BJR

Received: 5 March 2015 | Revised: 9 April 2015 | Accepted: 15 April 2015

doi: 10.1259/bjr.20150191

Cite this article as:
Caffery LJ, Armfield NR, Smith AC. Radiological interpretation of images displayed on tablet computers: a systematic review. Br J Radiol 2015; 88:20150191.

FULL PAPER

Radiological interpretation of images displayed on tablet computers: a systematic review

L J CAFFERY, PhD, N R ARMFIELD, PhD and A C SMITH, PhD

Centre for Online Health, The University of Queensland, Brisbane, QLD, Australia

Address correspondence to: Dr Liam Caffery
E-mail: *l.caffery@uq.edu.au*

Objective: To review the published evidence and to determine if radiological diagnostic accuracy is compromised when images are displayed on a tablet computer and thereby inform practice on using tablet computers for radiological interpretation by on-call radiologists.
Methods: We searched the PubMed and EMBASE databases for studies on the diagnostic accuracy or diagnostic reliability of images interpreted on tablet computers. Studies were screened for inclusion based on predetermined inclusion and exclusion criteria. Studies were assessed for quality and risk of bias using Quality Appraisal of Diagnostic Reliability Studies or the revised Quality Assessment of Diagnostic Accuracy Studies tool. Treatment of studies was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).
Results: 11 studies met the inclusion criteria. 10 of these studies tested the Apple iPad® (Apple, Cupertino, CA). The included studies reported high sensitivity (84–98%), specificity (74–100%) and accuracy rates (98–100%) for radiological diagnosis. There was no statistically significant difference in accuracy between a tablet computer and a digital imaging and communication in medicine-calibrated control display. There was a near complete consensus from authors on the non-inferiority of diagnostic accuracy of images displayed on a tablet computer. All of the included studies were judged to be at risk of bias.
Conclusion: Our findings suggest that the diagnostic accuracy of radiological interpretation is not compromised by using a tablet computer. This result is only relevant to the Apple iPad and to the modalities of CT, MRI and plain radiography.
Advances in knowledge: The iPad may be appropriate for an on-call radiologist to use for radiological interpretation.

Consumer tablet computers can be used to access and display digital radiographic images for the purpose of radiological interpretation. Because tablet computers are portable, they have a potential role in remote, emergency diagnostic radiology services. There has been limited acceptance of smartphones for radiological interpretation because of their small screen size and limited display resolution.[1] Tablet computers offer similar portability to a smartphone but with high-resolution displays and a larger viewing size.[2] Hence, a tablet computer may be a more suitable display device for on-call radiologists.

The luminance and contrast properties of computer displays can vary considerably causing inconsistent display of images between devices. The accepted process for achieving consistent display of medical images is by calibration of the display device to the digital imaging and communication in medicine (DICOM) greyscale display function (GSDF).[3] Conformance to the GSDF has been shown to improve diagnostic accuracy.[4,5] A primary display is a dedicated medical display device and is used by radiologists for primary diagnosis. Whereas, a secondary display is often a commercial-off-the-shelf computer display. Established guidelines recommend conformance to the GSDF should be better than 10% and 20% for primary and secondary displays, respectively.[6] Whilst both primary and secondary liquid crystal display (LCD) devices can be calibrated to the GSDF, it is not possible to calibrate a tablet computer, which may potentially compromise accuracy.[7] Despite the inability to calibrate the display, high levels of diagnostic accuracy have been reported when using tablet computers.[8–10] Hence, there is contradictory information to inform practice on the use of tablet computers for radiological interpretation.

To date, there has been no attempt to synthesize the existing research evidence pertaining to diagnostic accuracy or diagnostic reliability of using tablet computers for radiological interpretation. The aim of this study was to systematically review the published literature to determine if diagnostic accuracy is compromised when images are displayed on

a tablet computer, which would in turn inform practice on the appropriateness of an on-call radiologist using a tablet computer for radiological interpretation.

METHODS AND MATERIALS

Search strategy

We searched the PubMed and EMBASE databases using a combination of keywords, Medical Subject Headings (MeSH) and Emtree terms for radiology, teleradiology and tablet computers (Table 1). The MeSH and Emtree terms for tablet computers are *handheld computer* and *microcomputer*, respectively. The results were constrained to the articles published in the past 10 years. Searches were conducted in January 2015.

Inclusion and exclusion criteria

We included studies published in peer-reviewed journals that examined either the diagnostic accuracy or the diagnostic reliability of radiological interpretation of images displayed on a tablet computer. Diagnostic reliability refers to the agreement between two or more observations of the same entity and is often reported as interrater or intrarater reliability.[11] Whereas, diagnostic accuracy is the likelihood of the interpretation being correct when compared with an independent standard.[12] For the purpose of this review, we defined a tablet computer as a hand-held or portable computer with a screen size of 7-inches or more. This criterion excluded studies of images displayed on smartphones, personal digital assistants and the Apple iPod® (Apple, Cupertino, CA). The modalities of diagnostic radiology, namely, plain film radiography, CT, ultrasonography, nuclear imaging or MRI were included. Dental imaging was excluded as it was not considered likely to be reported by an on-call radiologist. Studies that tested imaging that was performed on patients were included. Studies where the imaging was on phantoms or synthesized were excluded. Studies that were reported in languages other than English, conference proceedings, commentary and letters to the editor were also excluded.

Selection process

Two reviewers screened the title and abstract of studies to determine eligibility for inclusion. Screening the full text of articles was performed if the abstract did not provide sufficient information to judge eligibility. Uncertainty of inclusion was resolved by consensus discussion.

Data extraction and quality assessment

The full text of studies that met the inclusion criteria was obtained and data extracted. Data were extracted on study characteristics (year, country where the study was conducted, methodology, case metrics, reader selection and reporting instrument), outcome measures, technology (intervention display device and reference standard display device), results summary and secondary observations on the use of tablet computers for radiological diagnosis.

To evaluate the included studies for quality and risk of bias, we used two methods. The Quality Appraisal of Reliability Studies (QAREL)[11] was used to assess diagnostic reliability studies, and the revised Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2)[13] was used to assess diagnostic accuracy studies. Two tools were necessary because studies of diagnostic reliability contain unique design features that are not represented on tools that assess the quality of studies investigating diagnostic accuracy.[14] The QUADAS-2 tool must be tailored to each review by adding or removing signalling questions.[13] Table 2 lists the signalling questions that have been added or removed during tailoring of the QUADAS-2 tool to our review.

One reviewer independently performed data extraction, and quality and risk of bias assessment. A second reviewer validated the recorded information.

Analysis

For multireader, multicase (MRMC) receiver operating characteristic (ROC) studies, we calculated the 95% confidence interval (when not reported by the study's author) to aid comparability of area under the ROC curve measures. Similarly, the 95% confidence interval for sensitivity and specificity was calculated. Synthesis of results was performed narratively. Reporting of the findings of this review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

RESULTS

We identified 11 studies that met our inclusion criteria. The results for each stage of our search and screening processes are shown in the PRISMA flow diagram (Figure 1).

Table 1. Query syntax

Database	Syntax
PubMed	((("radiology"[MeSH Terms] OR "radiology"[All Fields]) OR ("radiography"[MeSH Terms] OR "radiography"[All Fields]) OR ("teleradiology"[MeSH Terms] OR "teleradiology"[All Fields])) AND (("computers, handheld"[MeSH Terms] OR "computers, handheld"[All Fields]) OR "handheld device"[All Fields] OR "mobile device"[All Fields] OR "tablet computer"[All Fields] OR ipad[All Fields]) AND "last 10 years"[PDat])
Embase	((("radiology"/exp OR radiology) OR ("teleradiology"/exp OR teleradiology)) AND ("microcomputer"/exp OR "microcomputer") AND [2005–2015]/py)

Table 2. Review-specific modifications to the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool

Signalling questions	Modification to QUADAS-2 tool[a]
Domain 1: case selection	
Was a consecutive or random sample of cases selected?	NC
Was a case–control design avoided?	NC
Did the study avoid inappropriate exclusions?	NC
Was a spectrum of disease severity included in the case selection?	+
Was a sample size of 50 or more cases used in the study?	+
Domain 2: index test	
Were the index test results interpreted without knowledge of the results of reference test?	NC
If a threshold was used, was it pre-specified	−
Was an entire rating scale rather than a binary scale used to record index text diagnosis?	+
Was the instrument used to record their diagnosis calibrated with the instrument used by the gold standard readers?	+
Was the study reader selection for the index test representative of a radiologist population (in terms of number of readers and range of experience)?	+
Was the study reader for the index test given the same referral information and previous imaging as the reference test reader?	+
Was the index test diagnosis not limited to a type(s) of pathology?	+
Was the index test read with comparable monitor luminance and ambient lighting to the reference test?	+
Was the case order randomized?	+
Were all index test outcomes reported?	+
Domain 3: reference test	
Is the reference standard likely to correctly classify the target condition?	−
Were the reference standard results interpreted without the knowledge of the results of the index test?	NC
Did all cases have a reference test?	NC
Was the gold standard diagnosis validated—for example, by consensus of multiple radiologists or review of clinical notes?	+
Domain 4: flow and timing	
Was there an appropriate interval between index test and reference standard?	−
Was there an appropriate interval between index test reading and reference test reading to address retained information?	+
Did all cases receive a reference standard?	NC
Was the same reference test used for all cases?	NC
Were all cases included in the analysis?	NC

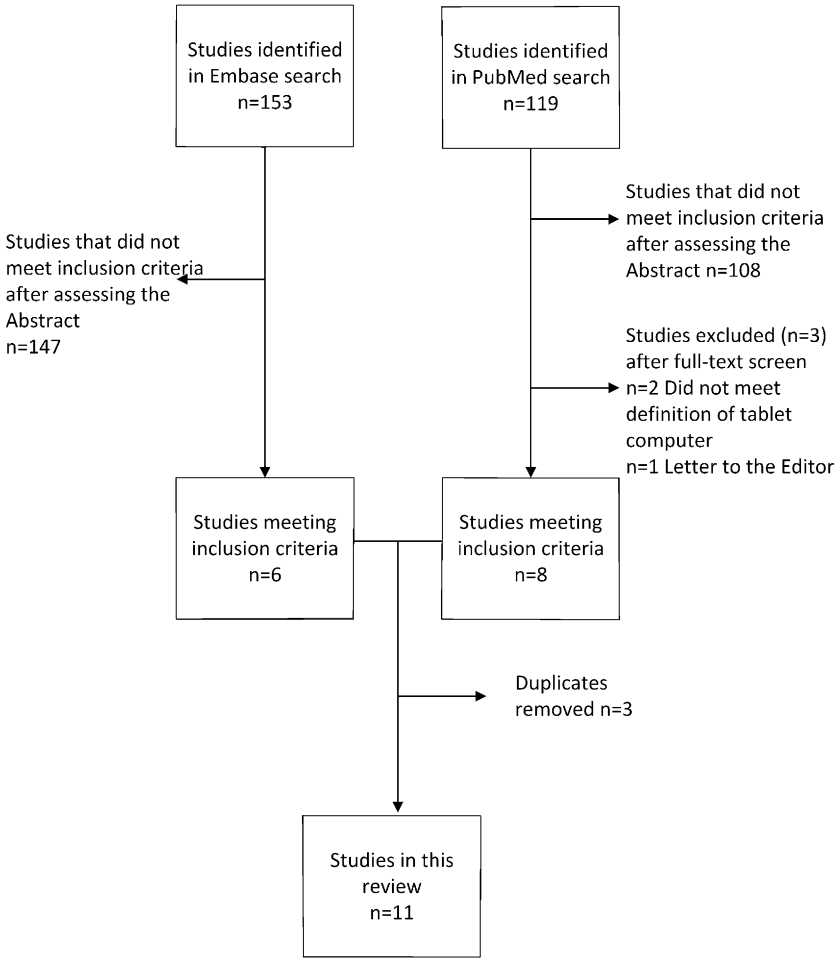
[a]+, signalling question added; −, signalling question removed; NC, no change from default tool.

Study characteristics

The Apple iPad (Apple) was the intervention display in 10 studies. These studies were published between 2011 and 2013. One study evaluated an iPad with a retina display (screen matrix of 2048 × 1536 pixels).[15] The resolution of the iPads in the other nine studies was 1024 × 768 pixels. The only other tablet computer tested was a Hewlett-Packard® TC1000 (Hewlett-Packard, Palo Alto, CA) tablet with a resolution of 1024 × 768 pixels.[16] This study was published in 2005.

Studies originated from eight countries (United States, Ireland, Germany, Singapore, India, Taiwan, Republic of Korea and

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Japan). Two of the included studies evaluated large matrix plain film radiographs while the remaining studies evaluated small matrix (CT and MRI) images (Table 3). Radiologists were the readers in all but one study. Emergency department physicians were used in the remaining study (Table 4).[8] Half of the studies compared the interpretation of images displayed on a tablet computer with the interpretation of images displayed on a primary picture archiving and communication system display. The remaining studies used a secondary LCD as the reference standard display (Table 5). The ambient lighting was controlled in six of the studies,[9,10,15,17,19] whereas other authors intentionally used conventional lighting conditions to imitate conditions under which the tablet computer would be used.[2,8] The remaining studies did not state lighting conditions.

Diagnostic accuracy and diagnostic reliability

Eight of the included studies were diagnostic accuracy studies and three were diagnostic reliability studies. Different methodologies were used to test diagnostic accuracy. The diagnostic accuracy was assessed using MRMC ROC curve in four of the included studies.[8,9,16,19] In these studies, the diagnosis from images displayed on a tablet computer was compared with the gold standard diagnosis (Table 4). The difference in area under the binormal ROC curve (AUROC) was used to test significance (Figure 2). No difference of statistical significance was found in five of the six studies (Table 6). Yoshimura et al[19] did find the AUROC was significantly smaller for an iPad than for a gamma 2.2-calibrated LCD. The same author found no statistical difference between an iPad and DICOM GSDF-calibrated LCD. Lee et al[16] found that the tablet computer performed better than the control display when assessing abdominal radiographs for ureteral calculi, whereas the control display (a GSDF-calibrated cathode ray tube monitor) was superior to the tablet for diagnosis of the renal calculi.

The included studies reported high sensitivity and specificity for tablet computers with values ranging from 84% to 98% and from 74% to 100%, respectively (Figure 3).[8–10,18] Two of the studies that measured sensitivity and specificity also performed significance testing. No significant difference in sensitivity and specificity between a tablet computer and control display was found.[8,18] Johnson et al[18] reported the same accuracy rate (98%) for interpretation of CT scans for pulmonary emboli performed on an iPad and on a primary display. Panughpath et al[10] reported an accuracy rate of 99.86% and 99.92% for the iPad and a secondary display, respectively, for the detection of intracranial haemorrhage on CT. In the studies performed by John et al[2] and McLaughlin et al,[1] the study readers produced

Table 3. Case selection summary

Study	Examination	Pathology	Case selection rationale	Case sample size	Pathology positive (%)	Pathology negative (%)
Abboud et al[17]	CXR	TB	Random selection of CXR from a pool of 500 TB screening cases	240	40 (17%)	200 (83%)
McNulty et al[9]	MRI lumbar spine; MRI cervical spine	Four possible pathologies—spinal cord compression, spinal cord oedema, cauda equina syndrome, spinal cord haemorrhage	Arbitrarily selected cases from actual emergency MRI for spinal trauma. Selection designed to include pathology from emergency presentations plus normal control cases	31	13 (42%)	18 (58%)
Tewes et al[15]	CT head; CTPA	Cerebral infarction; segmental or subsegmental PE	Arbitrarily selected cases designed to emulate typical ED cases and pathologies. Cases were actual cases performed as out-of-hours emergency imaging and included both positive and negative cases	40 CT head; 40 CTPA	20 (50%) CT head; 20 (50%) CTPA	20 (50%) CT head; 20 (50%) CTPA
Panughpath et al[10]	CT head	ICH	Random selection from an emergency radiology imaging database	100	27 (27%)	73 (73%)
Johnson et al[18]	CTPA	Pulmonary embolism	Existing set of 50 cases of imaging for suspected PE originally compiled for QA program. The selection included both positive and negative. Positive cases ranged in subtleness of pathology from easy (main pulmonary artery) to difficult (subsegmental thrombi)	50	25 (50%)	25 (50%)
Yoshimura et al[19]	CT head	Cerebral infraction	Arbitrarily selected cases after searching reporting database and electronic medical record for	97	47 (48%)	50 (52%)

(Continued)

Table 3. (Continued)

Study	Examination	Pathology	Case selection rationale	Case sample size	Pathology positive (%)	Pathology negative (%)
			cases of suspected cerebral infarction. Selection included both positive and negative cases			
Park et al[8]	CT head	Intracranial haemorrhage	Arbitrarily selected cases from actual CT head performed in ED for trauma or headache. Cases had subtle radiological signs of ICH. Subtle meant 1st or 2nd year ED resident had missed the ICH. 10 cases were paediatric to reflect real practice of emergency radiology	100	50 (50%)	50 (50%)
John et al[2]	CT and MRI	Common ED pathologies	Arbitrarily selected cases of common after-hours pathology that had been clinically interpreted after hours by one particular senior radiologist. The interpretation of this radiologist was used as the gold standard diagnosis	88 (79 CT, 9 MRI)	64 (73%)	24 (27%)
McLaughlin et al[1]	CT head	Various	100 consecutive CT brain studies referred from the ED	100	57 (57%)	43 (43%)
Bhatia et al[20]	CTBA; MR spine cervical; MR spine thoracic; MRI spine lumbar; MRI brain	Acute ischaemic event	Arbitrarily selected cases from patients that undergone ED imaging for an acute central nervous system event	50 CTBA; 50 MRI brain; 50 MRI spine	26 (52%) CTBA; not stated MRI brain; not stated MRI spine	24 (48%) CTBA; not stated MRI brain; not stated MRI spine
Lee et al[16]	AXR	Urolithiasis	Consecutive cases for patients referred for intravenous urography	160 renal systems (80 AXR)	28 (18%) renal stone; 24 (15%) ureteric stone	132 (82%) kidney; 136 (85%) ureters

AXR, abdominal radiograph; CTBA, CT brain angiography; CTPA, CT pulmonary angiogram; CXR, chest radiograph; ED, emergency department; ICH, intracranial haemorrhage; PE, pulmonary embolism; TB, tuberculosis; QA, quality assurance.

Table 4. Reader metrics

Study	Reference standard	Number of reference standard readers	Index test readers and profession	Index test reader attributes	Index test instrument
Abboud et al[17]	NA	NA	Five radiologists	Two chest fellowship, two fellowship trainees and one resident	Binary (positive/ negative)
McNulty et al[9]	Radiologist working in consensus	Two	13 radiologists	12 board-certified neuroradiology experts, 1 board-certified spinal and musculoskeletal expert	Six-point confidence scale
Tewes et al[15]	NA	NA	Three radiologists	Three radiologists with 3, 4 and 6 years' experience, respectively	Five-point confidence scale
Panughpath et al[10]	Discordant studies assessed by fellowship-trained neuroradiologist	Two	Two radiologists	Not reported	Binary (positive/ negative) plus type categories, *e.g.* extradural, subdural, subarachnoid, intraparenchymal or intraventribular
Johnson et al[18]	Cases reported as clinically positive were further reviewed	Three (one initial clinical radiologist plus two additional)	Two radiologists	Two fellowship trained but junior radiologists	Binary (positive/ negative)
Yoshimura et al[19]	Radiologists working in consensus plus accuracy of report confirmed by MRI and clinical records	Two	Nine radiologists	Six general radiologists and three neuroradiologists with 3–17 years' experience	Continuous confidence scale
Park et al[8]	Neuroradiologist	One	Five emergency department physicians	Three attending and two senior residents	Five-point confidence scale
John et al[2]	Clinical report of non-study senior radiologist	One	Three radiologists	Three attending with at least 10 years' experience each	Descriptive report. Non-study radiologist classified discrepant diagnosis as major or minor
McLaughlin et al[1]	Clinical report of radiologist	One	Two radiologists	Two radiologists with 5 and 16 years' experience	Descriptive report. Discrepancies classified according to American College of Radiologist's RadPeer classification system
Bhatia et al[20]	NA	NA	Five radiologists	One board-certified neuroradiologist, three fourth-year radiology residents and one second-year radiology resident	Binary (positive/ negative) plus type categories, *e.g.* disc herniation and/or reason category, *e.g.* gradient echo signal abnormality
Lee et al[16]	Results of IVU study (as opposed to plain AXR) and clinical records	One	Two radiologists	Not reported	Five-point confidence scale

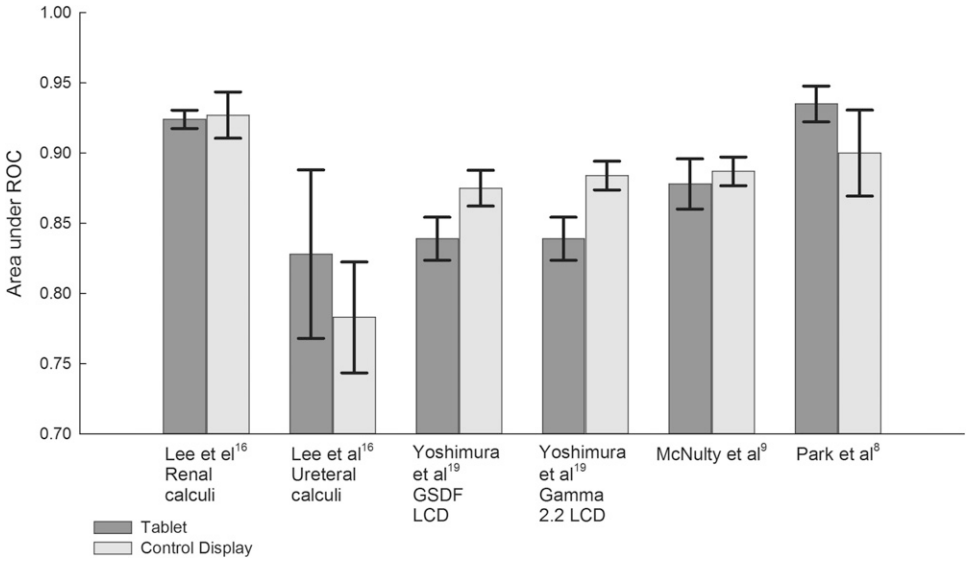
AXR, abdominal radiograph; IVU, intravenous urography; NA, not applicable.

Table 5. Results summary

Study	Reference standard display	Radiographic examination	Interrater reliability tablet (K)	Interrater reliability reference standard display (K)	Intrarater reliability (K)	Sensitivity tablet	Sensitivity reference standard display	Specificity tablet	Specificity reference standard display	AUC ROC tablet	AUC ROC reference standard display
Abboud et al[17]	Secondary LCD	Chest radiograph	0.865	0.817	0.830	NR	NR	NR	NR	NR	NR
Tewes et al[15]	Primary display	CT head	0.330	0.320	0.520	NR	NR	NR	NR	NR	NR
	Primary display	CTPA	0.690	0.600	0.670	NR	NR	NR	NR	NR	NR
Bhatia et al[20]	Secondary LCD	CT brain angiography, MRI brain and MR spine	NR	NR	>0.75	NR	NR	NR	NR	NR	NR
Park et al[8]	Secondary LCD—GSDF calibrated	CT head	NR	NR	0.597	0.850	0.850	0.790	0.800	0.935	0.900
McNulty et al[9]	Secondary LCD—GSDF calibrated	MRI spine including lumbar and cervical	NR	NR	NR	0.842	0.824	0.740	0.786	0.878	0.887
Panughpath et al[10]	Secondary LCD	CT head	NR	NR	NR	0.960	NR	1.00	NR	NR	NR
Johnson et al[18]	Primary display	CTPA	NR	NR	NR	0.980	1.00	0.980	0.960	NR	NR
Yoshimura et al[19]	Primary display—GSDF calibrated	CT head	NR	NR	NR	NR	NR	NR	NR	0.839	0.875
	Primary display—gamma calibrated	CT head	NR	NR	NR	NR	NR	NR	NR	0.839	0.884
Lee et al[16]	Primary—CRT	AXR (renal calculi)	NR	NR	NR	NR	NR	NR	NR	0.924	0.927
	Primary—CRT	AXR (ureteral calculi)	NR	NR	NR	NR	NR	NR	NR	0.828	0.783

AUC, area under the curve; AXR, abdominal radiograph; CRT, cathode ray tube; CTPA, CT pulmonary angiogram; GSDF, greyscale display function; LCD, liquid crystal display; NR, not reported; ROC, receiver operating characteristic.

Figure 2. Comparison between mean binormal area under the receiver operating characteristic (ROC) curves for tablet and control display (error bars are 95% confidence intervals). GSDF, greyscale display function; LCD, liquid crystal display.



a descriptive diagnosis that was compared with the formal clinical report. Discrepancies between the reader's diagnosis and clinical diagnosis were classified by a non-study reader. John et al[2] reported 3.4% major (finding would affect immediate clinical management) discrepancy rate and a 5.6% minor (would not affect immediate clinical management) discrepancy rate when using a tablet computer. McLaughlin et al[1] categorized discrepancies according to the American College of Radiologist's RadPeer classification system. There were 12 errors (3 clinically significant and 9 not clinically significant) when using the control display and 7 errors (3 clinically significant and 4 not clinically significant) when using a tablet computer. Interrater reliability was almost identical between index and reference standard for the studies by Abboud et al[17] and Tewes et al[15] (Table 5).

Quality and risk of bias

The eight diagnostic accuracy studies[1,2,8–10,15,18,19] were assessed with the QUADAS-2 tool for quality and risk of bias. All eight studies were judged *high* or *unclear* in at least one domain. Proportions of studies for each of the risk of bias classifications are shown in Figure 4. A large proportion of these studies (88%) were judged to have a high risk of bias for the index test domain. This was owing to a number of reasons, including the index test readers not having sufficient number or range of experience to represent a radiologist population; the index test readers not having the same referral information and access to previous imaging as the reference test readers; the index test reader's diagnosis being limited to known type of pathology; and the monitor luminance and ambient lighting of the index test not being comparable to the reference test.

All studies in the reference standard domain and most studies (88%) in both the index test and case selection domain were judged to be applicable to the review question for all QUADAS-2 domains (Figure 5).

The three diagnostic reliability studies[15,17,20] were assessed using the QAREL tool. All studies had at least one item judged to indicate a poor quality in the study. The reasons for poor quality included readers not being representative of the review's population, blinding of referral and clinical information, non-randomization of reading order and the use of binary scales by study readers.

Synthesis

The results of studies included in this review could not be statistically combined. This was owing to the heterogeneity of study designs (accuracy and reliability studies), methodologies, display characteristics of intervention tablet computer and reference standard display device, profession of readers, lighting conditions and radiographic modality (large matrix and small matrix) all being evaluated.

DISCUSSION

This review revealed a near complete consensus from the study authors on the non-inferiority of diagnostic accuracy of images displayed on a tablet computer. The included studies reported high sensitivity (84–98%), specificity (74–100%) and accuracy rates (98–100%) when using a tablet computer for radiological diagnosis. There was no statistically significant difference in accuracy between tablet computers and GSDF-calibrated control displays. All of the included studies were judged to be at risk of bias. The included studies were judged to have high applicability to the review question.

The MRMC ROC method has been used in four of the included studies.[8,9,16,19] All authors of MRMC ROC studies have used an entire rating scale rather than a binary scale that conforms with best practice recommendations.[21] A number of studies have validated the gold standard diagnosis by using multiple readers or reviewing clinical notes (Table 4). The use of multiple readers has been shown to increase the reliability of radiological

Table 6. Significance test summary

Study	Significance test	Significance level	*p*-value	Author's conclusions
McNulty et al[9]	DBM MRMC difference in mean AUC	5%	p(random readers and case) = 0.6696; p(fixed readers random case) = 0.5961; p(random readers fixed cases) = 0.6696	NSD diagnostic accuracy between tablet and secondary display
Tewes et al[15]	Wilcoxon (U) rank-sum test for Likert scale evaluations; t-test (t_1) for difference in mean correlation coefficient and t-test (t_2) for difference in mean kappa score	5%	$p(U) > 0.05$; $p(t_1) > 0.05$; $p(t_2) > 0.05$	NSD between the tablet and primary display for both CT head and CTPA
Panughpath et al[10]	Fisher's exact test	NR	$p(F) < 1.00$	NSD between the table and secondary display for the detection of intracranial haemorrhage
Johnson et al[18]	Difference in se, sp and ac	NR	p(se) = 1.0; p(sp) = 1.0; p(ac) = 1.0	NSD in sensitivity, specificity and accuracy between a tablet and primary display
Yoshimura et al[19]	DBM MRMC difference in mean AUC (tablet *vs* GSDF-calibrated primary display); DBM MRMC difference in mean AUC (tablet *vs* gamma-calibrated primary display; ANOVA	5%	p(tablet *vs* GSDF-calibrated primary display) > 0.05; p(tablet *vs* gamma-calibrated primary display) < 0.05; p(ANOVA) = 0.06	NSD between tablet and GSDF primary display; AUC was statistically smaller for the tablet when compared with gamma-calibrated primary display; ANOVA showed NSD between all three displays
Park et al[8]	McNemar's test for difference in se and sp DBM MRMC difference mean AUC	NR	p(se) = 1.00; p(sp) = 0.885; p(AUC) = 0.183	NSD between iPad® and calibrated secondary liquid crystal display
Lee et al[16]	Difference in AUC	5%	NR	NSD between tablet and primary CRT for the detection of urolithiasis

ac, accuracy; ANOVA, analysis of variance; AUC, area under receiver operating characteristic curve; CRT, cathode ray tube; CTPA, CT pulmonary angiogram; DBM, Dorfman-Berbaum-Metz; GSDF, greyscale display function; MRMC, multireader multicase; NR, not reported; NSD, no significant difference; se, sensitivity; sp, specificity.
iPad in the table refers to the Apple iPad (Apple, Cupertino, CA).

diagnosis.[22] Park et al[8] found the interpretation on the iPad had greater diagnostic accuracy than did a GSDF-calibrated secondary LCD. However, in this study, the luminance of the reference test LCD was set to 170 cd m^{-2} compared with the 400 cd m^{-2} on the iPad. Furthermore, the reading was performed under conventional lighting. Hence, it would be reasonable to expect inferior performance from a low luminance monitor in high ambient lighting. Yoshimura et al[19] concluded that the diagnostic

Figure 3. Sensitivity and specificity of interpretation on tablet display [error bars are 95% confidence intervals (CIs)].

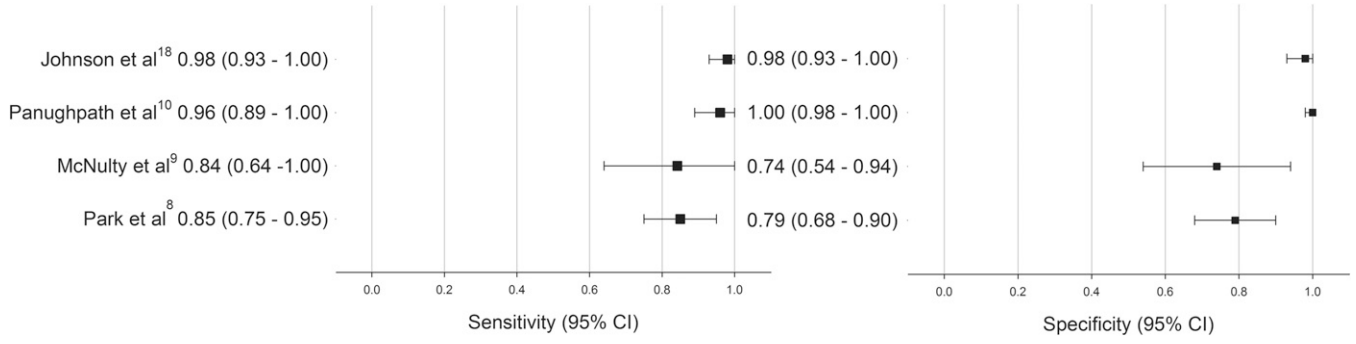
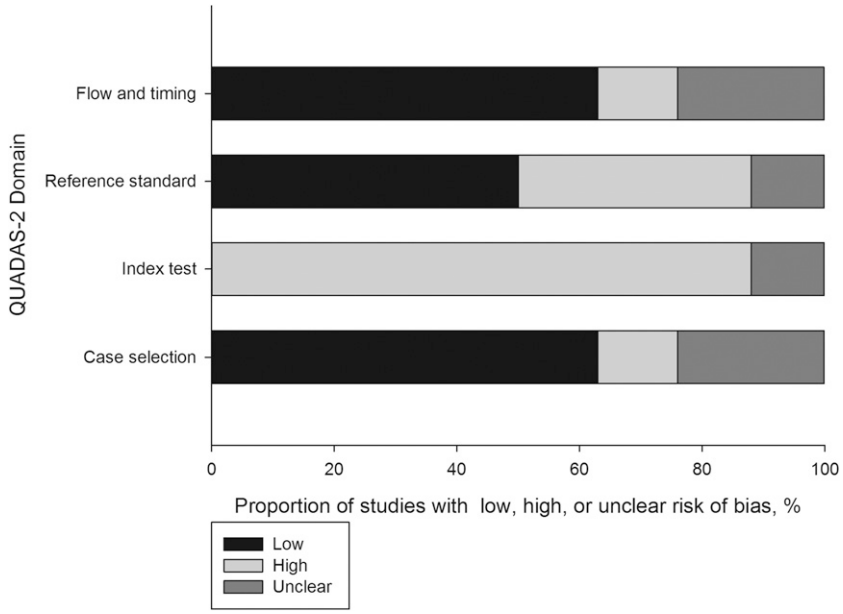
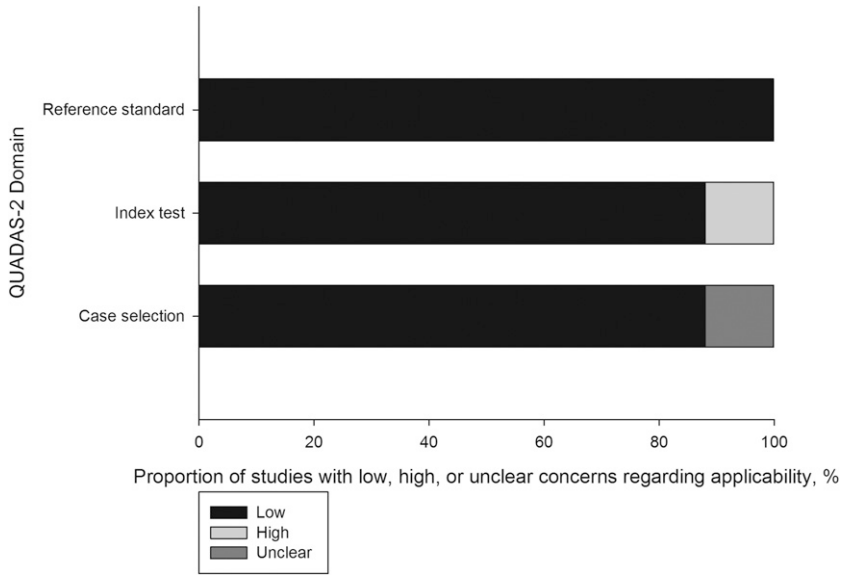


Figure 4. Summary of Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) assessments for risk of bias.



accuracy of images displayed on an iPad was significantly less than when displayed on a Gamma 2.2-calibrated LCD, but not significantly less than when displayed on a GSDF-calibrated monitor. DICOM GSDF is the most widely used calibration technique used in radiology today.[23] There appears to be some limitation in the methods used by McLaughlin et al[1] who classified discrepancies. The limitation has been caused by the index test and reference test instrument not being calibrated. This lack of calibration has resulted in inconsistency when grading discrepancies for incidental findings, *e.g.* mucosal thickening. This inconsistency obscures whether discrepancies are attributable to the display or a difference in radiologist's reporting style. Most authors used arbitrarily selected cases, which may not be a true representation of positive-to-negative case ratios. The use of only subtle pathology in one study may have resulted in under reporting of the accuracy rate.[8] In many of the studies, the reader was blinded to clinical details of the referral and did not have access to prior imaging. In other studies, the readers were aware of a limited type of pathology in which the studies needed to be reported. Both scenarios may affect interpretative performance. The reader population was not representative of a radiologist population in a number of studies—for example, McNulty et al[9] used 13 subspecialists as the study readers; whereas, only two readers were used in other studies.[1,10,16,18] The use of a subspecialist may result in overreporting of diagnostic accuracy. The authors of nearly all studies have minimized bias of retained information by including a time delay between readings. To further reduce bias from the retained

Figure 5. Summary of the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) assessment for applicability.



information, some authors have randomized the reading order (tablet *vs* control) and randomized the case order.[17]

Many authors identified the likely application of a tablet computer was for remote on-call interpretation of emergency imaging and chose cases and pathology relevant to this situation.[1,2,8–10,15,18,20] Hence, the included studies were judged to have a high applicability to the review question. The exceptions were the study by Abboud et al,[17] which tested the reliability of TB screening; the use of emergency department physicians as the study readers by Park et al;[8] and the accuracy of diagnosing urolithiasis from the abdominal radiographs[16] (which has in a large part been replaced by CT imaging[24]).

The studies were undertaken in eight countries indicating the international interest in the use of tablet computers for diagnostic radiology. The US Food and Drug Administration have cleared the use of the Apple iPad for primary radiological diagnosis. However, the clearance is limited to use with a Mobile MIM (MIM Software, Cleveland, OH) software application, small-matrix images and situations where there is no access to a primary diagnostic display.[25] To the best of our knowledge, there is no similar approval in any other country. In the UK and Germany, regulatory guidelines prevent tablet computers being used for primary diagnosis owing to screen size.[15,26]

Secondary observations on the use of tablet computers were elicited in many of the studies. The various software applications were criticised by a number of authors for the difficulty in scrolling and touch movements, cumbersome user interface (especially when trying to compare previous imaging), the lack of post-processing tools and instability (especially for large studies).[1,2,8] Limitations in network coverage and speed, potential effect of ambient lighting on diagnosis were also noted.[8,15,20] The inability to access clinical systems for referral information and prior studies[1] and the increased time to perform a read compared with primary workstation[19] were other limitations of using tablet computers. The portability and fast boot time were seen as the major advantage of tablet computers.[8,10,27]

This review has a number of limitations. There were only a small number of heterogeneous studies, which made consolidation of results difficult. The small sample size and the large number of studies judged to be at risk of bias may reduce confidence in the findings of this review. The iPad was assessed in 10 out of 11 studies. Hence, the findings are only applicable to this device. There is currently no evidence in favour of, or against, the use of any other makes or models of tablet computer. Similarly, the findings are only applicable to modalities tested in the included studies, namely CT, MRI and plain radiography.

CONCLUSION

In conclusion, the findings of this review suggest that the diagnostic accuracy of radiological interpretation was not compromised by using a tablet computer to interpret CT, MRI or plain radiography. This conclusion is only applicable to the Apple iPad and indicates this device is suitable for an on-call radiologist to display images for interpretation. The conclusions were based on studies that were judged to be at risk of bias. There were low concerns regarding the applicability of the included studies to the review question. The use of tablet computers for a primary diagnosis may be subject to local regulatory guidelines. When considering the usage of a tablet computer for on-call radiology, the user also needs to assess the software functionality—for example, access to referral information and previous imaging; software stability and network performance. These have all been identified by the authors of included studies as potential impediments in using a tablet computer for radiological interpretation.

REFERENCES

1. Mc Laughlin P, Neill SO, Fanning N, McGarrigle AM, Connor OJ, Wyse G, et al. Emergency CT brain: preliminary interpretation with a tablet device: image quality and diagnostic performance of the Apple iPad. *Emerg Radiol* 2012; **19**: 127–33. doi: 10.1007/s10140-011-1011-2
2. John S, Poh AC, Lim TC, Chan EH, Chong le R. The iPad tablet computer for mobile on-call radiology diagnosis? Auditing discrepancy in CT and MRI reporting. *J Digit Imaging* 2012; **25**: 628–34. doi: 10.1007/s10278-012-9485-3
3. Fetterly KA, Blume HR, Flynn MJ, Samei E. Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays. *J Digit Imaging* 2008; **21**: 193–207. doi: 10.1007/s10278-007-9022-y
4. Buls N, Shabana W, Verbeek P, Pevenage P, De Mey J. Influence of display quality on radiologists' performance in the detection of lung nodules on radiographs. *Br J Radiol* 2007; **80**: 738–43. doi: 10.1259/bjr/48049509
5. Krupinski EA, Roehrig H. The influence of a perceptually linearized display on observer performance and visual search. *Acad Radiol* 2000; **7**: 8–13. doi: 10.1016/s1076-6332(00)80437-7
6. Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, et al; AAPM TG18. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. *Med Phys* 2005; **32**: 1205–25. doi: 10.1118/1.1861159
7. Yamazaki A, Liu P, Cheng WC, Badano A. Image quality characteristics of handheld display devices for medical imaging. *PLoS One* 2013; **8**: e79243. doi: 10.1371/journal.pone.0079243
8. Park JB, Choi HJ, Lee JH, Kang BS. An assessment of the iPad 2 as a CT teleradiology tool using brain CT with subtle intracranial hemorrhage under conventional illumination. *J Digit Imaging* 2013; **26**: 683–90. doi: 10.1007/s10278-013-9580-0
9. McNulty JP, Ryan JT, Evanoff MG, Rainford LA. Flexible image evaluation: iPad *versus* secondary-class monitors for review of MR spinal emergency cases, a comparative study. *Acad Radiol* 2012; **19**: 1023–8. doi: 10.1016/j.acra.2012.02.021
10. Panughpath SG, Kumar S, Kalyanpur A. Utility of mobile devices in the computerized tomography evaluation of intracranial hemorrhage. *Indian J Radiol Imaging* 2013; **23**: 4–7. doi: 10.4103/0971-3026.113610
11. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool

for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010; **63**: 854–61. doi: 10.1016/j.jclinepi.2009.10.002
12. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003; **228**: 303–8. doi: 10.1148/radiol.2282011860
13. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–36. doi: 10.7326/0003-4819-155-8-201110180-00009
14. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**: 25. doi: 10.1186/1471-2288-3-25
15. Tewes S, Rodt T, Marquardt S, Evangelidou E, Wacker FK, von Falck C. Evaluation of the use of a tablet computer with a high-resolution display for interpreting emergency CT scans. *Rofo* 2013; **185**: 1063–9. doi: 10.1055/s-0033-1350155
16. Lee W-J, Lee H-T, Ching Y-T, Tsai C-H, Liu H-M, Chen S-J. Tablet PC as a PACS workstation: observer performance evaluation. *Chin J Radiol* 2005; **30**: 277–81.
17. Abboud S, Weiss F, Siegel E, Jeudy J. TB or Not TB: interreader and intrareader variability in screening diagnosis on an iPad *versus* a traditional display. *J Am Coll Radiol* 2013; **10**: 42–4. doi: 10.1016/j.jacr.2012.07.019
18. Johnson PT, Zimmerman SL, Heath D, Eng J, Horton KM, Scott WW, et al. The iPad as a mobile device for CT display and interpretation: diagnostic accuracy for identification of pulmonary embolism. *Emerg Radiol* 2012; **19**: 323–7. doi: 10.1007/s10140-012-1037-0
19. Yoshimura K, Nihashi T, Ikeda M, Ando Y, Kawai H, Kawakami K, et al. Comparison of liquid crystal display monitors calibrated with gray-scale standard display function and with γ 2.2 and iPad: observer performance in detection of cerebral infarction on brain CT. *AJR Am J Roentgenol* 2013; **200**: 1304–9. doi: 10.2214/AJR.12.9096
20. Bhatia A, Patel S, Pantol G, Yen-ying W, Plitnikas M, Hancock C. Intra and inter-observer reliability of mobile tablet PACS viewer system *vs.* standard PACS viewing stations-diagnosis of acute central nervous system events. *Open J Rad* 2013; **3**: 91–8. doi: 10.4236/ojrad.2013.32014
21. Skaron A, Li K, Zhou XH. Statistical methods for MRMC ROC studies. *Acad Radiol* 2012; **19**: 1499–507. doi: 10.1016/j.acra.2012.09.005
22. Espeland A, Vetti N, Kråkenes J. Are two readers more reliable than one? A study of upper neck ligament scoring on magnetic resonance images. *BMC Med Imaging* 2013; **13**: 4. doi: 10.1186/1471-2342-13-4
23. Hirschorn DS, Krupinski EA, Flynn MJ. PACS displays: how to select the right display technology. *J Am Coll Radiol* 2014; **11**: 1270–6. doi: 10.1016/j.jacr.2014.09.016
24. Levine JA, Neitlich J, Verga M, Dalrymple N, Smith RC. Ureteral calculi in patients with flank pain: correlation of plain radiography with unenhanced helical CT. *Radiology* 1997; **204**: 27–31. doi: 10.1148/radiology.204.1.9205218
25. US Food and Drug Administration. FDA clears first diagnostic radiology application for mobile devices. 2011. [Updated 4 February 2011; cited 1 January 2015.] Available from: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm242295.htm
26. Royal College of Radiologists, Board of the Faculty of Clinical Radiology. Picture archiving and communication systems (PACS) and guidelines on diagnostic display devices. [Updated 1 November 2012; cited 1 January 2015.] Available from: http://www.rcr.ac.uk/sites/default/files/docs/radiology/pdf/BFCR%2812%2916_PACS_DDD.pdf
27. Yoshimura K, Shimamoto K, Ikeda M, Ichikawa K, Naganawa S. A comparative contrast perception phantom image of brain CT study between high-grade and low-grade liquid crystal displays (LCDs) in electronic medical charts. *Phys Med* 2011; **27**: 109–16. doi: 10.1016/j.ejmp.2010.06.001