



CREaTE

Canterbury Research and Theses Environment

Canterbury Christ Church University's repository of research outputs

<http://create.canterbury.ac.uk>

Please cite this publication as follows:

Lockwood, P. and Piper, K. (2015) AFROC analysis of reporting radiographer's performance in CT head interpretation. *Radiography*, 21 (3). e90-e95.

Link to official URL (if available):

<http://dx.doi.org/10.1016/j.radi.2015.04.001>

This version is made available in accordance with publishers' policies. All material made available by CReaTE is protected by intellectual property law, including copyright law. Any use made of the contents should comply with the relevant law.

Contact: create.library@canterbury.ac.uk



AFROC analysis of reporting radiographer's performance in CT head interpretation

P. Lockwood^a, K. Piper^b

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

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

Abstract

Aim: A preliminary small scale study to assess the diagnostic performance of a limited group of reporting radiographers and consultant radiologists in clinical practice undertaking computer tomography (CT) head interpretation.

Method: A multiple reader multiple case (MRMC) alternative free response receiver operating characteristic (AFROC) methodology was applied for this study. Utilising an image bank of 30 CT head examinations, with a 1:1 ratio of normal to abnormal cases. A reference standard was established by double reporting the original reports using two additional independent consultant radiologists with arbitration of discordance by the researcher. Twelve observers from six southern National Health Service (NHS) trusts were invited to participate. The results were compared for accuracy, agreement, sensitivity, specificity. Data analysis used AFROC and area under the curve (AUC) with standard error against the ground truth.

Results: The reporting radiographers results demonstrated a mean sensitivity rate of 88.7% (95% CI 82.3 to 95.1%), specificity 95.6% (96% CI 90.1 to 100%) and accuracy of 92.2% (95% CI 89.3 to 95%). The consultant radiologists mean sensitivity rate was 83.35% (95% CI 80 to 86.7%), specificity 90% (95% CI 86.7 to 93.3%) and accuracy of 86.65% (95% CI 83.3 to 90%). Observer performance between the two groups was compared with AFROC, AUC, and standard error analysis ($p=0.94$, SE 0.202).

Conclusion: The findings of this research indicate that within a limited study, a small group of reporting radiographers demonstrated high levels of diagnostic accuracy in the interpretation of CT head examinations that was equivalent to a small selection of consultant radiologists.

Highlights

- We assessed reporting radiographers and consultant radiologists in a clinical setting.
- This was a small scale retrospective multi-reader multi-case multi-site study.
- AFROC used lesion-based decisions rather than case-based decisions.
- Within a limited study the observer performance was high in CT head interpretation.
- Parallels were drawn with published results from other CT head interpretation studies.

Introduction

1
2 The number of head injured patients attending district general hospitals has been estimated by the
3 United Kingdom (UK) Acquired Brain Injury Forum¹ during 2011-2012 to be around 353,059 UK
4 patients. These figures estimate around 558 per 100,000 of the population experience head injuries
5 each year. This represents a 33.5% increase in the last ten years (10,000-20,000 per year in the UK)
6 of admissions for severe traumatic brain injuries.
7

8
9
10
11 Both the National Healthcare Service (NHS) and the Department of Health (DoH)^{2, 3, 4, 5, 6} have a
12 strong ethos of developing and improving patient outcomes and service delivery. With the NHS
13 currently undertaking the 'Nicholson Challenge' (2006-2015)⁷ to generate extra productivity and
14 service quality improvement, set by Sir David Nicholson. Within radiology additional NHS drivers for
15 change include pressures from DoH targets of the acute 4 hour waiting time⁸, cancer 'referral to
16 treatment' 18 week target waits⁶, and the National Diagnostics Imaging Board⁹ policies on reporting
17 targets. Specifically within computed tomography (CT) as a modality, National Institute for Health
18 and Clinical Excellence (NICE) guidelines^{10,11} of reporting turnaround timeframes for stroke and head
19 injury examinations have changed historic working practices with the need for urgent 30 minute to 1
20 hour verbal and written CT head reports. This coupled with an increase in the amount of CT
21 examinations that have increased by 33.5% a year since 2008¹² have emphasised the need to re-
22 evaluate how the service delivery can accommodate future pressure.
23

24
25
26
27
28
29
30
31
32
33
34 Barriers to improving current working practices include staff shortages to implement new guidelines,
35 and the current dilemma of implementing a full 7 day service delivery with restricted service
36 capacity. Within diagnostic imaging, the Royal College of Radiologists (RCR) Clinical Radiology
37 Workforce Report¹² recommended a level of 47 consultant radiologists per million of the population
38 for the UK. The reported RCR¹² level for the south of England was 30 per million, the lowest of all
39 regional variations. With a deficit of 210 unfilled NHS consultant radiologist posts in the UK, the
40 RCR¹² report advised that the current consultant radiologist workforce does not meet the required
41 needs of the radiology service demand. The report indicated 85% of UK radiology departments
42 reporting workload was not being adequately completed by the consultant radiologist workforce.
43 The RCR¹² estimated the shortfall in reporting to be 47% of all examinations were left unreported in
44 2011, whilst recommending the best approach to tackle the deficit has been the adoption of
45 reporting radiographers.
46
47
48
49
50
51
52
53
54

55
56 In identifying potential ways to reduce reporting delays and increase service provision, a skills mix of
57 reporting has been promoted and endorsed jointly by the RCR and the Society and College of
58 Radiographers (SCoR)^{13, 14, 15}. Examples of such an approach have been demonstrated in surveys by
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

the SCoR^{16, 17} which showed at least 17 NHS trusts in the UK had adopted and supported role extension of reporting radiographers to supplement their service provision by 2012¹⁶. This has helped to improve service delivery of reporting traumatic injuries and assisted in the early detection of pathological conditions and cancers^{2, 4, 5}.

Aims and Objectives

The study hypothesis predicted reporting radiographers would have a diagnostic accuracy comparable or equal to consultant radiologists in CT head interpretation in a clinical setting. To answer the hypothesis, the research study set inter and intra-participant objectives within the study: Identifying statistical interpretation results for variation or equivalence rates between two groups of participants (consultant radiologists and reporting radiographers) undertaking the same image bank analysis.

Methodology

The design followed a multiple reader multiple case (MRMC) retrospective study of CT head interpretation by reporting radiographers ($n=6/6$) and consultant radiologists ($n=2/6$) at 6 NHS hospitals within the southern region of the UK.

Chang¹⁸ suggests that any experimental study which evaluates the efficiency of reporting standards by Bayesian analysis must use an explicitly defined reference standard. The study adopted a retrospective method using patient cases with known true disease status from a collection of 125 cases previously obtained by the University for teaching and research. This had been additionally double reported by two independent consultant radiologists. Brealey¹⁹ and Robinson²⁰ advise that employing a triple approach to obtaining a retrospective reference standard enforces validity of the reference standard.

Brealey¹⁹ discusses issues of internal validity of research as the amount and range of presenting conditions used in the control group (image case bank) in diagnostic performance studies. The CT head examinations reflected a suitable range of subtle and textbook examples to determine high levels of accuracy to remove internal validity concerns. Displaying a fair representation of pathologies as recommended by Robinson et al²⁰ and Brealey¹⁹, and similar to methods used in studies by Briggs et al²¹, McCarron et al²², Ery et al²³, Strub et al²⁴, and Gallagher et al²⁵. Concerning the relative frequency of cases with and without disease in the study sample, Brealey¹⁹; Metz²⁶; Brealey and Scally²⁷; Thompson et al²⁸; and Piper, Paterson and Ryan²⁹ endorsed a balanced approach to the ratio of normal to abnormal conditions (1:1).

1 The test bank was reviewed within the participant's clinical departments under ambient lighting
2 settings for radiological reporting environments. The images were displayed on a Toshiba Windows
3 Notebook Laptop with a Liquid Crystal Display (LCD) monitor with resolution of 1280x1024. The
4 laptop had been calibrated to the Digital Imaging and Communications in Medicine (DICOM) part 14
5 Greyscale Standard Display Function (GSDF) with the VeriLUM software programme³⁰. Quality checks
6 were performed on the Laptop LCD monitor prior to each test with a standard diagnostic imaging
7 Society of Motion Picture and Television Engineers (SMPTE) reference pattern for spatial uniformity
8 of luminance and temporal luminance stability as recommended by the RCR³¹. An independent
9 PACS system of iQ-View software programme³² was used to display the cases in a sequential order.

10
11
12
13
14
15
16 The recruitment criteria of participants required completion of SCoR accredited training and
17 qualification in CT Head reporting, with completion of a period of post training experience of
18 independent reporting within an NHS hospital trust. Obuchowski³³ proposes designs of an MRMC
19 Phase 1 pilot study only requires a small selection of 10-50 cases, of which we choose 30 cases from
20 the bank of 125 cases to be double reported. Obuchowski^{33, 34} also suggests in MRMC studies of
21 difficult cases in terms of disease prevalence and appearances should include between 5-10
22 observers to compare groups of observer's performance.

23
24
25
26
27
28
29 Six reporting radiographers were invited to participate ($n=6/6$ completed the study), and six
30 consultant radiologists were invited to participate in the study ($n=2/6$ completed the study). Each
31 participant was provided with a copy of instructions detailing the patient history, presenting
32 symptoms, age, gender and referral source, for each case. The participants received these
33 instructions in person by the researcher and were collected after each participant session for
34 compiling of the raw data.

35
36
37
38
39
40
41 The study required participants to record their findings as either normal or abnormal. If the case was
42 normal they marked the case 0, and moved on to the next case. If the participant deemed the case
43 to be abnormal, they recorded a score of 1-4 (very low to very high confidence of an abnormality)
44 and recorded the name of the pathological condition seen, the anatomical location of the
45 condition/disease and their confidence score of the interpreted pathology. The confidence
46 classification score and free response text allowed the results to be analysed by true positive (TP),
47 true negative (TN), false positive (FP), and false negative (FN). Allowing calculations of accuracy,
48 agreement, sensitivity and specificity using a method adopted by Piper, Ryan and Paterson²⁹ and
49 Piper, Buscall and Thomas³⁵. When considering the accuracy of interpreting radiographic
50 examinations, Obuchowski³⁴ suggests high accuracy to be 90% (specificity / sensitivity 80%).
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Statistical evaluation employed alternative free response receiver operating characteristic (AFROC)
2 curve analysis and area under the curve (AUC) comparison to measure performance. In MRMC
3 studies the use of the AFROC method is ideal when the amount of abnormalities and locations are
4 required to be identified, and ranked each against values according to the confidence levels.
5 Particular attention to the location of the lesion identified to within an acceptance radius (proximity
6 criterion emanating for the centre of the suspected lesion-location (LL) Thompson et al³⁶) allowed
7 the researcher to class the participant's responses as LL (true location of abnormality =TP) or non-
8 lesion (NL) location (wrong location of abnormality = FP or FN).
9

10 Chakraborty³⁹ cautions that the conventional receiver operating characteristic (ROC) paradigm does
11 not distinguish statistical differences for incorrect location (FP), if multiple lesions are present the
12 ROC would classify a TP result even if all the abnormalities were not identified or anatomical location
13 described correctly. Significant clinical implications which may impact on treatment cannot be
14 accounted for in this scenario. Chakraborty³⁹ advocates AFROC curves over conventional ROC curves,
15 as they provide an increased power due to lesion localization.
16

17 Jackknife free-response ROC (AJFROC) calculations were considered for the data analysis but were
18 rejected on the grounds that the output and statistical tests assume paired analysis of two
19 modalities not readers. The use of single modalities violates the assumption of the calculations.
20 Additionally a test run produced a zero score for the incorrect localisation fraction (ILF), thus it had
21 in this instance no power advantage over AFROC analysis.
22

23 Conventional ROC plotting generates a curve using the axis of true positive fraction (TPF) in this case
24 sensitivity, versus false positive fraction (FPF) which is calculated as 1-specificity (Thompson et al³⁶).
25 AFROC plotting uses a mixture of conventional ROC methodology and free response ROC (FROC)
26 calculations. FROC is a variant of ROC which was designed to reduce the ROC limitations of a binary
27 yes/no answer and instead determine scoring of multiple lesions per case with unlimited location
28 identification (Thompson et al³⁶). FROC calculations replace the FPF with non-lesion fractions (NLF)
29 on the x-axis, and number of lesions (lesion location fraction (LLF) on the y-axis. AFROC is a
30 combination of both paradigms and uses LLF on the y-axis (the same as FROC) and FPF on the x-axis
31 (the same as conventional ROC calculations) Thompson et al³⁶.
32

33 The study was approved by the university research ethics and governance committee and
34 conformed to Section 33 of the UK Data Protection³⁸. All the cases had been obtained from a pre-
35 existing DICOM digital teaching library (DTL). The radiology source data (identifying narrative
36 elements including staff names, hospital name, and identifying patient data) had been manually
37 removed to anonymise the images. This practice follows Cosson and Willis³⁹ guidance from the
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 National Information Governance Board for Health and Social Care, and the General Medicine
2 Council⁴⁰.

3 4 **Results**

5
6 The results for the reporting radiographers ($n=6$, Ranked RR1-RR6) from six NHS hospitals judged
7 against the reference standard are shown in Table 1. The conjectured accuracy predictor by
8 Obuchowski³⁷ for intra-observer variability listed high accuracy to be 90% (specificity / sensitivity
9 80%). For the reporting radiographers, 4 out of 6 scored higher than 90% in accuracy (the lowest score
10 was 88.3%, mean 92.2%), for sensitivity 5 out of 6 scored over 80% (lowest score 78%, mean 88.7%),
11 and for specificity all scored over 80% (lowest score 86.7%, mean 95.6%). Comparison of the AUC
12 was calculated using MedCalc⁴¹ to obtain individual AFROC plotting (Graph 1 and 2, and Table 2), and
13 a mean AUC value of 0.903 (95% CI 0.835 to 0.948). MedCalc⁴¹ calculations to produce the AUC used
14 methodology by Metz⁴², Griner et al⁴³ and Zweig and Campbell⁴⁴ which advised would give increased
15 power and sensitivity to the results from this method than from using traditional t-test comparison
16 calculations.
17

18
19 Further calculations using MedCalc⁴¹ which applied DeLong, DeLong and Clarke-Pearson⁴⁵, Hanley
20 and Haijian-Tilaki⁴⁶ and Hanley and McNeil^{47, 48} sampling comparison methodology produced a
21 reporting radiographer mean standard error (SE) analysis of 0.020033.
22
23

24
25 The results for the consultant radiologists ($n=2$, Ranked CR1-CR2) judged against the reference
26 standard are shown in Table 3 and 4. The consultant radiologists for sensitivity scored 80% and
27 86.7% respectively, for specificity all scored over 80% (86.7%, and 93.3%), accuracy was judged to be
28 83.3% and 90%. Comparison of the AUC was calculated using MedCalc⁴¹ to obtain individual AFROC
29 plotting (Graphs 3 and 4, and Table 3 and 4), and a mean AUC value of 0.888 (95% CI 0.817 to 0.936)
30 and a SE of 0.026. A test of the comparison between the RR and CR AUC and SE, resulted in $p=0.9408$
31 and SE 0.202, inferring that the AUC was not statistically different between the cohorts.
32
33

34 35 **Discussion**

36
37 A common issue with conventional ROC scoring of participants raw data is the potential for
38 degenerative data. Metz²⁶ discussed controversies of converting raw data into ROC curve plotting;
39 where the data scale is too discrete and implies it contains degenerate data to produce
40 inappropriate ROC curve shapes and AUC calculations. In response to these practical issues Metz²⁶
41 advised to use AFROC plotting to obtain AUC scores for valid statistical significance in MRMC studies
42 of reader variation. This decision process requires the participant's results to be scored against the
43 amount of lesions and locations present in the images banks. In this study the participant's case
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 bank of 30 CT head examples contained a possible 115 scores of LL or NL with associated location
2 and confidence scores to give an accurate description of the participant's diagnostic threshold.
3

4 Obuchowski³³ and Chakraborty³⁷ recommend ROC curves and AUC as a global measure of accuracy
5 and performance. In pathology interpretation where false negative scores could have significant
6 complications, a high sensitivity (TP rate) and specificity (TN rate) is recommended. Obuchowski³⁴
7 advises the use of sensitivity at a FP score equal to or less than 0.10 (specificity >0.90). This high level
8 of sensitivity and specificity in ROC studies has been set to a standard that reflects the seriousness of
9 the interpretation of pathology on patient outcomes and treatments (avoidance of surgery or other
10 diagnostic tests, hospital stay, or abandonment of clinical treatment). Fineburg et al⁴⁹, Fryback and
11 Thornbury⁵⁰ and Brealey¹⁹ emphasize the interpretation of imaging in the chain of clinical efficacy
12 must set high standards to reduce the risk of error and harmful patient outcomes.
13
14
15
16
17
18
19
20

21 Six electronic databases (Cochrane, Medline, Europe Pubmed Central, CINAHL, ScienceDirect and
22 Google Scholar) were searched to find comparative CT head interpretation studies. The literature
23 search located 45 papers; only one non-peer review journal paper displayed the results of a
24 reporting radiographer's CT head interpretation study⁵¹. The paper did not provide sufficient details
25 as to the methodology, data, sample size or statistical analysis used, although the limited results
26 displayed a high sensitivity and specificity.
27
28
29
30

31 The review of literature evaluating consultant radiologist's interpretation of CT head scans allowed
32 analysis of the summary estimates to calculate a broad estimation of the combined results. The most
33 statistically detailed study found was Erly et al⁵² who studied 15 consultant radiologists reviewing
34 716 CT head scans (649 were normal). The results produced an agreement level of 95%, sensitivity
35 85.7%, specificity 99.7% and accuracy of 99.4%.
36
37
38
39
40

41 Further published studies found limited statistical details on diagnostic thresholds for consultant
42 radiologist's interpretation CT head examinations. Nagaraja et al⁵³, studied 6 consultant radiologists
43 reviewing 270 paediatric CT head examinations of subtle fractures and congenital abnormalities,
44 found 84.1% agreement and 15.9% disagreement. Le et al⁵⁴ on the findings of 10 consultant
45 radiologists reviewing 1,736 cases of which 48 were reported as discordant, gave a concordance rate
46 of 97.2%. A similar study by Briggs et al²¹ produced a 66% agreement and 44% discordance rate.
47 McCarron et al²² studied 9 consultant radiologists reviewing 77 CT head examinations, obtained an
48 agreement of 86.6%. Schriger et al⁵⁵ used a multiple site, multiple case methodology of 36
49 consultant radiologists reviewing 56 CT scans established an accuracy of 83%.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

When considering the accuracy of interpretation, Obuchowski³⁴ suggests high accuracy to be 90% (specificity / sensitivity 80%). The literature search and analysis provided a reasonable estimation of consultant radiologists from the published literature reviewed studies. The averaged estimated consultant radiologist reference standard was 83% accuracy, and 85.5% agreement (95% CI 73.0 to 97.0%, $p < 0.271$) from results by Schriger et al⁵⁵; Ery et al⁵²; Le et al⁵⁴, Briggs et al²¹; Nagaraja et al⁵³ and McCarron et al²². The literature showed that the majority of consultant radiologist study results had not supplied sufficient data to accurately calculate a pooled sensitivity or specificity for consultant radiologists. In comparison from the limited small sample of observers which is not generalizable to the greater population, our preliminary study found reporting radiographers mean accuracy to be 92.2%, and from the small sample of consultant radiologists 86.6%, which is above the mean of the published literature.

Conclusion

The overall aim of this limited scale preliminary research was to achieve an understanding of the degree of image interpretation accuracy of a small sample of CT head reporting radiographers and consultant radiologists in a clinical environment. Particularly, the relationship between the calibre of results (intra observer analysis) and in comparison (inter observer analysis) to each other and the published consultant radiologists diagnostic threshold.

The study findings suggested that a small sample of reporting radiographers displayed a high level of accuracy in the interpretation of CT head examinations, which was equivalent to a small sample of consultant radiologists, and were consistent with the published findings of other studies in this field. It is recommended further funded research needs to be undertaken to establish the degree of accuracy of a larger sample of participants. Further research would also encourage debate on role extension of radiographers reporting, and foster discussion on improving and modernising the workforce roles for future service delivery within this modality.

References

1. United Kingdom Acquired Brain Injury Forum. Life after Brain Injury: A Way Forward- Evidence Base; 2012. Available at: http://www.ukabif.org.uk/uploads/UKABIF/Life_After_Brain_Injury.pdf
2. Department of Health. The NHS Plan: A plan for investment, A plan for reform. London: HMSO; 2000.
3. Department of Health. Radiography skills mix: a report on the four-tier service delivery model. London. London: HMSO; 2003.
4. Department of Health. Equity and excellence: Liberating the NHS. London: HMSO; 2010.
5. Department of Health. Improving outcomes: A Strategy for Cancer. London: HMSO; 2011.
6. Department of Health. Referral to treatment: consultant-led waiting times. London: HMSO; 2014.
7. Department of Health. The Year: NHS Chief Executive's Annual Report for 2008-09. London: HMSO; 2009.
8. The King's Fund. How is the health and social care system performing?. London, The King's Fund; 2013.
9. National Diagnostics Imaging Board. Delivering the 18 week patient pathway. London: Radiology reporting times guidance; 2008.
10. National Institute for Health and Care Excellence. Stoke: Diagnosis and initial management of acute stroke and transient ischaemic attack (TIA). CG68. London: National Institute for Health and Care Excellence; 2008.
11. National Institute for Health and Care Excellence. Head injury: Triage, assessment, investigation and early management of head injury in children, young people and adults. CG176. London: National Institute for Health and Care Excellence; 2014.
12. The Royal College of Radiologists. Clinical Radiology UK Workforce Report 2011. London: The Royal College of Radiologists; 2012.
13. The Royal College of Radiologists and the Society and College of Radiographers. Team working in clinical imaging. London: The Royal College of Radiologists and the Society and College of Radiographers; 2012.
14. Department of Health. Radiography skills mix: a report on the four-tier service delivery model. London. London: HMSO; 2003.
15. The Society of Radiographers. Preliminary Clinical Evaluation and Clinical Reporting by Radiographers: Policy and Practice Guidance, London, The College of Radiographers; 2013.
16. The Society and College of Radiographers. Scope of radiographic practice survey 2012, London, The Society and College of Radiographers; 2012.
17. The Society and College of Radiographers. Diagnostic Radiography UK Workforce Report 2014, London, The Society and College of Radiographers; 2014.
18. Chang PJ. Bayesian Analysis Revisited: A Radiologists Guide. American Journal of Roentgenology 1998, 152: 721-727.
19. Brealey S. Measuring the effects of image interpretation: an evaluative framework. Clinical Radiology 2001, 56(5): 341-347.

- 1 20. Robinson PJ, Wilson D, Coral A, Murphy A, Verow P. Variation between experienced observers in
2 the interpretation of accident and emergency radiographs. *The British Journal of Radiology* 1999,
3 72(856): 323-330.
- 4 21. Briggs GM, Flynn PA, Worthington M, Rennie I, McKinstry CS. The role of specialist
5 neuroradiology second opinion reporting: is there added value?. *Clinical Radiology* 2008, 63: 791-
6 795.
- 7
8 22. McCarron MO, Sands C, McCarron P. Neuroimaging reports in a general hospital: Results from a
9 quality-improvement program. *Clinical Neurology and Neurosurgery* 2010, 112(1): 54-58.
- 10 23. Erly WK, Berger WG, Krupinski E, Seegar JF, Guisto JA. Radiology resident evaluation of head CT
11 scan orders in the emergency department. *American Journal of Neuroradiology* 2002, 23: 103–107.
- 12 24. Strub WM, Vagal AA, Tomsick T, Moulton JS. Overnight resident preliminary interpretations on
13 CT examinations: should the process continue?. *Emergency Radiology* 2006, 13(1): 19-23.
- 14 25. Gallagher FA, Tay KY, Vowler SL, Szutowicz H, Cross JJ, McAuley DJ, Antoun NM. Comparing the
15 accuracy of initial head CT reporting by radiologists, radiology trainees, neuroradiographers and
16 emergency doctors. *British Journal of Radiology* 2011, 84(1007): 1040-1045.
- 17 26. Metz CE. Practical Aspects of CAD Research: Assessment Methodologies for CAD', Proceedings of
18 the 45th Annual Meeting of American Association of Physicists in Medicine, San Diego, America;
19 2003.
- 20 27. Brealey S, Scally AJ. Methodological approaches to evaluating the practice of radiographers'
21 interpretation of images, A Review. *Radiography* 2008, 14 (1): 46-54.
- 22 28. Thompson JD, Manning DJ, Hogg P. The value of observer performance studies in dose
23 optimization: A focus on free response receiver operating characteristic methods. *The Journal of*
24 *Nuclear Medicine Technology* 2013, 41: 57-64.
- 25 29. Piper K, Ryan C, Paterson A. The Implementation of a Radiographic Reporting Service, for trauma
26 examinations of the skeletal system, in 4 National Health Service Trusts. Project Report SPGS 438.
27 South Thames Regional Office (NHSE); 1999.
- 28 30. IMAGE Smith, Inc. (VeriLUM (Version 5.2.1); 2008. Available at:
29 <http://verilum.software.informer.com/>
- 30 31. The Royal College of Radiologists. Picture archiving and communication systems (PACS) and
31 guidelines on diagnostic display devices, second edition. London, The Royal College of Radiologists;
32 2012.
- 33 32. IMAGE Information Systems. iQ-View (Version 2.8.0); 2013. Available at: [http://www.k-](http://www.k-pacs.net/10555.html)
34 [pacs.net/10555.html](http://www.k-pacs.net/10555.html)
- 35 33. Obuchowski NA. How many observers are needed in clinical studies of medical imaging?.
36 *American Journal of Roentgenology* 2004, 182 (4): 867-869.
- 37 34. Obuchowski NA. Sample size tables for receiver operating characteristic studies. *American*
38 *Journal of Roentgenology* 2000, 175(3): 603-608.
- 39 35. Piper K, Buscall K, Thomas, N. MRI reporting by radiographers: Findings of an accredited
40 postgraduate programme. *Radiography* 2010; 16: 136-142.
- 41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
36. Thompson JD, Manning DJ, Hogg P. Analysing data from observer studies in medical imaging research: An introductory guide to free-response techniques. *Radiography* 2014, 20 (4): 295-299.
 37. Chakraborty DP. Statistical power in observer-performance studies: comparison of the receiver operating characteristic and free-response methods in tasks involving localization. *Academic Radiology* 2002, 9(2): 147-156.
 38. Data Protection Act. Chapter 29. London: HMSO; 1998.
 39. Cosson P, Willis N. Digital teaching library (DTL) development for radiography education. *Radiography* 2011, 18(2): 112-116.
 40. General Medical Council. Supplementary Guidance: making and using visual and audio recordings of patients. London. GMC; 2013.
 41. MedCalc Statistical Software version 14.8.1 (MedCalc Software bvba, Ostend, Belgium); 2014. Available at: <http://www.medcalc.org/>
 42. Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* 1978, 8: 283-298.
 43. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. *Annals of Internal Medicine* 1981, 94: 555-600.
 44. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 1993, 39: 561-577.
 45. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988, 44: 837-845.
 46. Hanley JA, Hajian-Tilaki KO. Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: an update. *Academic Radiology* 1997, 4: 49-58.
 47. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982, 143: 29-36.
 48. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983, 148(3): 839-843.
 49. Fineburg H, Bauman R, Sosman M. Computerized cranial tomography: Effect on Diagnostic and Therapeutic Plans. *Journal of the American Medical Association* 1977, 238(3): 224-227.
 50. Fryback DG, Thornbury JR. The efficiency of diagnostic imaging. *Medical Decision Making* 1991, 11: 88-94.
 51. Craven CM. Radiographer reporting: CT head scans. *Synergy* 2003, November: 15-19.
 52. Ery WK, Ashdown BC, Lucio RW, Seegar JF, Alcalá JN. Elevation of Emergency CT scans of the head: is this a community standard?. *American Journal of Radiology* 2003, 180, June: 1727-1730.
 53. Nagaraja S, Ullah Q, Lee KJ, Bickle I, Hon LQ, Griffiths PD, Raghavan A, Flynn P, Connolly DJA. Discrepancy in reporting among specialist registrars and the role of a paediatric neuroradiologist in reporting paediatric CT head examinations. *Clinical Radiology* 2009, 64(9): 891-896.
 54. Le AH, Licurse A, Catanzano TM. 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.

55. Schriger DL, Kalafut M, Starkman S, Krueger M, Saver JL. Cranial computed tomography interpretation in acute stroke. The Journal of the American Medical Association 1998, 279(16): 1293-1297.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Acknowledgements

The authors 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.

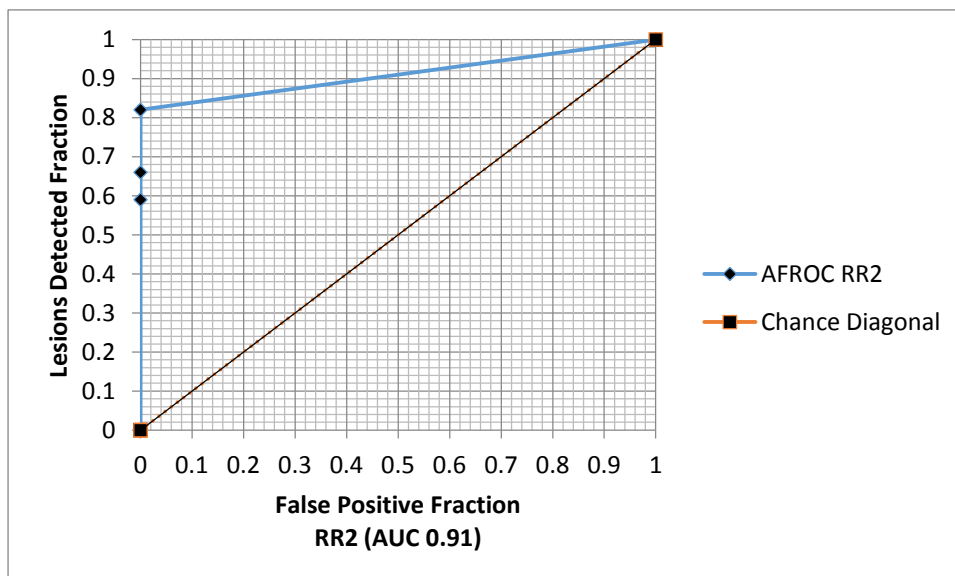
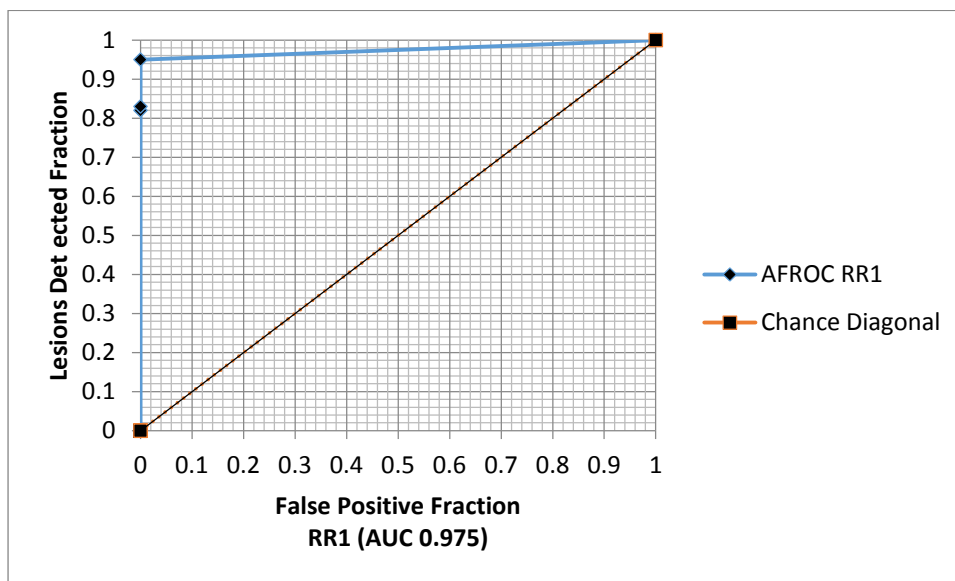
Conflict of interest statement

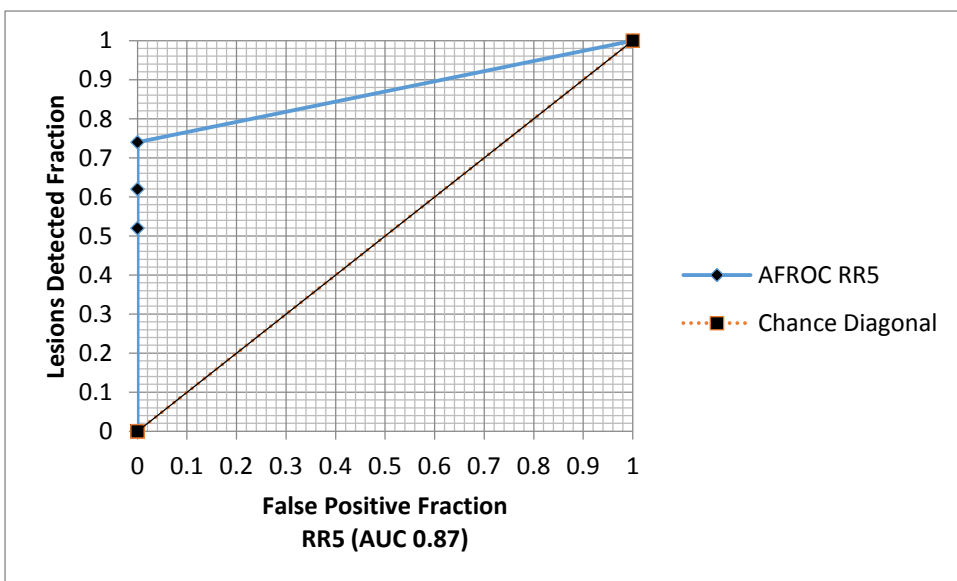
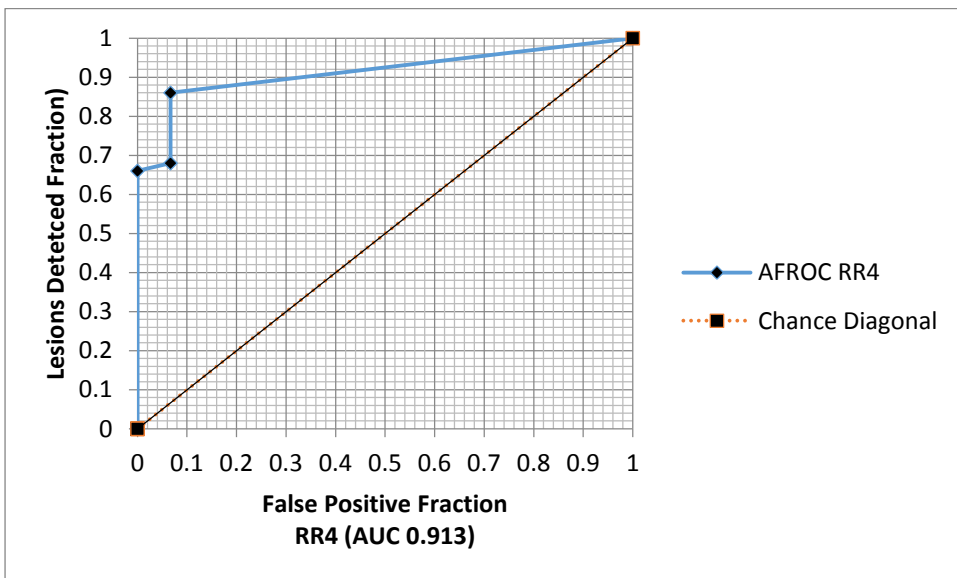
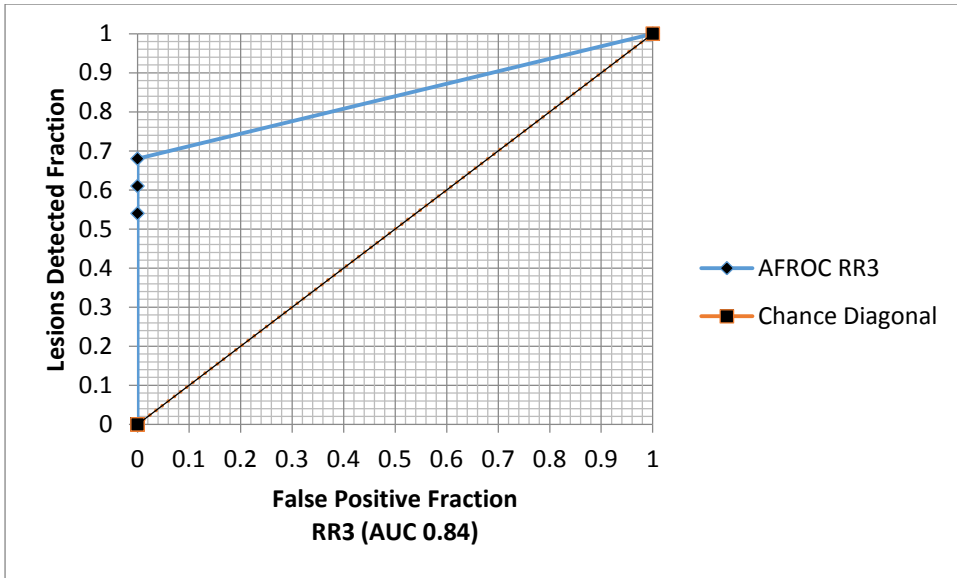
The author *a* is a programme leader and author *b* programme director for the postgraduate CT head reporting course at CCCU.

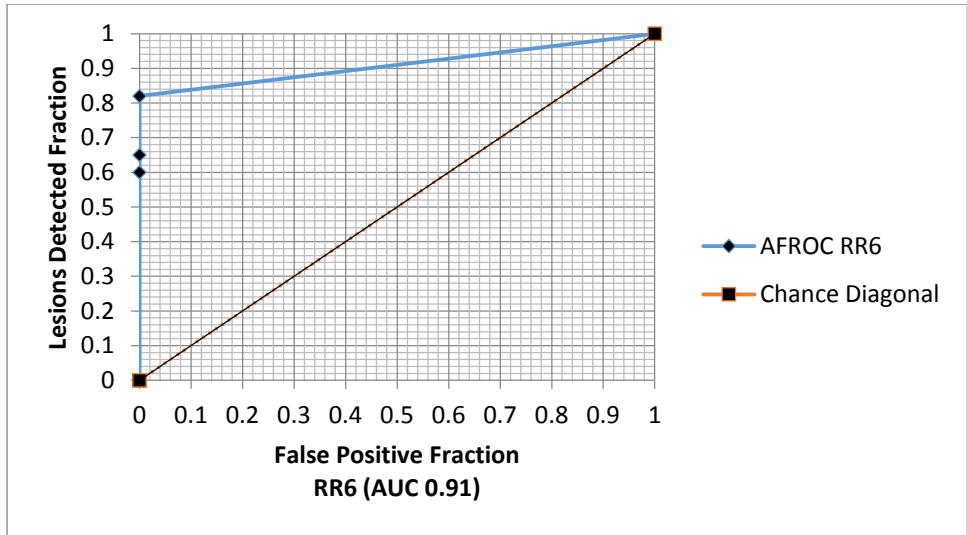
Tables

Participant	Sensitivity	Specificity	Accuracy	FP	FN	TP	TN
RR1	96.6	91.8	94.1	1.25	1.25	14.25	14
RR2	88.1	98.4	93.3	0.25	1.75	13	15
RR3	78	98.4	88.3	0.25	3.25	11.5	15
RR4	91.7	86.7	89.2	2	1.25	13.75	13
RR5	90	100	95	0	1.5	13.5	15
RR6	88.1	98.4	93.3	0.25	1.75	13	15
Mean	88.75	95.61	92.2	0.66	1.79	13.66	14.5

Table 1 Reporting radiographers results compared to the study reference standard



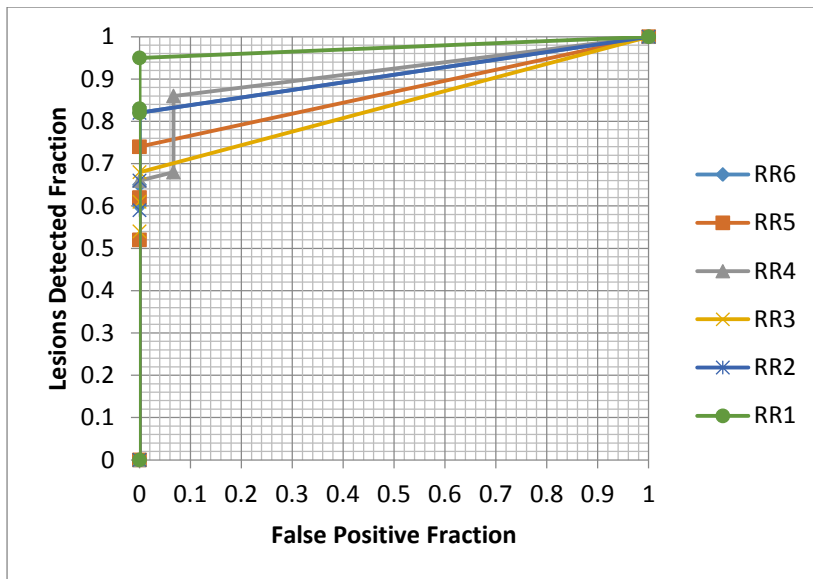




Graph 1 Reporting radiographer AFROC and AUC results.

	AUROC	SE	95% CI
RR1	0.975	0.0110	0.927 to 0.995
RR2	0.910	0.0193	0.842 to 0.955
RR3	0.840	0.0234	0.760 to 0.902
RR4	0.913	0.0252	0.845 to 0.902
RR5	0.870	0.0220	0.795 to 0.925
RR6	0.910	0.0193	0.842 to 0.955
Mean	0.975	0.0365	0.835 to 0.939

Table 2 Reporting radiographer AFROC results



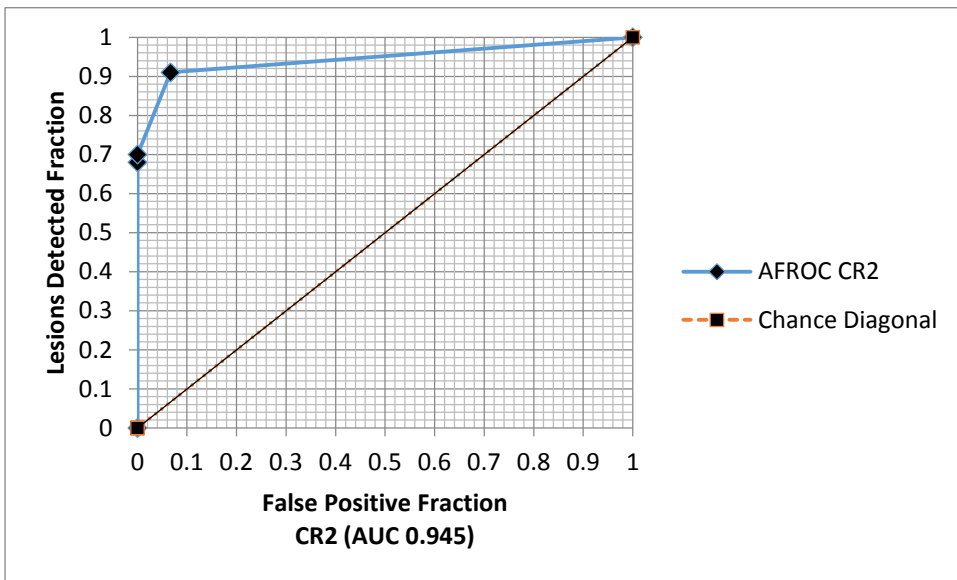
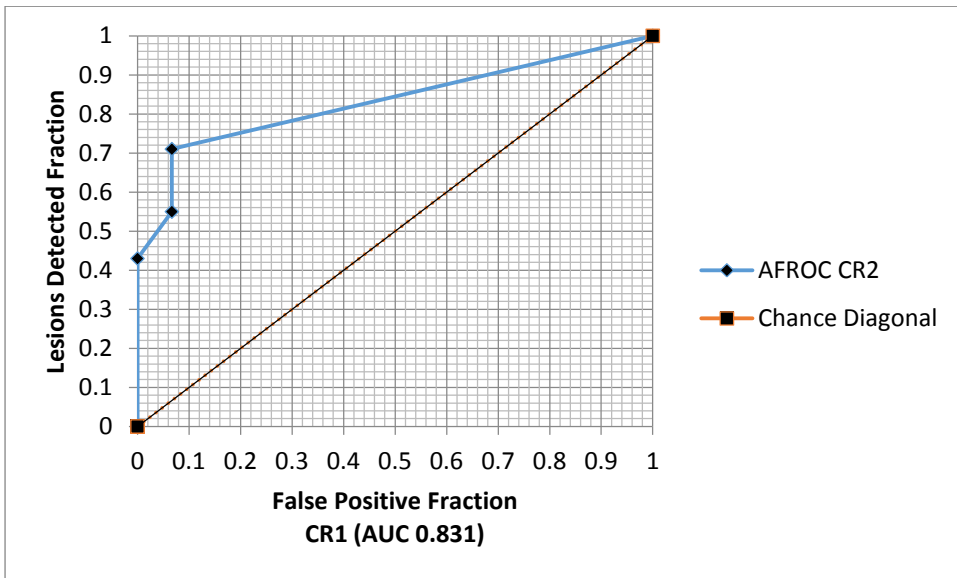
Graph 2 Comparison of reporting radiographer AFROC results.

Participant	Sensitivity	Specificity	Accuracy	FP	FN	TP	TN
CR1	80	86.7	83.3	2	3	12	13
CR2	86.7	93.3	90	1	2	13	14
Mean	83.35	90	86.65	1.5	2.5	12.5	13.5

Table 3 Consultant radiologist results compared to the study reference standard

	AUROC	SE	95% CI
CR1	0.831	0.0338	0.749 to 0.864
CR2	0.945	0.0182	0.886 to 0.979
Mean	0.888	0.026	0.817 to 0.921

Table 4 Consultant radiologist AFROC results



Graph 3 Consultant radiologist AFROC results.



Graph 4 Comparison of consultant radiologist AFROC results.