Is there an advantage to using computer aided detection for the early detection of pulmonary nodules within chest X-Ray imaging?

HABER, M., DRAKE, Amelia and NIGHTINGALE, JulieJ

‘Is there an advantage to using Computer Aided Detection for the early detection of Pulmonary Nodules within Chest X-Ray imaging?’

Introduction

Lung or pulmonary nodules (PN) are commonly found as opacities on radiological lung imaging. These opacities are indicative of a range of ailments ranging from benign infection to tumour, malignant lung cancer or metastasis [1]. Chest X-rays (CXRs) are the most commonly performed examinations within a medical imaging department, being the most readily available, and the cheapest modality to examine the lungs [2]. They also offer the additional advantage of being a low dose examination, particularly when compared to Computed Tomography (CT) [3]. Moreover, CXRs remain the first line investigation for suspected lung cancer and hence play an important role in the potential diagnosis of PNs, as the modality is correlated to a lesions’ early discovery, possibly resulting in a better prognosis for patients [4]. A drawback of the modality, however, is the poor sensitivity to PNs of diameter smaller than 2cm [2]. Additionally CXR image interpretation is subject to interpreter error, leading to erroneous diagnoses and patient mismanagement but potentially also to medicolegal repercussions [3; 5]. Indeed overlooked lung cancers on CXRs were quoted to be the 6th most frequent reason of medicolegal action against interpreters who were noted to miss 90% of nodules on CXR and 10% on CT [6].
The error rate of 20-60% for missed PNs on CXRs has remained the same for years [3;7].To reduce the risk of misdiagnosis and malpractice suits, software technologies such as computer aided detection (CAD) systems were developed to aid interpretation for CXRs as well as for other anatomical areas such as breast imaging [8].
A literature search was carried out to identify studies within this field. Many quantitative studies and narrative reviews were published. CAD in its early stages of development was defined as

pattern recognition software that identifies suspicious features on the image and then alerts the reporting clinician in an effort to reduce false negative readings [9]. This often involves the reporting clinician initially reviewing the image, then activating CAD software, and then re-evaluating any flagged areas for concern prior to writing the definitive report [9]. In effect the CAD becomes a 'second reader'. However early research in this field shows contradictory results. While some studies show that CAD technology improves the detection sensitivity of reasonably elusive PNs [10;11], others suggested that although CAD did improve PN detection (94% accuracy) it was at the cost of a high False Positive (FP) rate leading to unnecessary further imaging and higher radiation dose to patients [12]. Most previous work has compared CXR to CT for identifying lung lesions, including an earlier systematic review by Amir & Lehmann [13]. Adding to the complexity of this debate is the more recent drive towards the use of Artificial Intelligence (AI), in particular Machine Learning and Deep Learning convolutional neural networks (CNN). Within radiology contexts, AI refers to the more advanced components of deep learning/CNN, capable of being 'trained' to recognize subtle patterns in medical images, essentially triaging normal and abnormal examinations. A recent landmark study by Annarumma et al. demonstrated that real-time triaging of adult chest radiographs with use of an artificial intelligence system is feasible, with clinically acceptable performance [14]. With radiologist workforce shortages evident in many countries, CXR reporting backlogs are commonplace, with many CXRs going unreported or reported too late to influence clinical management [15]. Annarumma et al.'s study demonstrates that there is potential for AI to triage CXRs, highlighting the images that require priority reporting [14].

Within the field of radiology, AI has been restricted to the role of a rapidly expanding research topic to this point in time [16], with 10% of submissions to the prestigious Radiology journal

being AI-related in 2019 [16]. However it appears likely that AI-dependent medical imaging in day-to-day practice is not too far away [16]. At the current time, however, it is prudent to systematically review the efficacy of CAD systems that have been available for use in clinical practice, often based on early machine learning algorithms [17]. However there is currently no published evidence demonstrating how widespread is the use of these CAD tools.

A systematic review seeks to appraise and synthesize the evidence obtained through wide-ranging searches [18].To formulate a focused research question, the 'PICOT' principle was used [19] (Table 1). Consequently, the following research question was created; *'Is there an advantage to using CAD for the early detection of PNs within CXR imaging?'*

Population/Patient	Patients with lung cancer
Intervention	CXR screening
Comparison	CXR with the use of CAD diagnosis
Outcome	Early detection of PNs
Type of Study	Systematic review of published quantitative literature.

Table 1 PICOT principle.

Materials and methods

Search for published literature

For a sensitive and specific search, synonyms and different acronyms were identified a priori [20]. Table 2 details the search term variations used within the search for both systematic reviews and other quantitative studies.

Column Terms Combined with	Patient Condition AND	Intervention AND	Comparative Intervention AND	Outcomes AND
OR	Lung Cancer	Chest X-Ray	CAD	Early detection
OR	Pulmonary Nodule	Chest X-ray	Computer-Aided Detection	Improved prognosis
OR	Solitary Pulmonary Nodule	CXR	Computer-Aided Diagnosis	Reduced mortality
OR	PN	Chest Radiograph	Computer Aided Detection	Better Diagnosis
OR	SPN	Chest Radiography	Computer Aided Diagnosis	Early Diagnosis
OR	Lung Tumour	Thoracic Radiograph	Computer Assisted Diagnosis	Treatment
	Lung Tumour		Computer Assisted Detection	

Table 2 Search Terms

The Boolean search developed in Table 2 was employed within seven electronic databases [21; 22]. Databases used were: Cochrane Library, TRIP database, EBSCO Host Medline, Scopus, Science Direct Elsevier, CINAHL, and NCBI- PUBMED.

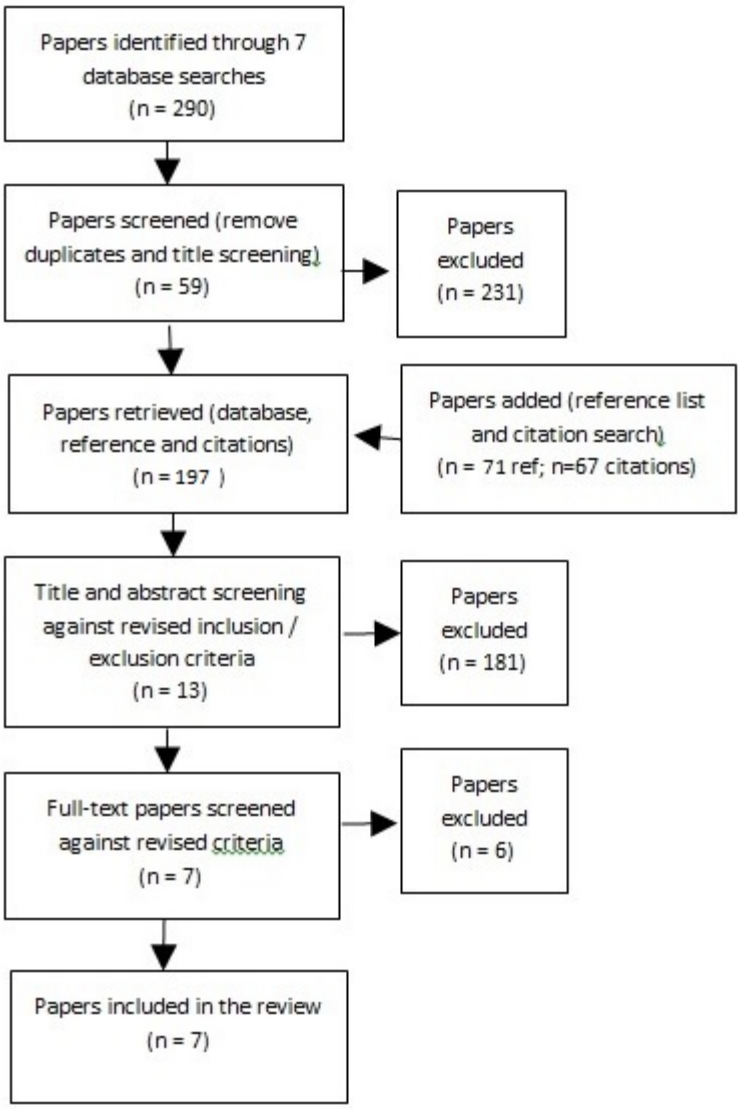

Fig.1. PRISMA Flow Chart of article search strategy

A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart (Fig. 1), details the article search strategy [21]. After the list of articles was obtained, the resulting articles' reference lists were scanned for other potential journal publications [22], producing two additional articles which were deemed relevant [10; 12]. The other articles retrieved were all either duplicates or irrelevant literature-based studies not revolving around CAD and CXRs in conjunction. By searching through the reference list and citations an

additional 69 and 67 articles were retrieved respectively. This further increased the number of articles to 197. A brief title and abstract review of these articles at this stage identified that some articles were not directly relevant, and a more focused inclusion and exclusion criteria were needed. Time constraints also required further narrowing of the criteria. This included restricting the date range to articles published after 2010, to coincide with the development of CAD as an established tool. Additionally this also automatically excluded any conventional CXRs due to the technological advancements within more recent years. Table 3 details the inclusion and exclusion criteria applied.

Inclusion Criteria	**Exclusion Criteria**	**Number of Articles Excluded**	**Number of Articles Left**
2010-2017	Articles before 2010	115	78
CAD as the only technique	No Artificial Neural Network (ANN), Tomosynthesis, or Digital Subtraction etc.	31	47
Cohort, RCTs, CCTs and Systematic Reviews	Narrative Reviews	13	34
CXR imaging modality only	No PET, Mammography, CT, MRI or Ultrasound	8	26
Journal Articles or peer reviewed published poster abstracts	Books or Book Chapters	1	25
PNs in General	No Specific Disease to be Diagnosed	9	16
No time constraints on reporting	Time constraints on reporting CXRs	1	15
PA radiographs only	No AP or Lateral radiographs	1	14
CAD PN detection	CAD Detection by Chance or True Findings	1	13
Patients with PNs who are adults	Paediatric patients >16	0	0
Digital radiography only	Conventional CXRs to be included	0	0
Studies whose reporting radiologists are experienced (>/=2 years)	Studies whose reporters are only inexperienced radiologists, trainees or residents	0	0

Table 3 Inclusion and Exclusion Criteria of Search Strategy

To ensure that the selection of primary studies was standardized and not tainted by selection bias, a *'research paper selection form'* was filled out for each article [20; 23]. All the inclusion criteria detailed in Table 3 and Table 4 were revisited for these articles.

Out of the 13 studies, 6 studies were excluded due to various exclusion criteria [12; 24; 25; 26; 27; 28]. Four of these studies [12; 24; 25; 26] used a lateral view CXR in addition to the PA for the determination of PNs. This not only increased the odds of detection but also differed from the clinical setting within which a PA CXR is usually performed. Additionally, one study [27] was eliminated after it was discovered to be a narrative review, not immediately apparent on title review. The final article [28] was also eliminated on the basis of having more than 5 exclusion criteria and therefore too many variables.

Type of Studies	**Exclusion**	**Inclusion**
Quantitative	Books	Systematic reviews
	Case Studies	RCTs
	Narrative Reviews	CCTs
	Qualitative Studies	Cohort Studies
	Article abstracts	Case- Control Studies

Table 4 Inclusion and Exclusion criteria of reviewed studies

Data extraction

Seven articles remained which were scrutinised to extract key data (Table 5) that were assembled following the PICOT strategy [20]. The determination of whether CAD improved accuracy and/or sensitivity of detection of PNs was the purpose of all primary studies selected. These included patients who were over the age of 16, therefore eliminating any paediatric patients.

Additionally, all studies had the same intervention i.e. the PA CXR and the same comparative i.e. PA CXR with CAD.

The common diagnostic tool against which CAD was assessed was CT, as this had been found to be superior in diagnostic quality than CXRs [29] (Table 6). Furthermore, histological results from lung biopsies were used in all studies with the exception of [30] and [31]. With respect to the comparative intervention i.e. CAD, different systems were used depending on the year of publication (Table 7).

Study/Year	Population	Intervention	Comparative Intervention	Outcomes
Meziane et al. (2011)	>45 years. 100 patients with PNs and 100 control patients	PA CXR	PA CXR with CAD; RapidScreen 1.1 and OnGuard 3.0	CAD did not improve the performance of Chest and General Radiologists.
Moore et al. (2011)	16-88 years. . Retrospective cohort study of 240 patients incl. negative and positive findings	PA CXR	PA CXR with commercial CAD	CAD for PN detection when used by an experienced radiologist proved good sensitivity, accuracy and specificity.
Lee et al. (2012)	21-86 years. 100 PN patients and 100 control patients.	PA CXR	PA CXR with commercial CAD	CAD could possibly improve PN detection.
Meziane et al. (2012)	45 years. 100 patients with PNs and 100 controls.	PA CXR	PA CXR with 4 versions of CAD; RapidScreen 1.1, OnGuard 3.0, 4.0. and 5.0	Latest version of CAD software shows good detection rate with lower FP rate.
Kligerman, Cai and White (2013)	>46-90years 81 patients and 215 controls	PA CXR	PA CXR with CAD OnGuard 5.1	The use of OnGuard 5.1 software improved reader accuracy and sensitivity.
Mazzone et al. (2013)	40–75 years 710 patients and 713 controls	PA CXR	PA CXR with CAD OnGuard 5.0	Undetermined outcomes. The RCT was stopped early due to slow recruitment.
Frolkis and Gilkeson (2014)	44-91 years. 41 patients. No control mentioned	PA CXR	PA CXR with CAD OnGuard 5.2	Superior detection of lung nodules with CAD with fewer FPs.

Table 5 Data Extraction (i)

Due to the lack of homogeneity within the primary papers, such as statistical tests used as well as the variation in the kind of results presented, a meta-analysis could not be undertaken and therefore results could not be accurately synthesised [32].

Study/Year	PN Population	Control	Gold Standard	Intervention	No. of Raters	Rater Experience Mean Years	Comparative Intervention /Algorithm	Statistical Test Used
Meziane et al. (2011)	100 patients	100	CT and Histology	PA CXR	18	N/A	CAD; RapidScreen 1.1 and OnGuard 3.0	McNemar Test
Moore et al. (2011)	240 patients with negative and positive findings	N/A	CTA	PA CXR	1	5	Commercially Available CAD	Spearman Rank Correlation
Lee et al. (2012)	100 patients	100	CT and Histology	PA CXR	10	N/A	Commercially Available CAD	JAFROC analysis and McNemar Test
Meziane et al. (2012)	100 patients	100	CT and Histology	PA CXR	2	N/A	4 versions of CAD; RapidScreen 1.1, OnGuard 3.0, 4.0. and 5.0	McNemar Test
Kligerman, Cai and White (2013)	81 patients	215	CT and Histology	PA CXR	11	3.8	CAD OnGuard 5.1	2-tailed *t* Test and Wilcoxon Signed Rank Test
Mazzone et al. (2013)	710 patients	7l3	CT	PA CXR	2	N/A	CAD OnGuard 5.0	Two-Sample *t*-Test /Wilcoxon Two-Sample Test
Frolkis and Gilkeson (2014)	41 patients	N/A	CT and Histology	PA CXR	1	15	CAD OnGuard 5.2	N/S

Table 6 Data Extraction (ii) (N/S Not Stated, N/A Non Applicable).

Study	Sensitivity CAD only	FP/ Image	Sensitivity Radiologist	FP/ Image	Sensitivity CAD and Radiologists	Confidence Interval (CI)	P Value (Improvement of Radiologist Performance with use of CAD)
Meziane et al. 2011)	RapidScreen: 44.2%; OnGuard: 62.5%	Rapid Screen:3.9 OnGuard 3.3	0.68	N/S	OnGuard 3.0: 0.70	95%	(RapidScreen 1.1) 0.283 (OnGuard 3.0) 0.013
Moore et al. (2011)	0.707 71%	0.48	N/S	0.48	N/S	95%	N/S
Lee et al. (2012)	59%	0.19	85.2%	0.17	87%	N/S	0.02-1.00
Meziane et al. (2012)	Rapid Screen 1.1 : 44.2% OnGuard 3.0: 62.5% OnGuard 4.0: 62.5% OnGuard 5.0: 64.4%	Rapid Screen 1.1 : 3.9OnGuard 3.0: 3.3 OnGuard 4.0: 2.6 OnGuard 5.0: 2.0	N/A	N/A	N/A	95%	Rapid Screen 1.1 vs OnGuard 3.0 <0.0001 OnGuard 3.0 vs OnGuard 4.0 <0.0001 OnGuard 4.0 vs OnGuard 5.0 <0.0001
Kligerman, Cai and White (2013)	49.4%	1.8	0.69	N/S	0.74	95%	0.007
Mazzone et al. (2013)	N/S	N/S	N/S	N/S	0.978	N/S	N/A
Frolkis and Gilkeson (2014)	67%	0.75	N/S	N/S	N/S	N/S	N/S

Table 7 Percentage CAD Sensitivity vs FP / Image

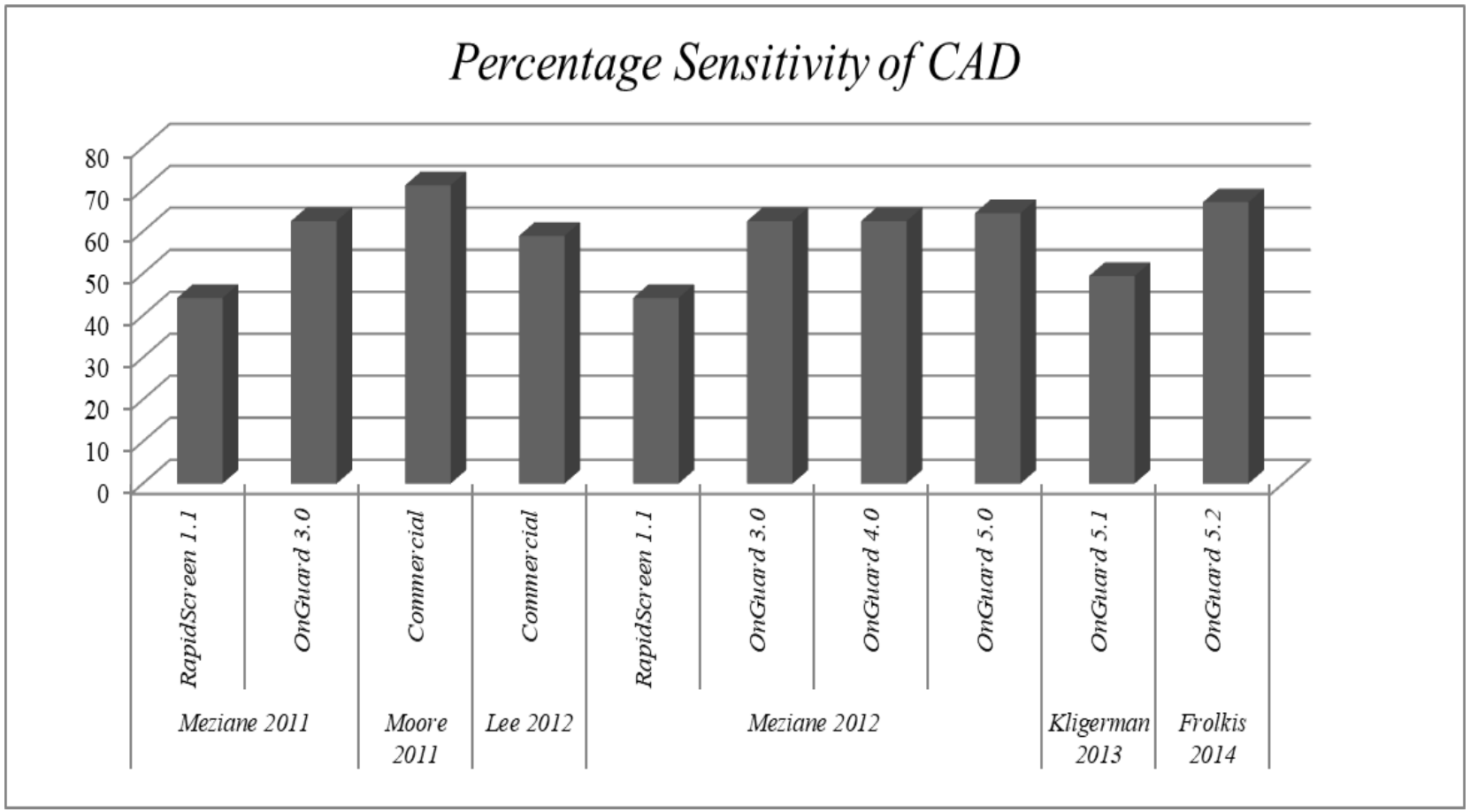

Fig. 2 Percentage sensitivity of CAD according to respective studies

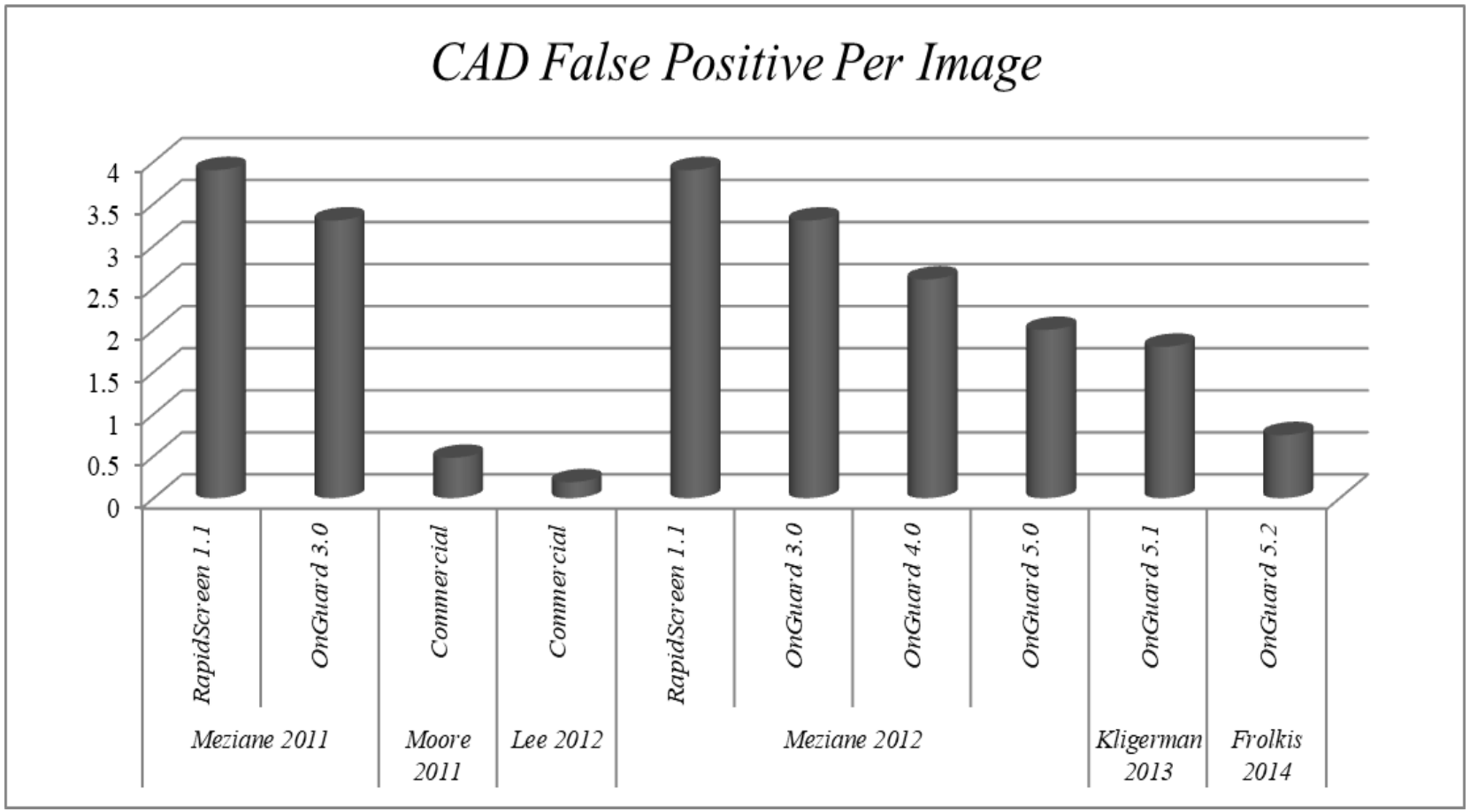

Fig. 3 CAD FP per image according to respective studies

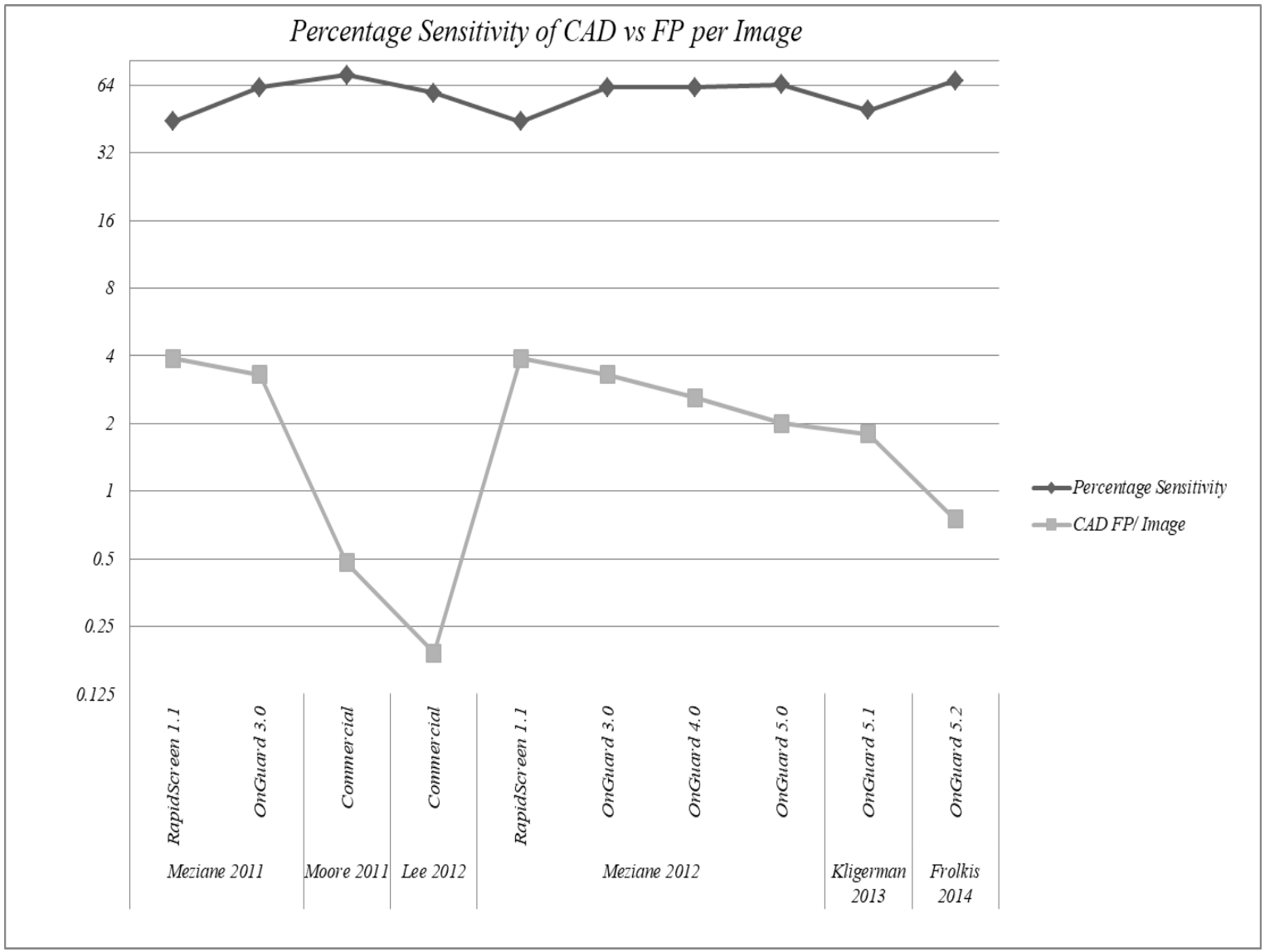

Fig. 4 Relationship between percentage sensitivity of CAD and FP rate

The QUADAS-2 tool was used to inform analysis as it is suited for 'Quality Assessment of Diagnostic Accuracy Studies' [33]. Seven primary studies were included within this review, nonetheless only six of these presented their result findings. The only numerical data available within the Randomised Control Trial (RCT) [30] detailed the sensitivity of CAD and interpreting radiologists as an ensemble without any mention of the sensitivity of CAD and radiologists independently [30]. Moreover this study was terminated early due to slow recruitment of patients and hence its findings were inconclusive. This led to the exclusion of the latter study from the data synthesis. The average CAD sensitivity result of the rest of these studies, excluding the latter, was 58.67% (range; 44.2%- 71%) alongside a mean 2.22 (range; 0.19- 3.9) FP rates per image (Fig. 2-3).

Fig. 4 suggested that the higher the percentage sensitivity of CAD gave rise to a lower FP rate per image detected. To test this relationship and answer the question as to whether there is correlation, a regression analysis on the data available was performed [34]. Unfortunately, it failed to confirm correlation, (Fig. 5) though it may be worthwhile to revisit this relationship with a larger dataset.

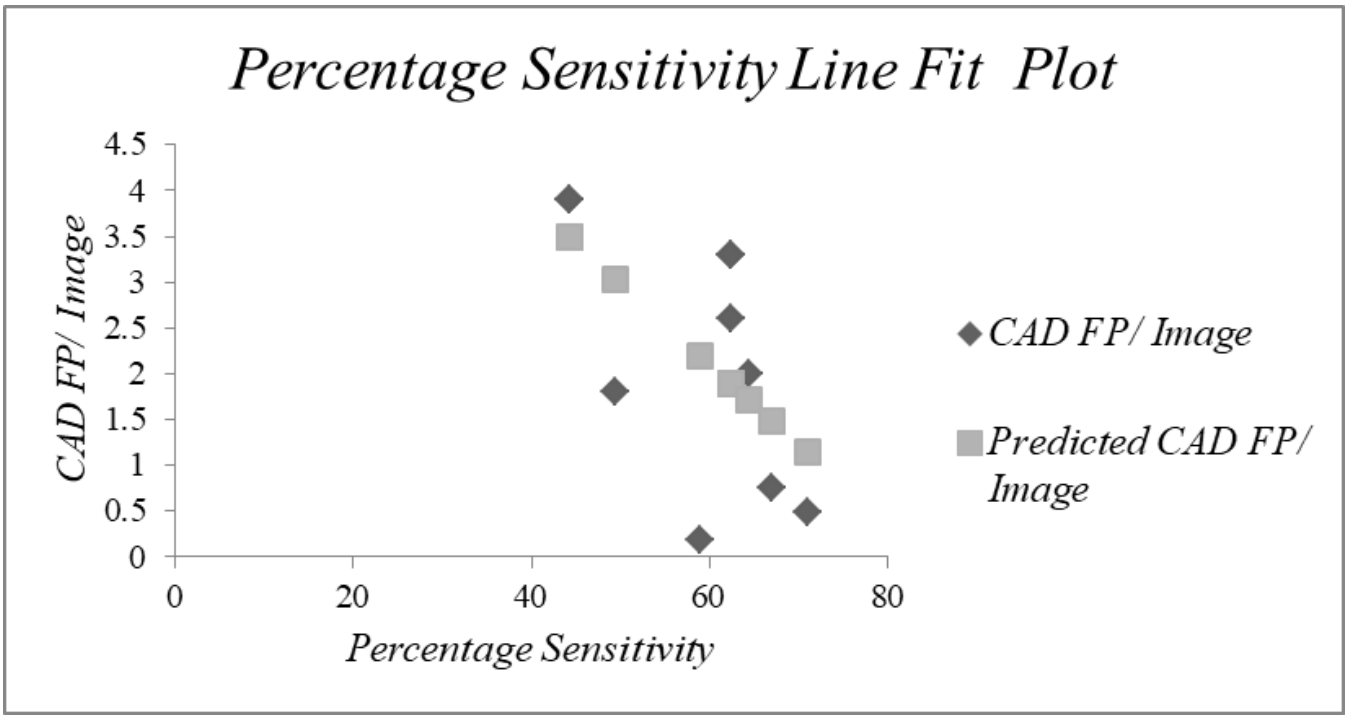

Fig. 5 Percentage sensitivity Line Fit Plot

Encouragingly, most of these more recent studies (2011-2013) suggest that advancement in technologies have improved the algorithms resulting in recommendations for the use of CAD for the improved detection of PNs. Only two of the seven studies in this review did not agree with these findings [30; 35], with the latter study not having been concluded due to slow recruitment of patients. Another study suggested that lack of experience of the reporter may still lead to a FP impact and subsequent unnecessary follow-ups, imaging and biopsies [31].

Discussion

Variation in included studies

The revised inclusion and exclusion criteria used within the searches resulted in seven studies being included in the review. While the criteria would have been better applied at the commencement of the search strategy, prior to reference list and citation searches of the retrieved articles, it yielded an appropriate range of articles for review. Seven articles may be seen as a small number however additional augmentation of the results of the electronic searches, including hand searching of key journals, reference lists and citations of included papers, allows assurance that all relevant articles were included in this systematic review.

Although all studies included a PA CXR with CAD, they were performed in different years and countries. Inevitably the CXRs included in this review were performed with different X-ray acquisition and display systems, introducing additional variables. On some occasions [35] different acquisition systems were used within the same study itself; 125 Siemens Healthcare, 68 Canon and seven other manufacturers' equipment. Most studies were retrospective with retrieved and analysed data from different systems and hospitals. All CXRs should adhere to minimum quality standards [36]; however it is known that the selection of different imaging parameters such as higher kVs can influence the resultant image, and different acquisition and display mechanisms can influence the ability to optimise the image for interpretation. It is therefore conceivable that different systems gave varying quality which led to diagnostic inconsistency within diagnoses. This is an inevitable limitation of clinical rather than experimental or quasi-experimental studies.

Additionally all studies used a different CAD system which, similarly, could have affected the research synthesis [10]. Ideally, the study would have compared the same CAD version

throughout all the literature such as to reduce variants as much as possible. An added variation was that while some studies included all types of nodules irrelevant of size, shape and location, a particular study [31] only included PNs sizes ranging from 0.5 to 1.5cm due to difficulty in detection. The latter credited this choice to the specific CAD's detection range stating that the algorithm ignored nodules outside these parameters.

Quality of included studies

Overall the reliability of the primary studies was questionable due to their lack of reliability tests. Out of the eight kinds of threats to internal validity, only two applied to the primary studies [37]. These were instrumentation threat due to population selection bias, and single group threat. The latter is concerned with the lack of a control group which was the case for two studies [31; 38]. Moreover, the control group needs to be matched to the study group, which was only the case for [11] showing strong internal validity. The rest of the studies especially [38] showed poor internal validity. One study [31] who did not include a control group instead performed a retrospective cohort study which included all patients who had undergone a CXR and CTA within a year irrespective of findings. This is regarded as a gold standard of observational research [36]. The latter study [31] had an improved internal validity when it came to specific inclusion criteria; this indicates very high internal validity but poor external validity as the study's population will differ greatly from normal settings [39]. On the other hand, the rest of the studies showed high external validity due to lack of specific inclusion criteria. The only study which showed high internal validity [11] had limited or low external validity since internal and external validity are inversely related [39; 40]. In one study [31] internal validity was once again questioned due to a dubious declaration, which conceded to the fact that all images were included within the study

irrelevant of image quality. Nevertheless the external validity of this study was in turn strengthened as this simulated the clinical setting more closely. Contradictorily however, one study [11] excluded poor quality images such as not to effect the overall study's outcome by adhering to strict medically established criteria. While this strengthened internal validity, external validity was compromised due to a disparity from the real clinical setting. Owing to the lack of specific inclusion criteria in the rest of the studies the overall external validity was considered high whereas internal validity was considered to be low.

Implications for Practice

In the six primary studies the combined patient population including controls, totalled 1,173. The average CAD sensitivity was 58.67% (range; 44.2%- 71%) alongside a mean 2.22 (range; 0.19-3.9) FP rates per image. A regression analysis performed to evaluate the relationship between CAD sensitivity and CAD FP rate per image showed that there was no correlation between the two factors after obtaining a *p*-value of 0.07. All the primary studies which were included within this review collectively showed that there is the need for further research for a definite conclusion with regards to the benefit of CAD within the context of PN diagnosis.

Because PNs can represent a plethora of diseases [41], the authors ensured that the patients included did not represent only one specific kind of disease to minimise bias and over positive results. However PNs may be associated with early lung cancer. Since lung cancer is the leading cause of death worldwide [42], its early detection should be taken seriously. In a 2014 survey a patient diagnosed with stage I lung cancer had a one-year net survival of 83% while a patient diagnosed with a stage IV had a survival rate of 17% [43].

The CXR is a simple, cheap and effective tool used as an initial diagnostic test for patients with clinical symptoms, making it the most common imaging procedure in most radiology departments. However image interpretation is not without problems, with an error rate of 4% often quoted, even amongst experienced reporters [44]. When this rate is taken into a worldwide perspective, it would equate to 40 million interpreter errors per year [44]. Even more worrying in some countries, including the United Kingdom, is a backlog in reporting which means that CXRs can go unreported, or reported too late to influence patient management [15]. A number of solutions have been proposed including expanding the role of reporting radiographers to include CXR interpretation, though this is not universally embedded or accepted. Nevertheless radiographers have demonstrated effectiveness in this role [45-47]. In parallel with these workforce solutions, artificial intelligence opportunities are being explored with some potentially promising results, particularly in using AI in a large volume triaging situation [48]. While not yet in widespread clinical use, imaging technology companies are working to bring these new technologies to market. So what future is there for the 'second reader' CAD solutions, if there is a fundamental lack of capacity for the 'first readers'? CAD needs to be efficient as well as effective, not adding any noticeable time to the reporting process where interpreters are in short supply. As well as being sensitive to PNs, CAD must also demonstrate low false positive rates; high FP rates cause unnecessary stress to the patient and their families, as well as additional unnecessary further imaging, possible histology or further consultation costs. Unfortunately the false positive rates identified in many of the included studies were high (up to 3.9 FP per image), although these reduced with subsequent revisions of the CAD technology. Each of these FP 'flags', as well as any true positive flags, requires additional interpreter time to review the potential lesion.

Lung cancer screening has become a major public health priority in many countries; in countries with limited access to CT, the role of CXR with and without CAD have been explored as a potential screening tool, but the complex reporting requirements and high FP rates make it non-viable [48;49]. While the sensitivity and false positive rates of CAD technology has continued to improve over time, it is likely that AI, with the enhanced potential to 'learn' to triage more effectively, will rapidly replace any current clinical applications for CXR CAD.

Conclusion

This systematic review aimed to identify whether there was an advantage to using Computer Aided Detection (CAD) to support CXR interpretation of pulmonary nodules; our findings were inconclusive. From initial 290 articles retrieved, seven studies were included in the review following a systematic screening process. The average CAD sensitivity in these studies was 58.67% (range; 44.2%- 71%) alongside a mean 2.22 (range; 0.19- 3.9) FP rates per image. No correlation between CAD sensitivity and false positive rates was identified. The findings suggest that further work is needed with larger sample sizes to improve confidence in synthesised findings. While future studies to evaluate CAD in the detection of PNs could be recommended, the recent research related to the higher potential effectiveness of Artificial Intelligence systems to support CXR interpretation suggests that this may no longer be an appropriate recommendation. Future research in either CAD or AI should explore and evaluate the risk versus benefit of computer-assisted technologies, as well as the impact on the imaging workforce and workflow. These technologies offer huge potential for diagnosis at an earlier stage, with a focus on saving more lives and improving the quality of life for those diagnosed with disease.

References

[1] Ost DE, Gould MK. Decision making in patients with pulmonary nodules. *American Journal of Respiratory and Critical Care Medicine* 2012;185(4):363–372

[2] Oda S, Awai K, Funama Y, Utsonomia D, Yanaga Y, Kawanaka K, et al. Effects of dual-energy subtraction chest radiography on detection of small pulmonary nodules with varying attenuation: receiver operating characteristic analysis using a phantom study. *Japan Journal of Radiology* 2010;28(3):214–219

[3] De Groot PM, Carter BW, Abbott GF, Wu CC. Pitfalls in chest radiographic interpretation: blind spots. *Seminars in Roentgenology* 2015;50(3):197-209

[4] Bradley SH, Abraham S, Callister EJ, Grice A, Hamilton WT, Rodrigeuz Lopez R, et al. Sensitivity of chest X-ray for detecting lung cancer in people presenting with symptoms: a systematic review. *British Journal of General Practice* 2019;69(689):e827-e835

[5] Waite S, Scott J, Gale B, Fuchs T, Kolla S, Redde D. Interpretive error in radiology. *American Journal of Roentgenology* 2016;208(4):739–749

[6] Fardanesh M, White C. Missed lung cancer on chest radiography and computed tomography. *Seminars in Ultrasound CT and MRI* 2012;33(4):280-287

[7] Greenhalgh R. (2016). *Missed lung cancers on chest radiographs.* [online]. Royal College of Radiologists. Last updated 24th October . https://www.rcr.ac.uk/audit/missed-lung-cancers-chest-radiographs

[8] Nover A, Jagtap S, Anjum W, Yegingil H, Shih W Y, Shih WH, et al. Modern breast cancer detection: a technological review. *International Journal of Biomedical Imaging* 2009;1-14

[9] Castellino RA. Computer aided detection (CAD): an overview. *Cancer Imaging* 2005;5(1):17–19

[10] Meziane M, Mazzone P, Novak E, Lieber ML, Lababede O, Phillips M, et al. A comparison of four versions of a computer-aided detection system for pulmonary nodules on chest radiographs. *Journal of Thoracic Imaging* 2012;27(1):58-64

[11] Kligerman S, Cai L, White CS. The effect of computer-aided detection on radiologist performance in the detection of lung cancers previously missed on a chest radiograph. *Journal of Thoracic Imaging* 2013;28(4):244-252

[12] de Hoop B, Schaefer-Prokop C, Gietema HA, de Jong PA, van Ginneken B, van Klaveren RJ, et al. Screening for lung cancer with digital chest radiography: sensitivity and number of secondary work-up CT examinations. *Radiology* 2010;255:629-637

[13] Amir G, Lehmann H (2016). After detection: the improved accuracy of lung cancer assessment using radiologic computer-aided diagnosis. *Academic Radiology* 2016;23:186–191

[14] Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated Triaging of Adult Chest Radiographs with Deep Artificial Neural Networks. *Radiology* 2019;291(1):196-202

[15] Care Quality Commission. *Radiology review: A national review of radiology reporting within the NHS in England.* July 2018. Available at https://www.cqc.org.uk/sites/default/files/20180718-radiology-reporting-review-report-final-for-web.pdf

[16] Pakdemirli E. Artificial intelligence in radiology: friend or foe? Where are we now and where are we heading?. *Acta Radiol Open* 2019;8:2058460119830222. Published 2019 Feb 21. doi:10.1177/2058460119830222

[17] Khan FA, Pande T, Tessema B, Song R, Benedetti A, Pai M, et al. Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy. *European Respiratory Journal* 2017;50: 1700953

[18] DUKE UNIVERSTY MEDICAL CENTRE (2017). *Systematic reviews: the process: types of reviews.* [online]. Last updated 17th July. http://guides.mclibrary.duke.edu/c.php?g=158155&p=1035849

[19] Khan KS, Kunz R, Kleijnen J. *Systematic Reviews to Support Evidence-Based Medicine.* 2nd ed. CRC Press, US;2011

[20] Bettany-Saltikov J. *How to do a systematic literature review in nursing: a step-by-step guide*. 1st ed. Maidenhead: Open University Press; 2012

[21] Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: *The PRISMA Statement*. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

[22] Centre for reviews and dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. 3rd ed. York Publishing Services Ltd; 2009

[23] Littell J. Evidence-based or biased? The quality of published reviews of evidence-based practices. *Children and Youth Services Review* 2008;30:1299-1317

[24] de Hoop B, De Boo DW, Gietema HA, van Hoorn F, Mearadji B, Schijf L, et al. Computer-aided detection of lung cancer on chest radiographs: effect on observer performance. *Radiology* 2010;257(2):532-540

[25] De Boo DW, Uffman M, Weber M, Bipat S, Boorsma EF, Scheerder MJ, et al. Computer-aided detection of small pulmonary nodules in chest radiographs: an observer study. *Academic Radiology* 2011;18(12):1507–1514

[26] De Boo D, van Hoorn F, van Schuppen J, Schijf L, Scheerder MJ, Freling NJ, et al. Observer training for computer-aided detection of pulmonary nodules in chest radiography. *European Radiology* 2012;22(8):1659–1664

[27] Li F, Engelmann R, Armato SG, MacMahon H. Computer-aided nodule detection system: results in an unselected series of consecutive chest radiographs. *Academic Radiology* 2015;22(4):475–480

[28] Samulski M, Snoeren PR, Platel B, van Ginneken B, Hogeweg L, Schaefer-Prokop C, et al. Computer-aided detection as a decision assistant in chest radiography. *Proceedings of SPIE* 2011;7966:796614-1-6

[29] Van Beek E, Mirsadree S, Murchison J. Lung cancer screening: computed tomography or chest radiographs? *World Journal of Radiology* 2015; 7(8):189-193

[30] Mazzone PJ, Obuchowski N, Phillips M, Risius B, Bazerbashi B, Meziane M. Lung cancer screening with computer aided detection chest radiography: design and results of a randomized, controlled trial. *PLOS ONE* 2013;8(3):e59650,1-7

[31] Moore W, Ripton- Snyder J, Wu G, Hendler C. Sensitivity and specificity of a CAD solution for lung nodule detection on chest radiograph with CTA correlation. *Journal of Digital Imaging* 2011;24(3):405-410

[32] Boland A, Cherry G, Dickson R. *Doing a systematic review a student's guide.* 2nd ed. SAGE Publications Ltd; 2017

[33] Whiting R, Westwood M, Deeks R, Leeflang S, Bossuyt PM, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36

[34] Vasaar M, Holzmann M. The retrospective chart review: important methodological considerations. *Journal of Educational Evaluation for Health Professions* 2013;10:1-7

[35] Meziane M, Obuchowski NA, Lababede O, Lieber ML, Philips M, Mazzone P. A Comparison of follow-up recommendations by chest radiologists, general radiologists, and pulmonologists using computer-aided detection to assess radiographs for actionable pulmonary nodules. *American Journal of Roentgenology* 2011;196(5):W542–W549

[36] Thiese MS. Observational and interventional study design types; an overview. *Biochemia Medica* 2014;24(2):199-2102013

[37] Tofthagen C. Threats to validity in retrospective studies. *Journal of the Advanced Practitioner in Oncology* 2012;3(3):181-3

[38] Frolkis C, Gilkeson R. Performance of a next generation computer-aided detection algorithm for detection of missed lung cancers on chest radiography. *Chest* 2014;146 (4):596A-596A

[39] Pannucci C, Wilkins E. Identifying and avoiding bias in research. *Plastic Reconstrive Surgery 2*010;26(2):619-625.

[40] Steckler A, McLeroy K.The importance of external validity. *American Journal of Public Health* 2008; 98 (1):9-10

[41] CLEVELAND CLINIC (2016). *Pulmonary nodules.* [online]. Last updated 25th March 2016. https://my.clevelandclinic.org/health/diseases/14799-pulmonary-nodules

[42] Parveen SS Kavitha C. A review on computer aided detection and diagnosis of lung cancer nodules. *International Journal of Computers & Technology* 2012;3(3):393-400

[43] Cancer Research UK 2016 https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/survival#ref-3

[44] Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reedle D. Interpreter Error in Radiology. *AJR* 2017;208:739–749

[45] Culpan G, Culpan AM, Docherty P, Denton E. Radiographer reporting: A literature review to support cancer workforce planning in England. *Radiography* 2019; 25:155-163.

[46] Woznitza N, Piper K, Burke S, Ellis S, Bothamley G. Agreement between expert thoracic radiologists and the chest radiograph reports provided by consultant radiologists and reporting radiographers in clinical practice: Review of a single clinical site. *Radiography* 2018; 24(3):234–239

[47] Bajre MK, Pennington M, Woznitza N, Beardmore C, Radhakrishnan M, Harris R, McCrone P. Expanding the role of radiographers in reporting suspected lung cancer: A cost-effectiveness analysis using a decision tree model. *Radiography* 2017; 23(4):273–278

[48] Mayo RC, Leung J. Artificial intelligence and deep learning – Radiology's next frontier? *Clinical Imaging* 2018; 49: 87–88

[49] Gossner J. Lung cancer screening-don't forget the chest radiograph. *World J Radiol* 2014; 6(9):116-118