Improving the Quality and Safety of Radiology Reporting: A Prospective Audit of Discrepancies amongst Consultant Radiologists

Abstract: (250 words max)

Objectives: To examine discrepancy rates over multiple annual cycles in reporting by consultant radiologists for the interpretation of acute CT examinations.

Methods: A prospective audit with peer review second reading of acute CT scans was implemented in one radiology department in the UK. A longitudinal audit over five years was performed to determine the discrepancy rate of acute and emergency CT scans conducted at weekends. The examinations comprised CT Head, Thorax, Abdomen and Pelvis, CT KUB, and CT Angiography scans. Discrepancies were scored as no discrepancy (zero), minor discrepancy (one), moderate discrepancy (two) or significant discrepancy (three). Seventeen consultants took part in the audit as primary reporters and second readers. All consultants were on-call radiologists, and the allocation of cases was randomised depending on the on-call rota. Results were reported annually to one consultant radiologist.

Results: A significant discrepancy rate of 1.2% (p=0.026) was found for consultant radiologists interpreting acute CT examinations based on 2951 second read CT scans.

Conclusions: To the best of our knowledge, this is the first study to establish significant discrepancy rates amongst consultant radiologists interpreting acute CT examinations over time.

Keywords: (3-5 keywords from MeSH database)

Discrepancy rates

Radiologists

CT scan

Introduction:

Discrepancies in the interpretation of radiological examinations are common and represent an inevitability of image interpretation in everyday radiology practice. Discrepancies can be due to misperceptions (for example, not detecting an abnormality) or misinterpretations (for example, providing an incomplete or incorrect differential diagnosis. Overall, discrepancy rates are estimated at between 3-5%, but retrospective double reading discrepancy rates can be as high as 30% in some instances [1]. Radiology is fundamentally an opinion-based specialty, contrary to the beliefs of many doctors and patients. Many factors may influence the degree of trust and confidence in which clinicians hold in the opinion, including whether the radiologist is known to them personally [2].

With regard to discrepancies of interpretation, two radiologists may hold different opinions about the interpretation of an image, but that does not make one radiologist correct and the other incorrect but instead raises the issue of the differences between perceived and actual reality. The fundamental accuracy of image interpretation is subject to many factors including the clinical information available at the time of the reporting (which may be limited or inaccurate), the experience of the radiologist, the area of practice or type of examination which a radiologist is interpreting, and also the amount and complexity of abnormality present on the image which the radiologist is attempting to interpret. However the impact of this interpretation is of direct relevance to subsequent diagnostic and management processes by medical and surgical practitioners. To date, there has been no prospective published ongoing evaluation of the performance and discrepancy rates of a cohort of experienced radiologists over a designated period.

With the reporting of our prospective and ongoing audit of discrepancy rates over a

3

five-year cycle from 2013 to 2017, we aim to assess for the average performance of consultant radiologists in a vital area, that of the interpretation of acute CT scans.

Materials and Methods:

As part of a quality improvement initiative in 2012, a radiology project involving prospective peer review second reading of a sample of acute CT scans was implemented in the Department of Radiology in South Tyneside and Sunderland NHS Foundation Trust to establish a value for the normal discrepancy rate within a cohort of consultant radiologists. The results of the double reader audit of a sample of acute CT reports over five consecutive years from 2013 to 2017 were analysed to obtain a tangible measurement of the normal error or discrepancy result within a consultant cohort in NHS practice, and as a consequence of this, to demonstrate any temporal relationship in discrepancy rates within this five-year period interval. There is no data on discrepancy rates prior to the commencement of this audit as this data is not routinely collected in UK hospitals.

<u>Examinations Audited</u>

This is a prospective, non-blinded, non-randomised audit of discrepancy rates in peer-reviewed second reading of acute and emergency CT scans performed and reported at weekends from a single radiology department in a large (950 bed) NHS Trust in the Northeast of England. The examinations comprised of CT Head, Thorax, Abdomen and Pelvis, CT KUB, and CT Angiography scans. The audit was performed in five consecutive annual cycles over five consecutive years between 2012 to 2017. This study was non-blinded; thus, the second reader was able to read

the original report, and the first reader was aware that their reports would be second read.

Reporters and Peer Reviewers

The primary reporters and second readers are a cohort of seventeen consultant radiologists who perform acute on-call duties in Sunderland Royal Hospital. From a pragmatic perspective, the audit was undertaken by the on-call radiologists iteratively providing details of specific cases to the next reporting radiologist designated to be on call over the following weekend. Therefore, although the radiologists who were reporting the examination were aware that they were the subject of a second read audit, the allocation of the cases to the second readers was dependent upon the on-call rota. None of the cases were reported by radiologists in training. The results were reported annually to a consultant radiologist colleague who collated the overall dataset as a means of standardising the data collection process of the audit.

The total number of CT scans performed in the department was also recorded for each year. All consultants who undertook on-call duties took part in the audit. All consultant radiologists taking part in the audit were primary reporters and acted as second readers.

Grading of Discrepancies

Discrepancies subjectively were numerically scored as no discrepancy (zero), minor discrepancy (one), moderate discrepancy (two) or significant discrepancy (three). This was in keeping with the RCR guidance on discrepancy analysis at the time of commencement of the audit in 2012; this guidance has subsequently altered with the

grading of discrepancies no longer recommended, and currently educational "learning from discrepancy" approaches are now being advocated. Both discrepancies of perception and interpretation were included in this audit.

Results:

The grades and number of discrepancies as scored are provided in Table 1. Overall, for the five-year cycle, the total discrepancy rate (moderate and significant discrepancies) from the audit was 1.2%. This is based on 2951 second read CT examinations in total. This represents 2.0% of a total of CT scans performed in this period.

With each year, there appears to be an improvement in the significant discrepancy rate from 2012 to 2017. The only outlying year is 2017 when the discrepancy rate rose by 0.1%, the reasons for this are unclear. The decrease in discrepancy rates between each year is statistically significant (X-squared = 11.09, df = 4, p-value = 0.02557.)

Discussion:

Despite debates, numerous publications, and editorials on this topic over decades, the normal discrepancy rate for a fully trained radiologist who has reached consultant level practice has never been accurately established. Indeed, as radiology is an opinion-based specialty, subject to human factors, a certain amount of discrepancy

between image interpreters is inevitable. Also, as has been saliently pointed out by other authors [1], clinical radiology remains the only specialty in which every image and subsequent diagnostic imaging report can be scrutinised in hindsight with relative ease, leading to the problem of extensive implicit hindsight bias. This is a prospective audit to assess discrepancy rates over a period of 5 years (5 audit cycles) in a cohort of general radiologists with different skill sets and experiences, all at consultant level. Our study demonstrated that the significant discrepancy rate for acute CT reporting is low at 1.2% overall and year on year, there was a statistically significant reduction in the discrepancy rates. Also, repeated audits over the years may well be a factor in improving discrepancies, although we have not addressed whether this translates into improvements in overall report quality and fewer clinically relevant discrepancies. We have also not identified why our discrepancy rates seem to have gradually but significantly decreased over the 5-year cycle. It is interesting to speculate that rather than becoming "more correct" we as a cohort of consultants are just becoming "more concordant" as a result of the audit process which raises interesting questions concerning the psychology underpinning radiological discrepancy and also the reported social dynamics within specific groups of radiologists. More optimistically, the theory could be advanced that experiential learning in the workplace, and repeated audit have improved performance.

Level of experience amongst radiologists

Several studies have reported discrepancy rates in trainee radiologists or residents in training. Ruma et al. [3] identified discrepancies in ultrasound, CT, and MR examinations in second to fourth-year radiology residents over nine months. The discrepancies were graded using a modified RADPEER grading systemin 21,482

examinations. The authors found 158 clinically significant discrepancies (which would equate to the most serious discrepancies), a discrepancy rate of 0.7%. The authors found no difference between the discrepancy rate across subspecialties or between adult or paediatric examinations. Within neuroradiological examination interpretation, there was a significantly higher discrepancy rate for MRI (1.5%) compared with CT (0.6%); within abdominal radiology, there was a significant higher discrepancy rate for CT (1.1%) compared with ultrasound (0.2%).

Some retrospective audits of trainee radiologists' reporting have demonstrated higher discrepancy rates. Terrablanche [4] reports rates of up to 17.1% (with a major discrepancy rate of 7.7%) demonstrated on the review of 1477 CT scans reported out of hours by registrars of varying experience with significant differences found between discrepancy rates on trauma scans as opposed to non-trauma scans.

Buskov et al. [5] compared the performance of trainee radiologists against reporting radiographers in a series of 1000 consecutive appendicular radiographs of the appendicular skeleton. They found sensitivities of 99% for plain film radiography and 94% for trainee radiologists and specificities of 97% for the radiographers and 99% for the trainee radiologists. Radiographers had a lower missed rate of fractures but did seem report to more findings as abnormal, therefore resulting in the lower specificity noted above. Of note, there is no information provided about the underlying level of an abnormal pathology within the test set of one thousand images [3] which may account for the extremely low discrepancy rates compared to the series assessing acute CT reporting [2].

Discrepancy Rates in Acute CT reporting

One series of the second reading of a small (60 examinations) random sample of CT examinations of the abdomen and pelvis (both emergency and elective cases) by three experienced specialist radiologists used the same grading scale as was used in this audit and found major discrepancy rates of between 26-32% [4]. Laurieten et al. [6] analysed 1071 consecutive abdominal, and pelvic CT scans from five Norwegian hospitals over 17 months. The request for the second read was initiated by the primary radiologist who may have been an inexperienced or experienced radiologist. Preliminary reports and final reports were compared for changes; the clinical importance of these changes was rated on a 5-point scale by at least two gastrointestinal surgeons with at least ten years of experience. Additionally, the severity of any changes was classified as increased, decreased, or unchanged. Of the reports, 14% were identified as those containing clinically significant changes, three critical, 35 major, and 108 intermediates. The severity of the changes increased in 118 (81% of reports) with critically significant changes. This retrospective but topical study demonstrates both the challenges and the opportunities available for improved quality assurance in reporting of acute radiology scans. The authors describe an ad hoc prospective double reading process initiated by the primary reader due to uncertainty about the presence or clinical relevance of the findings. The double reading process consumes a substantial amount (20-25%) of their professional resources. The secondary reader finalises the report with a mean delay of 20 hours from the preliminary interpretation [6]. In the UK, provision of a double reporting system for emergency CT cases is uncommon outside of training institutions with trainee reports being checked by a consultant. It is difficult to envisage how on the basis of the delay in report time alone, this double reporting

practice could be transposed into an NHS context, with ever increasing demands for more rapid imaging reports to guide clinical management and meet the ubiquitous four-hour waiting targets in the Emergency Departments, metrics which the NHS both tangibly measures and subsequently penalises. Equally, it is difficult to measure this as an accurate reflection of the discrepancy rate amongst radiologists in consultant practice as these cases are the ones the primary radiologist has chosen to send for a second opinion due to diagnostic uncertainty and which are the more complex or ambiguous cases. There is no measure of the radiologist's reporting accuracy across the entirety of their case mix for acute CT.

Abujadeh [8] reported the discrepancy rates amongst a series of CT abdomen and pelvis examinations amongst experienced radiologists. This study involved three expert radiologists double reading a random subset of sixty examinations, thirty of their reports and 30 of other radiologists' reports. CT scans were performed for a variety of common indications including evaluation of malignancy, pancreatitis, and urinary tract stones. Intra and inter-observer major discrepancy rates were between 26% and 32% respectively, which illustrates the high level of discordance (but not necessarily error) between reports amongst experienced radiologists.

The Royal College of Radiologists published the results of a national audit of reports of 4931 emergency abdominal CT scans for patients presenting with non-traumatic abdominal pain which included 2568 surgical and 2363 non-surgical cases which dated from 2013 [9]. The pooled major discrepancy rate was 2.3% for on-site consultant radiologists, 3.4% for registrar or trainee reports, and 6% for off-site (outsourced) reports. While the numbers of cases in this national audit are indeed

substantial and the efforts laudable, there are several methodological differences between this audit and the one we present. The cases in the RCR audit were selected from departments throughout the UK, although specifically how the cases from 2013 were selected at local level is not clear. Also, the case mix with almost equal numbers of surgical and non-surgical cases likely to be enriched with abnormalities compared to a more random sample. This factor, the level of abnormality in underlying sample, will also influence the discrepancy rates. Nevertheless, the results are broadly consistent with the findings of longitudinal audit.

A meta-analysis published in 2014 [10] included fifty-eight studies addressing the question of discrepancy rates in the reporting of CT examinations. The pooled discrepancy rate was 7.7%, and the major discrepancy rate was 2.4%. The pooled discrepancy rate was comparable for staff and residents (junior and senior radiologists.) The pooled discrepancy rates for head CT and spine CT were lower than those for chest CT and abdominal CT. Lack of blinding of the reference radiologist to the initial report was associated with a lower major discrepancy rate than when blinding was present. The authors concluded that there is considerable heterogeneity in discrepancy rates. Further work was deemed to be necessary particularly about the question of the blinding of the reference radiologists. Evidence from the audit is extremely valuable for annual appraisal and enables our department of radiologists to fulfil the Royal College of Radiologists peer review criteria for professional revalidation and the requirement for participation in audit or quality improvement criteria for this annual appraisal cycle [11].

<u>Limitations to the Audit</u>

1. Single Site Audit

This a single centre audit from a large teaching district general hospital within the Northeast of England.

2. Non-blinding to the original report

 As this was a pragmatic audit, the second reader was not blinded to the report from the first reader. The lack of blinding may be a source of bias as has been identified in a meta-analysis [10]. The practice of performing the audit in multiple cycles may have resulted in a bias of convergence in reporting styles and reporting decisions.

3. Lack of Clinical Assessment of Discrepancy

 The assessment of the severity of the discrepancy was made by the second reading radiologist. No clinical assessment of discrepancy severity in terms of consequent diagnosis and clinical management was made in this study, in contrast to the study by Lauritzen [6] in which the assessment of the clinical impact of such discrepancies was reported by a gastrointestinal surgeon, dependent on the outcomes of diagnostic imaging such as those we report here.

4. Discrepancies of the Second Reader

The second readers of the reports were all part of the same cohort as the first interpreters of the images which makes this audit unique. These radiologists were also as likely as the first readers to be prone to discrepancies. There may be discrepancies within the second reader's interpretations of the imaging.

Conclusion:

A significant discrepancy rate of 1.2% (p=0.026) was found for consultant radiologists interpreting acute CT examinations based on 2951 second read CT. To

the best of our knowledge, this is the first study to establish significant discrepancy rates amongst consultant radiologists interpreting acute CT examinations over time.

Author Contributions:
Data Collection: J Connor
Data Analysis: J Connor, K McGarry
Data Interpretation: J Connor, J Cox, Y Graham, C Hayes, P Mehrotra, K McGarry
Drafting/ Editing of Manuscript: J Connor, J Cox, Y Graham, C Hayes, P Mehrotra, K McGarry

Word count: 2893

References

1. Brady AP(2017) Error and discrepancy in radiology: inevitable or avoidable? Insights into Imaging; (8) 171–82.

2. Graham Y, Hayes C, Mehrotra P, et al.(2019) Clinicians' perceptions of the quality of outsourced radiology and actions taken around perceived imaging errors in practice
    European Radiology; 29(4) 1649-1654

3. Ruma J, Klein KA, Chong S, Wesolowski J, Kazerooni EA, Ellis JH, et al.(2011) Cross-sectional examination interpretation discrepancies between On-Call diagnostic radiology residents and subspecialty faculty radiologists: Analysis by imaging modality and subspecialty. J Am Coll Radiol.; 8(6) 409-414

4. Terreblanche OD, Andronikou S, Hlabangana LT, Brown T, Boshoff PE (2012) Should registrars be reporting after-hours CT scans? A calculation of error rate and the influencing factors in South Africa. Acta Radiol. 2012; 53(1) 61-68

5. Buskov L, Abild A, Christensen A, Holm O, Hansen C, Christensen H.(2013) Radiographers and trainee radiologists reporting accident radiographs: A comparative plain film-reading performance study. Clin Radiol.; 68(1) 55-58

6. Lauritzen PM, Andersen JG, Stokke MV, Tennstrand AL, Aamodt R, Heggelund T, et al.(2016) Radiologist-initiated double reading of abdominal CT: Retrospective analysis of the clinical importance of changes to radiology reports. BMJ Qual Safety;25(8):595–603

7. Sahni VA, Khorasani R.(2016) Radiology double reads. BMJ Quality and Safety. 25 (8):569-71

8. Abujudeh HH, Boland GW, Kaewlai R, Rabiner P, Halpern EF, Gazelle GS, et al. Abdominal and pelvic computed tomography (CT) interpretation: Discrepancy rates among experienced radiologists. Eur Radiol.

2010;20(8):1952–7.

9   Howlett D, Drinkwater K, Frost C et al. (2017) The accuracy of interpretation of emergency abdominal CT in adult patients who present with non-traumatic abdominal pain: results of a UK national audit Clinical Radiology (2017) 72(1) 41-51

10.   Wu MZ, McInnes MDF, Blair Macdonald D, Kielar AZ, Duigenan S. CT in Adults: Systematic Review and Meta-Analysis of Interpretation Discrepancy Rates. (2014) Radiology; 270(3):717–35

11.   The Royal College of Radiologists. Quality Assurance in radiology reporting: peer feedback. Clin Radiol. 2014;