# Observer performance in CT head reporting:

# An audit of 3,008 cases

P. Lockwood

*Allied Health Department, Canterbury Christ Church University, Medway Campus, Rowan Williams Building, 30 Pembroke Court, Chatham Maritime, Kent, ME4 4UF, UK*

**Keywords:** ***Reporting Radiographer, Computed Tomography, Image Interpretation, Diagnostic Accuracy.***

Word count: 2439

**Abstract**

*Aim:* To audit the reporting results of a cohort of radiographers (*n*=6) completing an accredited academic programme in clinical reporting of Computed Tomography (CT) head examinations.

*Methods:* An audit of retrospective academic image case banks and prospective random clinical workload case banks. Both the academic test banks and clinical workload banks included a wide range of normal and abnormal cases of different levels of difficulty and pathology. Abnormalities included: haemorrhage, fractures, lesions, infarctions, degeneration, and normal variants from a variety of referral sources. True positive and negative, as well as false positive and negative fractions were used to mark the reports, which were analysed for accuracy against a reference standard. Further interobserver variability was assessed using Cohens Kappa, one-way analysis of variance and Tukey for multiple comparisons and significance testing at 95% confidence intervals.

*Results:* The mean accuracy score for all radiographers (*n*=6) and reports (*n*=3,008) was 90.7% (95%CI 88.3%-93.0%). Mean sensitivity and specificity rate was 86.9% (95%CI 85.8%-88.2%), and 94% (95%CI 89.6%-98.3%) respectively. The most common errors were associated with herniation, lacunar infarctions and subtle fractures (false negatives) and involutional changes, subtle infarctions, and ventricular dilation (false positives).

*Conclusions:* The results suggest appropriately trained radiographers can successfully undertake to report CT head examinations to a high standard. The adoption of both academic and clinical workload image banks that reflect disease examples and the prevalence that may logically be encountered in practice offers the potential for an accurate measure of performance of radiographer's abilities.

Words 245

## Introduction

Demand for Computed Tomography (CT) examinations in English National Health Service (NHS) Trusts between March 2015 and February 2016[1] approximated 5,007,188 examinations. For the month of February 2016 alone, there were 104,667 on the waiting list, 37,734 planned tests, and 126,500 unscheduled tests[2]. The NHS Imaging and Radiodiagnostic activity report[3] assessed the number of CT investigations between April 2013 to March 2014 at 5.2 million. Demonstrating a 10% increase from the previous year[3], a 43.1% rise over five years[4], and 160% growth over a decade[3]. The Centre for Workforce Intelligence (CfWI)[5] describe the likely factors that influenced the increase of imaging was due to growing/aging populations, an escalation in cancer diagnosis and chronic illness, screening programmes, and extended working hours. For CT cranial imaging the growth of imaging has also risen due to endorsement by the National Institute for Health and Care Excellence (NICE)[6, 7]as the first line imaging of choice[8] due to being fast, non-invasive and the routine availability of CT.

NICE recommends traumatic CT head scans to be reported within 1 hour[6], and Stroke CT examinations to be scanned and reported with 1 hour[7,9]. The Royal College of Radiologists (RCR)[10] recommend a formal report for all diagnostic examinations within a maximum of 2 days but acknowledge due to workforce shortages this does not occur[4]. With the current number of registered radiologists in the United Kingdom (UK) totalling 2,997 (4.7 working time equivalent consultant radiologists per 100,000 population)[4]. Leading to delays in cancer and serious illness diagnosis, prolonged hospital stays, and the subsequent increased registration of radiology departments to NHS risk registers[10].

In October 2014 the RCR[10] highlighted a month delay in results in 25% of English NHS trusts, follow-up evidence in February 2015 indicated 71% of NHS trusts had delays of over a month for reporting. Revealing over 2,883 unreported CT scans, estimated for all English NHS trusts to be up to 3,693[10]. By May 2016 the RCR stated[11] the backlog had escalated to 263,318 test results being delayed by more than a month (including 4,408 CT reports), affecting 75% of English NHS Trusts. One current response to the crisis that the NHS is attempting is a short-term and costly resolution through outsourcing of reporting to private companies at a cost of £73.8million (2014-15)[11].

A more efficient and long term approach would be a larger investment in a skills mix of reporting as promoted and endorsed jointly by the RCR and the Society and College of Radiographers (SCoR) [12-15]. Examples of such an approach have been demonstrated through surveys by the SCoR[16,17] and Snaith et al[18] illustrating a minimum of 179 radiology departments in the UK adopting and supporting advanced practice reporting radiographers. The intention of the skills mix of reporting is to supplement current service provision[16], reduce backlogs and improve reporting of acute conditions and early detection of malignant pathologies[19-21]. This advanced practice includes CT head reporting, which has been appraised since 1997[22], and evidenced in NHS practice since 2007[15,23-27].

The aim of this study was to progress the work of previous papers[25,26] through auditing the reporting performance of the latest cohort of radiographers (*n*=6) completing a postgraduate clinical reporting course in CT Head examinations. The Postgraduate Certificate (PgC) Clinical Reporting (CT Head) programme is a three 20 credit, level 7 modular course, validated by the SCoR. The course involves part-time distance learning over one year, which includes academic teaching (by lecturers, consultant radiographers and reporting radiographers) and clinical department tutorials by mentors.

## Method

Ethics and governance approval for this study were agreed by the Faculty of Health and Wellbeing Research Ethics Committee, with adherence to RCR[28] and General Medical Council (GMC)[29] best practice guidelines.

An element of the student's proficiency in the PgC programme (May 2015- April 2016) involved retrospectively reporting multiple banks of CT head examinations under controlled examination conditions[30]. Using low level lighting and high definition reporting monitors[31] that meet the current RCR reporting specification standards[32] (42cm, 1280 x 1084 screen resolution, >170 cd/m2 luminance, ≥250:1 luminance contrast ratio). The case studies were displayed in Digital Imaging and Communications in Medicine (DICOM) format using KPACS software[33] to enable manipulation[32].

A retrospectively amassed and anonymised DICOM collection of CT head examinations with referral details (patient gender, age, clinical symptoms and history) and clinical reports from previous research[25,26] was used. Each test bank of CT cases had been reported independently by three experienced consultant radiologists (and blinded to each other's reports, to reduce verification and work-up bias[34]). Expected responses (compiled from the reports of three consultant radiologists[35]) for all of the examinations were then agreed and approved by the programme panel and external examiner (independent consultant radiologist), to verify that a suitable and robust range of examples were incorporated. A wide range of pathological examples was featured to adequately evaluate the students' knowledge and demonstration of competence. The test banks incorporated normal cases up to 50% (variation occurred in each test bank shown in Tables 1 and 2) and a wide range of subtle and characteristic abnormal (single and multi-site) pathological examples (disease prevalence ranged from 50-100% over the test banks are shown in Tables 1 and 2). Pathological cases included: hematomas, haemorrhages, cranial fractures, Ischaemic infarctions, primary and secondary malignant and benign cerebral lesions, post-surgical interventions, degenerative changes and normal variants (to reduce spectrum bias[34]).

The radiographers (RR1-6) were provided with demographic details, including gender, age (18-92), referral source (in-patient, out-patient, General Practitioner (GP), Accident and Emergency (A&E) and clinical history for each case. The radiographers were instructed to comment if they deemed the examination to be normal or abnormal, and provide a detailed report of findings (describing the exact anatomical location) to justify and support the diagnosis in the form of a free text response. Including any secondary effects of the primary condition, such as the mass effect on surrounding structures and sulci, herniation of anatomy (and direction of herniation), and if a lesion was identified, the size (in mm) and lesion outline (smooth, nodular, ring, irregular, and contrast enhancing characteristics).

A recent article by Hardy et al[36] discussed the issue and influence of prevalence bias of pathology upon standard academic test bank construction and the resulting accuracy of results. Hardy et al[36] advocated test bank designs to move away from previously established academic models[25,26,37] to representative local clinical workloads to reduce bias of high abnormality prevalence that may potentially over-estimate observer's competency in abnormality detection. The PgC programme used a second tier of observer performance measures of a local clinical workload bank, to reflect lower reported incidence of abnormal cases. Using prospective clinical worklists in CT (to reduce population bias[34]) from a variety of referral sources, including in-patient, out-patient, GP and A&E, (age range 17-98). Total sample size of cases used was important to reduce the risk of Type II error (performance may not be statistically significant but clinically important[38,39]), each radiographer (RR1-6) was required to report a minimum of 375 prospective reports at the local NHS Trust from the daily CT worklist as a course requirement. The disease prevalence ranged between 22.4-60.6%, as shown in Table 1. These reports were blind reported by a consultant radiologist and further reviewed and arbitrated by a consultant radiologist and qualified and experienced CT head reporting radiographer for concordance (to the reference standard[34,35]). The daily worklists used General Electric (GE) Centricity RIS-i 5.0; the scans were produced on the five CT scanners in the NHS Trust (a mix of GE 64 slice CT and Siemens Definition Flash 128 slice CT). The reports were completed in open plan reporting rooms, using EIZO RadiForce RX340 54cm 3MP (1,000 cd/$m^2$ luminance, 1400:1 luminance contrast ratio) LCD workstation monitors, running GE Centricity Picture archiving and communication systems (PACS) Radiology RA1000 Workstation and Exam Manager PACS software to review the examinations.

Responses were classified (using a 2x2 contingency table[40]) as true positive (TP), true negative (TN), false positive (FP) or false negative (FN), using fractions (whole and partial) as described in a previous study[25]. Sensitivity, specificity, and accuracy were calculated using standard measures of observer performance[25,40]; mean values were further analysed using Cohens Kappa statistic for inter-observer variability[30,34]. One-way analysis of variance (ANOVA) from summary data and Tukey post-hoc test was used for multiple comparisons and significance testing between radiographers and against the test bank reference standard at 95% confidence intervals (95%CI).

## Results

All radiographers (RR1-6) completed the retrospective academic test banks and achieved the minimum number of the prospective test reports from the local clinical worklists. The primary outcome measures calculated participant mean accuracy (90.7%, 88.3-93.0%), sensitivity (86.9%, 85.8-88.2%), and specificity levels (94%, 89.6-98.3%) shown in Table 3. No inter-observer statistical significance was noted ($p$=0.000) in the results. The inter-observer ANOVA and Tukey multiple comparison tests were also conducted and shown in Table 3. The Kappa values for radiographer performance (Tables 2 and 3) across all test banks combined ($k$=0.8114) and for individual test banks, displayed a high Kappa score of agreement[41] (Table 4).

A recognition of the limitations when comparing results between the academic test banks ($n$=738 cases) and the local clinical workload test banks ($n$=2270 cases) needs to consider the difference of sample size that can influence results although this was not an aim of this audit.

## Discussion

Errors of FN in the case banks included major discrepancies of midline shift and herniation, cranial linear fractures, subtle subdural hematomas, and small acute infarctions in the cerebellum. Minor discrepancies included small chronic ischaemia and lacunar infarctions, scalp hematomas, white matter changes, previous mastoid surgery, and included under-calling small associated linear fractures in major trauma cases.

A Study by Abujudeh et al. (2010)[42] using CT abdomen and pelvis examinations reflected the complexity of interpreting multiple pathologies in cross-sectional imaging (with volumetric data reconstruction in multiple planes) establishing major discrepancy rates of 25-32%. These perceptual errors in failure to detect disease occurred when multiple lesions were present and combined with a failure of 'satisfaction of search' patterns[42-47]. Pinto et al[48], Stephens et al[49] and Lee et al[50] estimate errors of searching can approximate to 30-43%, and although the reasons are multi-faceted, the main factors are misinterpretations. Indeed, it may be difficult to underpin the average error rate in CT reporting, and it may even be under-estimated nationally.

The most common FP errors comprised major discrepancies of subtle haemorrhage, middle cerebral artery thrombus, lacunar infarction, and small vessel disease. Minor discrepancies of cerebral calcification, ventricular dilation, and involutional changes were also recorded. Overcalling of FPs also frequently occurred in elderly patients, which included white matter changes, differentiation between normal and pathological has been regarded to affect sensitivity rates in previous studies[25,26]. It could be argued that interpretation of the test banks in the academic environment may influence decision making[51] with the low level of risk associated in this context, some may have been over cautious, with the diagnosis of pathological conditions.

There has been a paucity of evidence on inter-observer radiographer performance of CT head examinations to compare against from current literature[22,25,26]. Using academic image test banks in this study allows results to be comparable to previous radiographers results[25,26]. Data for inter-observer accuracy from clinical workload test banks allowed a predicative value of radiographer's abilities to perform in a clinical environment, where no exact comparable study is available. By using random, but representative case studies that conform to routine practice allows judgement of competency in clinical practice[40,52,53] with strong results (87.7%; cases *n*=753 and 92.8%; cases *n*=1517, Table 2). Research by Le et al[54], using a sample size of 10 radiologists and 5 first year fellows, reviewing *n*=3,886 cases from an emergency department referral source displayed a similar accuracy rate (97.3%), with discordance of 2.7%. Further research by Erly et al[55], of an equal sample size (15 radiologists) and smaller case bank sample of *n*=716 (despite a significantly lower disease prevalence of 6.5%, using an emergency only referral source) demonstrated an equivalent range of measures of agreement (95%). Other studies have shown comparable results, including Schriger et al[56] employing a larger sample size of 36 radiologists reviewing a smaller sample bank of *n*=56 CT scans (75% disease prevalence of stroke referrals) established a lower accuracy of 83%. Likewise, a smaller study by McCarron et al[57] reviewed 9 radiologists reading *n*=77 CT head examinations, obtained an agreement of 86.6%.

The method used evidences current best practice policy in training, to expose radiographers to both academic retrospective example test banks and clinical prospective workload banks of images to present a varying case mix of normal/abnormal to reduce prevalence bias on interpretation. High-prevalence case banks have shown in training to nurture a desirable sensitivity-specificity compromise[58,59] in cases where abnormalities have major health implications. Likewise, 'context bias'[60] has been shown to influence the interpretation and evaluation of varying prevalence (pre-set high and low abnormality) test banks, which illustrates the complexity in achieving unbiased performance levels. Although this is not without its critics and challenges to provide exact measures of accuracy in performance using varying levels of disease prevalence[50,61,62].

Defining a satisfactory level of performance for CT head reporting by radiographers is a difficult task and often dependent upon adequate sample size determination and statistical power. Scally and Brealey[39] use an expected example of 80% sensitivity and 95% specificity (60% disease from 250-335 cases, using 95% CI), but note these figures will alter with varying prevalence of disease and case number in test banks.

Furthermore, it would be useful for future research to re-evaluate these results through clinical audit after qualification. Following best practice frameworks by the SCoR[14] and RCR[63] quality improvement guidelines, to maintain a record of measure of performance. This is an important practice in rapidly developing modalities where the volume of data per patient is increasing per examination, raising the complexity of reporting, which can be a factor contributing to misinterpretation errors of clinically important findings in discordant CT examinations[64,65].

Conclusion

The radiographer's performance demonstrated similar results to previous research[25,26] on observer performance and competency. The discrimination parameter of using a prospective random clinical workload model for testing radiographer's interpretative findings provided similar results to the academic retrospective test results, and the differences did not provide statistically different results.

The data suggests appropriately trained radiographers can successfully undertake reporting of CT head examinations to a high standard. Although, a recognition of the limitations of the sample size of participants impacts upon the generalisable nature of the results. The adoption of both academic and clinical test banks that reflect disease examples and the prevalence that may logically be encountered in practice offers the potential for an accurate measure of performance of radiographer's abilities. Recommendations from this study include further research to review the post-qualified clinical audits of reports for quality, consistency and concordance.

**Table 1.** Disease prevalence across clinical workload test banks.

| Test Bank | Number of cases | RR1 | RR2 | RR3 | RR4 | RR5 | RR6 | Mean Disease Prevalence |
|---|---|---|---|---|---|---|---|---|
| Clinical Workload Mixed Test Bank 1 | *n*=125 Minimum | 59.3% (*n*=128) | 38.4% (*n*=125) | 32% (*n*=125) | 66.9% (*n*=125) | 22.4% (*n*=125) | 34.1% (*n*=125) | **42.1%** |
| Clinical Workload Mixed Test Bank 2 | *n*=250 Minimum | 49.4% (*n*=263) | 50% (*n*=250) | 44.2% (*n*=253) | 60.6% (*n*=250) | 31.6% (*n*=251) | 38.3% (*n*=250) | **45.6%** |

**Table 2.** Comparison of mean observer outcomes by test bank

| Type of bank | Test Bank | Disease | Amount | Accuracy | | | Sensitivity | | | Specificity | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Prevalence | of cases | Mean | 95% CI | SD | Mean | 95% CI | SD | Mean | 95% CI | SD |
| 1.Manufactured Test Bank | Normal Case Bank | 0% | 5 | 100 | 100-100 | 0 | 0 | 0 | 0 | 100 | 100-100 | 0 |
| 2.Manufactured Test Bank | Trauma Case Bank | 100% | 7 | 96.4 | 94.0-98.7 | 2.27 | 96.4 | 94.0-98.7 | 2.27 | 0 | 0 | 0 |
| 3.Manufactured Test Bank | Degenerative Case Bank | 70% | 10 | 65.8 | 54.9-76.6 | 10.3 | 67.8 | 44.5-91.0 | 22.12 | 61.0 | 20.1-100 | 38.96 |
| 4.Manufactured Test Bank | Stroke Case Bank | 100% | 13 | 90.0 | 83.5-96.4 | 6.13 | 90.0 | 83.5-96.4 | 6.13 | 0 | 0 | 0 |
| 5.Manufactured Test Bank | Tumour Case Bank | 100% | 8 | 91.1 | 86.7-95.4 | 4.17 | 91.1 | 86.7-95.4 | 4.17 | 0 | 0 | 0 |
| 6.Manufactured Test Bank | Mixed Case Bank 1 | 63% | 16 | 82.2 | 72.5-91.8 | 9.15 | 82.1 | 71.4-92.7 | 10.1 | 83.3 | 67.6-98.9 | 14.93 |
| 7.Manufactured Test Bank | Mixed Case Bank 2 | 50% | 24 | 95.4 | 87.4-100 | 7.6 | 95.6 | 87.6-100 | 7.59 | 95.2 | 86.5-100 | 8.19 |
| 8.Manufactured Test Bank | Mixed Case Bank 3 | 50% | 40 | 93.1 | 91.0-95.1 | 1.94 | 91.2 | 89.5-92.8 | 1.56 | 95.0 | 90.3-99.6 | 4.47 |
| 1.Clinical Workload Test Bank | Mixed Case Bank 1 | 42% (22.4-66.9%) | Min. 125 | 87.7 | 84.9-90.4 | 2.64 | 75.9 | 68.1-83.6 | 7.41 | 94.0 | 89.0-98.9 | 4.76 |
| 2.Clinical Workload Test Bank | Mixed Case Bank 2 | 45% (31.6-60.9%) | Min. 250 | 92.8 | 89.4-96.1 | 3.21 | 89.8 | 86.9-92.6 | 2.71 | 94.8 | 90.2-99.3 | 4.33 |

**Table 3.** Standardised performance results across all test banks

| Radiographers | Number | Accuracy | | | Sensitivity | | | Specificity | | | Cohens Kappa | | Cohens Kappa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | of cases | Mean | 95% CI | SD | Mean | 95% CI | SD | Mean | 95% CI | SD | Unweighted | 95% CI | Linear Weighted | 95% CI |
| RR1 | 514 | 86.8 | 83.5-89.7 | 10.68 | 87.5 | 84.5-90.1 | 28.94 | 86.0 | 82.2-89.2 | 45.55 | 0.7346 | 0.6756-0.7936 | 0.7346 | 0.6756-0.7936 |
| RR2 | 498 | 92.2 | 89.4-94.1 | 16.74 | 88.9 | 86.1-90.8 | 31.9 | 95.6 | 92.8-97.5 | 44.31 | 0.8444 | 0.7974-0.8914 | 0.8444 | 0.7975-0.8913 |
| RR3 | 501 | 91.3 | 88.5-93.2 | 9.75 | 85.7 | 82.6-87.8 | 29.64 | 96.0 | 93.5-97.8 | 46.26 | 0.8237 | 0.7736-0.8738 | 0.8237 | 0.7738-0.8736 |
| RR4 | 498 | 89.6 | 86.6-91.7 | 8.12 | 87.5 | 85.1-89.2 | 28.71 | 93.1 | 89.1-95.9 | 45.75 | 0.7831 | 0.7272-0.839 | 0.7831 | 0.7276-0.8386 |
| RR5 | 499 | 92.7 | 90.1-94.7 | 7.57 | 86.4 | 82.8-89.0 | 29.64 | 96.5 | 94.3-98.0 | 47.8 | 0.8427 | 0.7932-0.8922 | 0.8427 | 0.7933-0.8921 |
| RR6 | 498 | 92.3 | 89.6-93.9 | 10.26 | 85.8 | 82.8-87.7 | 30.99 | 97.2 | 94.8-98.6 | 46.96 | 0.8403 | 0.7919-0.8887 | 0.8403 | 0.7921-0.8885 |
| ***n*=6** | **3008** | **90.7** | **88.3-93.0** | **10.5** | **86.9** | **85.8-88.2** | **29.97** | **94.0** | **89.6-98.3** | **46.1** | **0.8114** | **0.765-0.858** | **0.8114** | **0.7599-0.8629** |

**Table 4.** Value of Cohens Kappa Strength of agreement[42]

| Kappa score | Agreement strength |
|---|---|
| <0.00 | Poor |
| 0.00-0.20 | Slight |
| 0.21-0.40 | Fair |
| 0.41-0.60 | Moderate |
| 0.61-0.80 | Substantial |
| 0.81-1.00 | Almost perfect |

References

1. NHS England. **February 2016 Statistical Report: Waiting times and activity for diagnostic tests and procedures**. January 2006 to February 2016. April 2016 https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2016/04/DWTA-Report-February-2016-yg25r.pdf (Accessed 17 June 2016).

2. NHS England. **Monthly Diagnostics Commissioner February 2016**. April 2016 https://www.england.nhs.uk/statistics/statistical-work-areas/diagnostics-waiting-times-and-activity/monthly-diagnostics-waiting-times-and-activity/monthly-diagnostics-data-2015-16/ (Accessed 17 June 2016).

3. Great Britain. **NHS Imaging and Radiodiagnostic activity 2013/2014. Leeds: NHS England Analytical Services (Operations)**; 2014 https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/04/KH12-release-2013-14.pdf (Accessed 17 June 2016).

4. Royal College of Radiologists. **Clinical Radiology UK Workforce Census Report 2012.** 2014. https://www.rcr.ac.uk/sites/default/files/BFCR(14)14_CR_Census.pdf (Accessed 17 June 2016).

5. Centre for Workforce Intelligence. **Securing the future workforce supply: Clinical radiology stocktake. London: Centre for Workforce Intelligence**2012. http://www.cfwi.org.uk/publications/securing-the-future-workforce-supply-clinical-radiology-stocktake/@@publication-detail (Accessed 17 June 2016).

6. The National Institute for Health and Care Excellence. **Head Injury: assessment and early management [CG176]** 2014. https://www.nice.org.uk/guidance/cg176 (Accessed 17 June 2016).

7. The National Institute for Health and Care Excellence**. Stroke [CG68] 2008**. https://www.nice.org.uk/guidance/cg68 (Accessed 17 June 2016).

8. Lollli, V., Pezzullo, M., Delpierre, I., Sadeghi, N. **MDCT imaging of traumatic brain injury**. British Journal of Radiology: 2016. 89 201508-49. http://www.birpublications.org/doi/abs/10.1259/bjr.20150849 (Accessed 17 June 2016).

9. The Department of Health. **National Stroke Strategy. 2007**. http://clahrc-gm.nihr.ac.uk/wp-content/uploads/DoH-National-Stroke-Strategy-2007.pdf (Accessed 17 June 2016).

10. The Royal College of Radiologists. **Unreported X-rays, computed tomography (CT) and magnetic resonance imaging (MRI) scans: Results of a snapshot survey of English National Health Service (NHS) trusts**. 2015.

https://www.rcr.ac.uk/sites/default/files/publication/Unreported_studies_Feb2015.pdf (Accessed 17 June 2016).

11. The Faculty of Clinical Radiology. **Diagnostic radiology – our patients are still waiting...** 2016. https://www.rcr.ac.uk/sites/default/files/backlog_survey_feb_2016.pdf (Accessed 17 June 2016).

12. The Royal College of Radiologists and the Society and College of Radiographers. **Team working in clinical imaging**. 2012 http://www.sor.org/sites/default/files/document-versions/BFCR(12)9_Team.pdf (Accessed 17 June 2016).

13. The Department of Health. **Radiography skills mix: a report on the four-tier service delivery model**. 2003. http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4007123 (Accessed 17 June 2016).

14. The Society of Radiographers. **Preliminary Clinical Evaluation and Clinical Reporting by Radiographers: Policy and Practice Guidance**. 2013. http://www.sor.org/learning/document-library/preliminary-clinical-evaluation-and-clinical-reporting-radiographers-policy-and-practice-guidance (Accessed 17 June 2016).

15. The Society of Radiographers. **The Role of the Radiographer in Stroke Management.** 2010. http://www.sor.org/learning/document-library/role-radiographer-stroke-management-0 (Accessed 17 June 2016).

16. The Society and College of Radiographers. **Scope of radiographic practice survey 2012**. 2012 http://www.sor.org/learning/document-library/scope-radiographic-practice-survey-2012 (Accessed 17 June 2016).

17. The Society and College of Radiographers. **Diagnostic Radiography UK Workforce Report 2014**. 2014. http://www.sor.org/learning/document-library/diagnostic-radiography-uk-workforce-report-2014 (Accessed 17 June 2016).

18. Snaith, B., Hardy, M., Lewis, E.F. Radiographer Reporting in the UK: A Longitudinal Analysis. *Radiography* 2015; 21.2:119-123. http://www.sciencedirect.com/science/article/pii/S1078817414001382 (Accessed 17 June 2016).

19. The Department of Health. **The NHS Plan: A plan for investment, a plan for reform**.2000. http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/publicationsandstatistics/publications/publicationspolicyandguidance/dh_4002960 (Accessed 17 June 2016).

20. The Department of Health. **Equity and excellence: Liberating the NHS.** 2010. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf (Accessed 17 June 2016).

21.The Department of Health. **Improving outcomes: A Strategy for Cancer**. 2011. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213785/dh_123394.pdf (Accessed 17 June 2016).

22. Craven, C, M., Blanshard, K, S. **Computed tomography (CT) head scans reported by an experienced CT radiographer**. *Radiography.* 1997 31; 3(2):105-11. http://www.sciencedirect.com/science/article/pii/S1078817497900158 (Accessed 17 June 2016).

23. Price, R, C., Le Masurier, S, B. **Longitudinal changes in extended roles in radiography: a new perspective**. *Radiography*. 2007, 28; 13(1):18-29. http://www.sciencedirect.com/science/article/pii/S1078817405001574 (Accessed 17 June 2016).

24. Clarke, R., Allen, D., Arnold, P., Snaith, B. **Implementing radiographic CT head reporting: The experiences of students and managers**. *Radiography*. 2014. 31; 20(2):117-20. http://www.sciencedirect.com/science/article/pii/S107881741300151X (Accessed 17 June 2016).

25. Lockwood, P., Piper, K., Pittock, L. **CT head reporting by radiographers: Results of an accredited postgraduate programme**. *Radiography*. 2015; 21(3):e85-9. http://www.sciencedirect.com/science/article/pii/S1078817414001540 (Accessed 17 June 2016).

26. Lockwood, P., Piper, K**. AFROC analysis of reporting radiographer's performance in CT head interpretation**. *Radiography*. 2015; 21(3):e90-5. http://www.sciencedirect.com/science/article/pii/S1078817415000504 (Accessed 17 June 2016).

27. Lockwood, P**. An economic evaluation of introducing a skills mix approach to CT head reporting in clinical practice**. *Radiography*. 2015; 22(2): 124-130. http://www.sciencedirect.com/science/article/pii/S1078817415001169 (Accessed 17 June 2016).

28. The Royal College of Radiologists. **Standards for patient consent particular to radiology.** 2nd ed.; 2012. https://www.rcr.ac.uk/standards-patient-consent-particular-radiology-second-edition (Accessed 17 June 2016).

29. The General Medical Council. **Supplementary Guidance. Making and using visual and audio recordings of patients**.2011 http://www.gmc-uk.org/Making_and_using_visual_and_audio_recordings_of_patients_2011.pdf_40338254.pdf (Accessed 17 June 2016).

30. Brealey, S. **Measuring the effects of image interpretation: an evaluative framework**. *Clinical radiology* 2001, 56(5), pp.341-347. http://www.sciencedirect.com/science/article/pii/S0009926001906786 (Accessed 17 June 2016).

31. Prime, N, J., Paterson, A, M., Henderson, P, I. **The development of a curriculum: a case study of six centres providing courses in radiographic reporting**. *Radiography* 1999; 5 (2):63-70. http://www.sciencedirect.com/science/article/pii/S1078817499900330 (Accessed 17 June 2016).

32. The Royal College of Radiologists. **Picture archiving and communication systems (PACS) and guidelines on diagnostic display devices.** 2012. https://www.rcr.ac.uk/sites/default/files/docs/radiology/pdf/BFCR(12)16_PACS_DDD.pdf (Accessed 17 June 2016).

33. IMAGE Information Systems. **K-PACS (Version 1.6.0).** 2008 http://www.k-pacs.net/10555.html (Accessed 17 June 2016).

34. Brealey, S., Scally, A, J. **Bias in plain film reading performance studies**. *The British journal of radiology*. 2001 Apr; 74 (880):307-16. http://www.birpublications.org/doi/abs/10.1259/bjr.74.880.740307 (Accessed 17 June 2016).

35. Brealey, S., Scally, A, J., Thomas, N, B. **Methodological standards in radiographer plain film reading performance studie**s. *The British journal of radiology*. 2002;75(890):107-13. http://www.birpublications.org/doi/abs/10.1259/bjr.75.890.750107 (Accessed 17 June 2016).

36. Hardy, M., Flintham, K., Snaith, B., Lewis, E, F. **The impact of image test bank construction on radiographic interpretation outcomes: A comparison study**. *Radiography*. 2015 22(2); 166-170. http://www.sciencedirect.com/science/article/pii/S1078817415001303 (Accessed 17 June 2016).

37. Piper, K., Paterson, A., Godfrey, R. **Accuracy of radiographers' reports in the interpretation of radiographic examination of the skeletal system: a review of 6796 cases**. *Radiography* 2005; 11(1):27-34. http://www.sciencedirect.com/science/article/pii/S1078817404000781 (Accessed 17 June 2016).

38. Brealey, S., Scally, A, J. **Methodological approaches to evaluating the practice of radiographers' interpretation of images: A review**. *Radiography.* 2008 31; 14:e46-54. http://www.sciencedirect.com/science/article/pii/S1078817408000023 (Accessed 17 June 2016).

39. Scally, A, J., Brealey, S. **Confidence intervals and sample size calculations for studies of film-reading performance**. *Clinical radiology.* 2003;31;58(3):238-46. http://www.sciencedirect.com/science/article/pii/S0009926002004579 (Accessed 17 June 2016).

40. Langlotz, C, P. **Fundamental Measures of Diagnostic Examination Performance: Usefulness for Clinical Decision Making and Research 1.** *Radiology*. 2003; 228(1):3-9. http://pubs.rsna.org/doi/abs/10.1148/radiol.2281011106 (Accessed 17 June 2016).

41. Landis, J.R., Koch, G, G. **The measurement of observer agreement for categorical data.** *Biometrics*. 1977 Mar 1:159-74. http://www.jstor.org/stable/2529310?seq=1#page_scan_tab_contents (Accessed 17 June 2016).

42. Abujudeh, H, H., Boland, G, W., Kaewlai, R., Rabinar, P., Halpern, E, F., Gazelle, G,S. and Thrall, J, H. **Abdominal and pelvic computed tomography (CT) interpretation: discrepancy rates among experienced radiologists**. *European Radiology* 2010, 20; 1952–1957. http://link.springer.com/article/10.1007/s00330-010-1763-1#/page-1 (Accessed 17 June 2016).

43. Ashman, C, J., Yu, J. Wolfman, S. **Satisfaction of search in osteoradiology**. *American Journal of Roentgenology* 2000, 175(2), 541-544. http://www.ajronline.org/doi/abs/10.2214/ajr.175.2.1750541 (Accessed 17 June 2016).

44. Renfrew, D, L., Franken Jr, E, A., Berbaum, K, S., Weigelt, F,H., Abu-Yousef, M,M. **Error in radiology: classification and lessons in 182 cases presented at a problem case conference**. *Radiology.* 1992; 183(1):145-50. http://pubs.rsna.org/doi/abs/10.1148/radiology.183.1.1549661 (Accessed 17 June 2016).

45. Berbaum, K, S., Franken Jr, E, A., Dorfman, D, D., Rooholamini, S, A., Kathol, M, H., Barloon, T, J., Behlke, F, M., Sato, Y., Lu, C, H., El-Khoury, G, Y., Flinckinger, F, W. **Satisfaction of search in diagnostic radiology**. *Investigative radiology*. 1990 Feb 1; 25(2):133-40. http://journals.lww.com/investigativeradiology/Abstract/1990/02000/Satisfaction_of_Search_in_Diagnostic_Radiology.6.aspx (Accessed 17 June 2016).

46. Drew, T., Võ, M, L., Wolfe, J, M. **The invisible gorilla strikes again sustained inattentional blindness in expert observers**. *Psychological science.* 2013; 1;24(9):1848-53. http://pss.sagepub.com/content/24/9/1848.short (Accessed 17 June 2016).

47. Venjakob, A, C., Marnitz, T., Phillips, P., Mello-Thoms, C, R. **Image Size Influences Visual Search and Perception of Hemorrhages When Reading Cranial CT An Eye-Tracking Study. Human Factors**. *The Journal of the Human Factors and Ergonomics Society*. 2016 Mar 2:0018720816630450. http://hfs.sagepub.com/content/early/2016/03/01/0018720816630450.abstract (Accessed 17 June 2016).

48. Pinto, A., Acompara, C., Pinto, F., Kourdioukova, E., Romano, L. Verstraete, K. **Learning from Diagnostic Errors: A Good Way to Improve Education in Radiology**. *European Journal of Radiology* 2011, 78 (3).372-376 http://www.sciencedirect.com/science/article/pii/S0720048X1000639X (Accessed 17 June 2016).

49. S. Stephens, S., Martin, I., Dixon, A, K. **Errors in abdominal computed tomography**. *Journal of Medical Imaging*, 3 (1989), 281–287 https://inis.iaea.org/search/search.aspx?orig_q=RN:21040018 (Accessed 17 June 2016).

50. Lee, C.S., Nagy, P.G., Weaver, S.J. and Newman-Toler, D.E. **Cognitive and System Factors Contributing to Diagnostic Errors in Radiology.** *American Journal of Radiology* 2013, 201(3)611-617. http://www.ajronline.org/doi/abs/10.2214/AJR.12.10375 (Accessed 17 June 2016).

51. Reed, W, M., Ryan, J, T., McEntee, M, F,. Evanoff,, M, G,. Brennan, P, C. **The effect of abnormality-prevalence expectation on expert observer performance and visual search**. *Radiology* 2011; 258 (3):938–43. http://pubs.rsna.org/doi/abs/10.1148/radiol.10101090 (Accessed 17 June 2016).

52. Pusic, M, V., Andrews, J, S., Kessler, D, O., Teng DC, Pecaric MR, Ruzal-Shapiro C, Boutis K. **Prevalence of abnormal cases in an image bank affects the learning of radiograph interpretation**. *Medical education* 2012, 1;46 (3):289-98. http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2923.2011.04165.x/full (Accessed 17 June 2016).

53. Nocum, D, J., Brennan, P, C., Huang, R, T., Reed, W, M. **The effect of abnormality-prevalence expectation on naïve observer performance and visual search**. *Radiography*. 2013;19 (3):196-9. http://www.sciencedirect.com/science/article/pii/S1078817413000448 (Accessed 17 June 2016).

54. Le, A, H., Licurse, A., Catanzano, T, M. **Interpretation of head CT scans in the emergency department by fellows versus general staff non-neuroradiologists: a closer look at the effectiveness of a quality control program**. *Emergency Radiology* 2007, 14(5): 311-316. http://link.springer.com/article/10.1007/s10140-007-0645-6#/page-1 (Accessed 17 June 2016).

55. Erly, W, K., Ashdown, B, C., Lucio, R, W., Carmody, R, F., Seeger, J, F., Alcala, J, N. **Evaluation of emergency CT scans of the head: is there a community standard?**. *American Journal of Roentgenology*. 2003;180 (6):1727-30. http://www.ajronline.org/doi/abs/10.2214/ajr.180.6.1801727 (Accessed 17 June 2016).

56. Schriger, D, L., Kalafut, M., Starkman, S., Krueger, M., Saver, J, L. **Cranial computed tomography interpretation in acute stroke**. *The Journal of the American Medical Association* 1998, 279(16): 1293-1297. http://jama.jamanetwork.com/article.aspx?articleid=187465 (Accessed 17 June 2016).

57. McCarron, M, O., Sands, C., McCarron, P. **Neuroimaging reports in a general hospital: Results from a quality-improvement program**. *Clinical Neurology and Neurosurgery* 2010, 112(1): 54-58. http://www.sciencedirect.com/science/article/pii/S0303846709002716 (Accessed 17 June 2016).

58. Green, D, M., Swets, J, A. **Signal Detection Theory and Psychophysics**. New York, NY: John Wiley & Sons 1967; 332–3.

59. Wolfe, J, M., Horowitz, T, S., Van Wert, M, J., Kenner, N, M., Place, S, S., Kibbi, N. **Low target prevalence is a stubborn source of errors in visual search tasks.** *Journal of Experimental Psychology* 2007; 136 (4):623–38. http://psycnet.apa.org/psycinfo/2007-16657-007 (Accessed 17 June 2016).

60. Egglin, T, K., Feinstein, A, R. **Context bias: a problem in diagnostic radiology**. *Jama*. 1996 4; 276(21):1752-5. http://jama.jamanetwork.com/article.aspx?articleid=411324 (Accessed 17 June 2016).

61. Mower, W, R. **Evaluating bias and variability in diagnostic test reports**. *Annals of Emergency Medicine* 1999; 33 (1):85–91. http://www.sciencedirect.com/science/article/pii/S0196064499704221 (Accessed 17 June 2016).

62. Gur, D., Rockette, H, E., Warfel, T., Lacomis, J, M., Fuhrman, C, R. **From the laboratory to the clinic: the "prevalence effect"**. *Academic Radiology* 2003, 10 (11); 1324–1326. http://www.sciencedirect.com/science/article/pii/S1076633203004665 (Accessed 17 June 2016).

63. The Royal College of Radiologists. **A guide to job planning in clinical radiology**; 2013 https://www.rcr.ac.uk/sites/default/files/BFCR(13)9_job_plans.pdf (Accessed 17 June 2016).

64. Lauritzen, P, M., Andersen, J, G., Stokke, M, V., Tennstrand, A,L., Aamodt, R., Heggelund, T., Dahl, F,A., Sandbæk, G., Hurlen, P. Gulbrandsen, P. **Radiologist-initiated double reading of abdominal CT: retrospective analysis of the clinical importance of changes to radiology reports**. *BMJ quality & safety* 2016; 0:1-9. http://qualitysafety.bmj.com/content/early/2016/03/24/bmjqs-2015-004536.abstract (Accessed 17 June 2016).

65. Lauritzen, P, M., Stavem, K., Andersen, J, G., Stokke, M,V., Tennstrand ,A, L., Bjerke, G., Hurlen, P., Sandbæk, G., Dahl, F, A., Gulbrandsen, P**. Double reading of current chest CT examinations: Clinical importance of changes to radiology reports**. *European journal of radiology* 2016; 85(1):199-204. http://www.sciencedirect.com/science/article/pii/S0720048X1530156X (Accessed 17 June 2016).

**Acknowledgements**

The author would like to thank all the radiographers and radiologists who took part in interpreting the image banks in this study, and for their time and dedication to the study.

**Funding**

There are no financial conflicts of interest.

**Conflict of interest statement**

The author *a* is a pathway lead for the postgraduate CT head reporting course at CCCU.