

# A systematic review of double reading in diagnostic radiology

By Dr. H Geijer & Dr. M Geijer

*The optimal outcome of diagnostic imaging would of course ideally be one hundred percent accuracy in all diagnoses, but unfortunately, in practice this is not possible. Instead, the pragmatic approach that is adopted is to keep the number of discrepancies as low as possible. One way to achieve this goal is via the system of double reading, namely the arrangement in which two observers review the same study before the report is finalized. However, it is evident that this approach has the side-effect of requiring considerable resources. In a report from Norway, double reading of one third of all radiologic examinations consumed 20-25 % of all available radiology working hours [1].*

*There are conflicting opinions on the role of double reading, which has prompted further investigation into its real value. This article summarizes the results of a recent systematic review of double reading in diagnostic radiology [2].*

## The Authors

Dr. Håkan Geijer, Department of Radiology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

Dr. Mats Geijer, Department of Radiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Corresponding author:

Dr. Håkan Geijer, Professor in Radiology.

Department of Radiology, Örebro University Hospital, SE-701 85 Örebro, Sweden

Email: hakan.geijer@regionorebrolan.se

Double reading can be performed in a number of scenarios:

A) Double reading of trainee or residents' cases by a more experienced colleague is common. The value — and necessity — of this is hardly debatable.

B) Perhaps the most common type of double reading occurs in screening, where two readers create separate reports that are then compared. However, screening is a particular situation which is different from that of clinical radiology since sensitivity must be very high, even at the cost of lower specificity. In screening, the logical "OR" rule is often applied, i.e. if one of the readers indicates a lesion, the case is regarded as positive.

C) Another type of double reading occurs when one of the readers is a computer, in so called computer aided diagnosis, CAD. This is often employed in screening settings when conventional double reading between two readers is not used.

D) In quality assurance and research settings, two readers might produce independent reports that are compared later.

E) Sometimes, a third reader might be employed to solve disagreements between the first and second reader. This is called arbitration when the third reader is aware of the specific disagreements and pseudo-arbitration when not.

F) Another special type of double reading is when a certain percentage of reports are double read retrospectively. This is of no benefit for the individual patient, but is performed entirely for quality assurance purposes. The American College of Radiologists (ACR) RADPEER program [3] is an example of this.

G) In clinical practice, which is the subject of our review, the first reader produces a report which is checked by a second reader before the report is finalized and sent to the referring physician [2]. Thus, both readers have direct influence on the treatment of the individual patient.

## REVIEW OF LITERATURE — METHODS

A systematic review of the literature was performed, searching for studies reporting error rates in double reading compared with single reading. The focus was on double reading between radiology specialists, so double reading of residents or double reading for screening purposes (i.e scenarios A and B above) were not included.

## RESULTS

We found 1610 eligible articles which were reduced to 46 articles after abstract and full-text reading. There was a high heterogeneity in the results which made a meta-analysis impossible.

Fifteen articles evaluated double reading in CT. In trauma CT, although the rate of discordances was high, namely 26-37 %, patient care was in fact changed in only 2.3 %, as reported in one study [4]. In abdominal CT, 17 % initial discordances resulted in treatment change in 3 % of cases after double reading [5]. Chest CT for pulmonary nodules can employ CAD as one of the readers. The results were in fact better than double reading between two human readers [6,7].

*“...A reasonable compromise might be to institute double reading only for selected, high-risk types of examinations ...”*

In general radiology, a large study involving over 11,000 cases found a mean rate of disagreements of 4.4 % [8] while another study of similar size found 2.7 % discordances, with missed findings being the most common discordance [9].

A large number of the selected articles dealt with double reading by a sub-specialty radiologist of initial reports produced by a general radiologist. The discrepancy rate was higher in this setting than in double reading between radiologists of the same degree of specialization. One article reported that a higher degree of sub-specialization resulted in higher discrepancy rates [10]. In chest radiology, a thoracic radiologist reported fewer indeterminate nodules than the general radiologist [11]. In CT and MRI of head and neck cancer patients, the subspecialized neuro-radiologist changed the interpretation in 41 % cases, mostly up-staging the tumor [12], while another study reported a clinically significant discrepancy rate of 12 %, most of these being perceptual errors [13].

## DISCUSSION

Our review found a wide variety of discrepancy rates in various studies, ranging between 0.4 % and 22 %, with minor discrepancies being more common. There was a lack of clear evidence for or against systematic double reading. In

general, double reading increased sensitivity at the cost of reduced specificity due to both more true positive and more false positive findings being reported in double reading. The advantage of double reading must, however, be balanced by the resources it consumes. A reasonable compromise might be to institute double reading only for selected, high-risk types of examinations such as trauma CT and possibly CT of the thorax and abdomen.

There were only two studies [4,5] which involved a reading arrangement which corresponded exactly to that of our main question, namely the value of double reading in a clinical context by a non-blinded second reader before the report is finalized. This resulted in a change in patient care in 2.3 - 4 % of cases. Double reading in CT has been evaluated in a recent systematic review [14] which found a major discrepancy rate of 2.4 %, and even lower when the second reader was non-blinded.

Thus, it seems that double reading discovers comparatively few discrepancies, but nevertheless consumes a considerable amount of available radiologists' resources. Double reading probably has the highest benefit in complicated studies that are read under pressure, such as polytrauma CT. More surprisingly, it seems that double reading by a sub-specialized radiologist results in more discrepancies. To an extent

*“...Radiology sub-specialization might be a more economical way to increase quality than double reading on a grand scale...”*

this might be due to the sub-specialist being more familiar with the demand from the clinicians in the particular field of expertise. This naturally begs the question: why are not all radiologists sub-specialized? With digital radiology, images can easily be transmitted over long distances, so there are really only organizational obstacles to this. Radiology sub-specialization might be a more economical way to increase quality than double reading on a grand scale.

## REFERENCES

1. Lauritzen PM, Hurlen P, Sandbaek G, Gulbrandsen P. Double reading rates and quality assurance practices in Norwegian hospital radiology departments: two parallel national surveys. *Acta Radiol* 2015; 56: 78.
2. Geijer H, Geijer M. Added value of double reading in diagnostic radiology, a systematic review. *Insights Imaging* 2018; 9: 287.
3. Jackson VP, Cushing T, Abujudeh HH, Borgstede JP, Chin KW, Grimes CK, et al. RADPEER scoring white paper. *J Am Coll Radiol* 2009; 6: 21.
4. Yoon LS, Haims AH, Brink JA, Rabinovici R, Forman HP. Evaluation of an emergency radiology quality assurance program at a level I trauma center: abdominal and pelvic CT studies. *Radiology* 2002; 224: 42.
5. Gollub MJ, Panicek DM, Bach AM, Penalver A, Castellino RA. Clinical importance of reinterpretation of body CT scans obtained elsewhere in patients referred for care at a tertiary cancer center. *Radiology* 1999; 210: 109.
6. Rubin GD, Lyo JK, Paik DS, Sherbondy AJ, Chow LC, Leung AN, et al. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. *Radiology* 2005; 234: 274.
7. Wormanns D, Beyer F, Diederich S, Ludwig K, Heindel W. Diagnostic performance of a commercially available computer-aided diagnosis system for automatic detection of pulmonary nodules: comparison with single and double reading. *RofO* 2004; 176: 953.
8. Siegle RL, Baram EM, Reuter SR, Clarke EA, Lancaster JL, McMahan CA. Rates of disagreement in imaging interpretation in a group of community hospitals. *Acad Radiol* 1998; 5: 148.
9. Harvey HB, Alkasab TK, Prabhakar AM, Halpern EF, Rosenthal DI, Pandharipande PV, et al. Radiologist Peer Review by Group Consensus. *J Am Coll Radiol* 2016; 13: 656.
10. Bell ME, Patel MD. The degree of abdominal imaging (AI) subspecialization of the reviewing radiologist significantly impacts the number of clinically relevant and incidental discrepancies identified during peer review of emergency after-hours body CT studies. *Abdom Imaging* 2014; 39: 1114.
11. Nordholm-Carstensen A, Jorgensen LN, Wille-Jorgensen PA, Hansen H, Harling H. Indeterminate pulmonary nodules in colorectal-cancer: do radiologists agree? *Ann Surg Oncol* 2015; 22: 543.
12. Loevner LA, Sonners AI, Schulman BJ, Slawek K, Weber RS, Rosenthal DI, et al. Reinterpretation of cross-sectional images in patients with head and neck cancer in the setting of a multidisciplinary cancer center. *AJNR Am J Neuroradiol* 2002; 23: 1622.
13. Kabadi SJ, Krishnaraj A. Strategies for Improving the Value of the Radiology Report: A Retrospective Analysis of Errors in Formally Over-read Studies. *J Am Coll Radiol* 2017; 14 : 459.
14. Wu MZ, McInnes MD, Macdonald DB, Kielar AZ, Duiganan S. CT in adults: systematic review and meta-analysis of interpretation discrepancy rates. *Radiology* 2014; 270: 717