiologist	45	0.6	0.2–2.1				
Oncology radiologist	136	0.8	0.4–1.5				
Other subspecialty	219	1.0	0.6–1.7				
Second reader, gender[d]	1023	1.6	1.0–2.4	0.04			
Second reader subspecialty	1004			0.003			0.002
Inexperienced consultant[f]	165	1.0					
General radiologist	293	0.5	0.3–0.9		0.6	0.3–1.1	
Chest radiologist	115	1.3	0.7–2.5		1.7	0.8–3.3	
Oncology radiologist	208	0.5	0.2–0.9		0.5	0.3–1.1	
Other subspecialty	223	0.4	0.2–0.9		0.5	0.2–1.0	

[a] OR: odds ratio.
[b] Urgency: non-urgent referral[f] vs. urgent (requested within 24 h).
[c] Time: during working hours (7 am–4 pm)[f] vs. not.
[d] Gender: male[f] vs. female.
[e] Admission status: outpatient[f] vs. inpatient.
[f] Reference in the logistic regression model.

peer review systems such as RADPEER, radiologists rate discrepancies mainly according to the magnitude of the error in question [12]. Such rating is subjective and may be perceived as punitive [17]. Previous studies on the scoring systems of the Royal College of Radiologists and RADPEER show slight to fair interrater agreement with a kappa of 0.17 and 0.2 respectively [18,19]. This is lower than the fair to moderate interrater agreement achieved with the clinical rating system in the present study [15]. Accordingly, our rating process was at least as reliable as others in use.

Our data stem from routine quality assurance as it is practiced, and the results should be representative of everyday clinical practice in these departments.

The first reader selected examinations for double reading. This selection may have been triggered by the complexity of a case or doubts of the first reader. Less experienced radiologists submitted more cases for double reading while more experienced radiologists tended to conduct the second reading, indicating that the task was not randomly assigned.

The selection of cases and readers may have resulted in a higher discrepancy rate than a random selection. This would have represented a bias, had our data been applied for benchmarking of performance. However, from a quality improvement perspective it is appropriate to focus the effort on a targeted selection where more discrepancies may be expected. Our results are a measure of how often this quality assurance method results in clinically important changes, rather than a measure of the performance of departments or individuals. Conversely, the non-random selection of cases in RADPEER-like systems may reduce discrepancy rates, as many radi-

ologists consider peer review "a waste of time" and "intentionally avoid cases requiring more time to peer review" [20]. Traditional peer review systems may also be vulnerable to underreporting of discrepancies in the same way that physicians are reluctant to participate in adverse event reporting, due to risk of liability exposure or professional embarrassment, burdensome reporting methods, perceptions of the clinical importance of adverse events, and lack of sense of ownership of the process [21].

Because we lack data on the total number of examinations interpreted and the proportion of examinations submitted for double reading by each radiologist, it is uncertain whether the lower rate of changes in reports made by female first readers is the result of a higher performance, or a tendency to submit more examinations for double reading. The higher rate of clinically important changes during second reading by subspecialist chest radiologist may partly be due to intentional routing of complex cases to these readers as well as their competence in detection, interpretation, and reporting. The finding that urgent examinations have higher odds of clinically important changes than non-urgent ones indicate that a targeted selection of examinations for double reading may increase the yield of discrepant cases, and thereby the benefit of the quality assurance measure. Hence, urgent cases should have priority for double reading. It might also be preferable that second readers are subspecialists when feasible. Establishing and refining objective selection criteria for double reading might further increase the likelihood of uncovering interpretation discrepancies, but would require additional studies.

In double reading of current examinations, as in the present study, discrepancies are uncovered with less delay, when patient treatment may still be corrected. This approach may reduce both harm to patients and the associated medico legal issues thereby addressing two reasons for radiologists' reluctance to participate wholeheartedly. For this reason we expect less underreporting in our data than in studies of peer review of previous examinations. It would also suggest that existing peer review systems might benefit from shifting attention from previous to current examinations, although certainly a more costly approach, and indeed the Royal College of Radiologists' recommendation is that review be carried out as soon as possible after the primary report so the risk of adverse outcome from a discordant finding is reduced [9].

Our study was limited to the written reports issued. Since these were acquired retrospectively, we could not consider any supplementary communication between radiologists and clinicians. Another limitation in our study is that some report changes may not result only from the second reading, because second readers may have gained additional information on patient development through clinical conferences, subsequent investigations, or the passing of time. Furthermore, the actual impact of the changes is unknown, as we neither had the resources nor sufficient access to patient medical records to establish this. It is questionable whether patient records can be relied on to establish this retrospectively as a course of action may change before it is recorded.

5. Conclusions

We conclude that a 9% rate of clinically important changes made during double reading suggests that some quality assurance of radiological interpretation is justified. A prospective approach may reduce the risk of harm to patients. Using expert second readers and targeting urgent cases may increase the yield of discrepant cases. Establishing objective selection criteria would require further studies.

Role of funding source

The funding sources had no involvement in study design, collection, analysis or interpretation of data, the writing of the report or the decision to submit for publication.

Acknowledgements

We thank Professor Amund Gulsvik for contributing to the classification of clinical issues and for his input on an early draft of the paper. We also thank Professor Vidar Søyseth for his contribution as the third clinical rater. The study received funding from the Norwegian Medical Association's fund for quality improvement and patient safety and from the Norwegian Society of Radiology's Research fund.

References

[1] G.S. Heriot, P. McKelvie, A.G. Pitman, Diagnostic errors in patients dying in hospital: radiology's contribution, J. Med. Imaging Radiat. Oncol. 53 (2) (2009) 188–193.
[2] L.G. Quekel, R. Goei, A.G. Kessels, E. van, J.M. ngelshoven, Detection of lung cancer on the chest radiograph: impact of previous films, clinical information, double reading, and dual reading, J. Clin. Epidemiol. 54 (11) (2001) 1146–1150.
[3] M.P. Federle, D. Gur, Double reading of certain examinations such as barium enemas and mammograms can increase sensitivity at the expense of specificity, AJR Am. J. Roentgenol. 164 (5) (1995) 1291–1292.
[4] C. Agostini, M. Durieux, L. Milot, I. Kamaoui, B. Floccard, B. Allaouchiche, et al., Value of double reading of whole body CT in polytrauma patients, J. Radiol. 89 (3 Pt 1) (2008) 325–330.
[5] L.S. Babiarz, D.M. Yousem, Quality control in neuroradiology: discrepancies in image interpretation among academic neuroradiologists, AJNR Am. J. Neuroradiol. 33 (1) (2012) 37–42.
[6] P. Hurlen, T. Østbye, A. Borthne, P. Gulbrandsen, Introducing PACS to the late majority. a longitudinal study, J. Digit. Imaging 23 (1) (2010) 87–94.
[7] G.C. Jakanani, R. Botchu, S. Gupta, J. Entwisle, A. Bajaj, Out of hours multidetector computed tomography pulmonary angiography: are specialist resident reports reliable? Acad. Radiol. 19 (2) (2012) 191–195.
[8] R. Ballard-Barbash, C. Klabunde, E. Paci, M. Broeders, E.A. Coleman, J. Fracheboud, et al., Breast cancer screening in 21 countries: delivery of services, notification of results and outcomes ascertainment, Eur. J. Cancer Prev. 8 (5) (1999) 417–426.
[9] The Royal College of Radiologists. Quality assurance in radiology reporting: peer feedback. London, UK, 2014.1–18.
[10] The Royal College of Radiologists. Standards for Learning from Discrepancies meetings. London, UK, 2014.1–16.
[11] S. Mahgerefteh, J.B. Kruskal, C.S. Yam, A. Blachar, J. Sosna, Peer review in diagnostic radiology: current state and a vision for the future, Radiographics 29 (5) (2009) 1221–1231.
[12] V.P. Jackson, T. Cushing, H.H. Abujudeh, J.P. Borgstede, K.W. Chin, C.K. Grimes, et al., RADPEER scoring white paper, J. Am. Coll. Radiol. 6 (1) (2009) 21–25.
[13] S. Diaz, O. Ekberg, The frequency of diagnostic errors in radiologic reports depends on the patient's age, Acta Radiol. 51 (8) (2010) 934–938.
[14] P.M. Lauritzen, P. Hurlen, G. Sandbaek, P. Gulbrandsen, Double reading rates and quality assurance practices in Norwegian hospital radiology departments: two parallel national surveys, Acta Radiol. 56 (1) (2015) 78–86.
[15] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1) (1977) 159–174.
[16] M.Z. Wu, M.D. McInnes, D.B. Macdonald, A.Z. Kielar, S. Duigenan, CT in adults: systematic review and meta-analysis of interpretation discrepancy rates, Radiology 270 (3) (2014) 717–735.
[17] D.B. Larson, J.J. Nance, Rethinking peer review: what aviation can teach radiology about performance improvement, Radiology 259 (3) (2011) 626–632.
[18] B. Mucci, H. Murray, A. Downie, K. Osborne, Interrater variation in scoring radiological discrepancies, Br. J. Radiol. 86 (1028) (2013) 20130245.
[19] L.C. Bender, K.F. Linnau, E.N. Meier, Y. Anzai, M.L. Gunn, Interrater agreement in the evaluation of discrepant imaging findings with the Radpeer system, AJR Am. J. Roentgenol. 199 (6) (2012) 1320–1327.
[20] R.L. Eisenberg, M.L. Cunningham, B. Siewert, J.B. Kruskal, Survey of faculty perceptions regarding a peer review system, J. Am. Coll. Radiol. 11 (4) (2014) 397–401.
[21] D.O. Farley, A. Haviland, S. Champagne, A.K. Jain, J.B. Battles, W.B. Munier, et al., Adverse-event-reporting practices by US hospitals: results of a national survey, Qual. Saf. Health Care 17 (6) (2008) 416–423.