PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Radiograph Accelerated Detection and Identification of Cancer in the Lung (RADICAL): A Mixed Methods study to assess the clinical effectiveness and acceptability of Qure.ai artificial intelligence software to prioritise chest X-ray (CXR) interpretation.
AUTHORS	Duncan, Sean; McConnachie, Alex; Blackwood, James; Stobo, David; Maclay, John; Wu, O; Germeni, Evi; Robert, Dennis; Bilgili, Banu; Kumar, Shamie; Hall, Mark; Lowe, David

VERSION 1 - REVIEW

REVIEWER NAME	Ueda, Daiju
REVIEWER AFFILIATION	Osaka City University, Diagnostic and Interventional Radiology
REVIEWER CONFLICT OF INTEREST	None
DATE REVIEW RETURNED	31-Oct-2023

GENERAL COMMENTS	Peer review for “Radiograph Accelerated Detection and Identification of Cancer in the Lung (RADICAL): A Mixed Methods study to assess the clinical effectiveness and acceptability of Qure.ai artificial intelligence software to prioritise chest X-ray (CXR) interpretation.” Title: The title is clear and concise, accurately reflecting the content and purpose of the study. Abstract: The abstract provides a comprehensive overview of the study. However, the specific hypotheses or research questions could be more clearly stated. Strengths and Limitations: This section is well-done, providing a balanced view of the study's potential strengths and limitations. Introduction: The introduction provides a clear and concise background, establishing the need for the study. The opportunity for AI is well presented and the objectives of the study are clearly stated. Methods and Analysis: The methods and analysis section is generally well written. However, there could be more detail regarding the statistical analysis. For example, it would be helpful to know what specific statistical tests will be used. Also, the section about the AI device data quality and availability could be more explicit about how the quality of the images will be assessed. Eligibility Criteria: The inclusion and exclusion criteria are clearly defined. Study Setting: The study setting is well described. However, it could

	be beneficial to provide some context about why these specific locations were chosen. Additionally, provide some clinical impact to use these locations. Study Intervention: The intervention is well described. However, it would be beneficial to provide more detail about how the software works and how it was developed. Technical Evaluation and Health Economic Evaluation: These sections are well written and provide important information about the study. However, more detail on the specific methods of these evaluations would be beneficial. Data Collection: The data collection process is well described. However, more information about the data management and quality control procedures would be beneficial. Ethics and Dissemination: This section is well written and provides important information about the ethical considerations and dissemination plan of the study. Overall, this is a well-written and thorough study protocol. However, additional details in certain sections would strengthen the manuscript.

REVIEWER NAME	Chauvie, Stéphane
REVIEWER AFFILIATION	Santa Croce e Carle Hospital
REVIEWER CONFLICT OF INTEREST	None
DATE REVIEW RETURNED	29-Dec-2023

GENERAL COMMENTS	pag 3 line 24 acronym is not defined for GP

REVIEWER NAME	Davis, Carolyn
REVIEWER AFFILIATION	Emory University
REVIEWER CONFLICT OF INTEREST	
DATE REVIEW RETURNED	06-Feb-2024

GENERAL COMMENTS	1. LDCT is used for lung cancer screening in high risk patients. It is established and accepted that chest x-ray is not recommended for lung cancer screening. Given the natural history and pathogenesis of lung cancer, can the authors address why they are proposing using chest xray as a screening tool and how do they account for those high-risk patients who underwent LDCT not CXR for screening? 2. CXR is a common imaging modality used to evaluate a range of chest concerns and risk factors for the development of lung cancer are well established. How do the authors account for this. 3. The authors state "Chest X-rays are the primary tool used to detect lung cancer [5] " based on weak evidence, can the authors address the basis of this claim? 4. Please clarify each group in more details including all risk factors and relevant patient factors established for risk assessment for lung cancer screening.

	5. Why do the authors focus on time from cxr acquisition to reporting? What is the goal of this and what evidence supports the authors chosen time frame? Why is this time frame different and shorter than even surveillance or follow-up imaging? 6. Authors mentioned that 2 weeks was considered a delay in reporting-why and what evidence is there that this changes or impacts outcomes? 7. Can the authors address the use of chest xray vs LDCT or CTDI for lung cancer screening and provide evidence to support their recommendation to use chest xray for lung cancer screening. 8. Can the authors explain patients who underwent biopsy or surgery vs CT. 9. How does this study propose to address diagnosing the various types of lung cancers (NSCLC vs SCLC etc), cancer of the chest (thymomma, thymic carcinoma etc) vs metastasis vs noncancerous conditions 10. How does this account for false positive rates and the effects on a patient a false positive screening test could potentially have? 11. Why have the authors decided to include all patients that are 18 years and older given the pathogenesis of lung cancer? 12. Please explain how the authors plan to stratify and identify people at high risk for developing the disease, as this is an essential goal and priority of any screening protocol. 13. What is the authors goals for earlier lung cancer screening? How do the authors address patients who are both candidates and non-candidates for curative intent therapy? 14. Why are the authors including potentially individuals at low to no risk for screening? 15. How are the authors accounting for patients under lung cancer surveillance? 16. How do the authors account for lead-time bias? 17. Can the authors explain how this will improve or change outcomes in lung cancers that may have a different natural history (radiographic vs clinically detected cancers). 18. How do the authors address the issue of overdiagnosis with this proposed model. 19. What about prior comparative images (an important part of lung cancer diagnosis) 20. Can the authors further explain their decision on selecting the 6 findings on cxr and why they state these 6 are findings to identify "malignancy" and the evidence to support this. How do they account for these findings in non-malignant states and how do they account for potential false negatives?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Daiju Ueda, Osaka City University
Comments to the Author:
Peer review for “Radiograph Accelerated Detection and Identification of Cancer in the Lung (RADICAL): A Mixed Methods study to assess the clinical effectiveness and acceptability of Qure.ai artificial intelligence software to prioritise chest X-ray (CXR) interpretation.”

Title: The title is clear and concise, accurately reflecting the content and purpose of the study.
Thank you. We aimed to make the title concise but also provide a basic overview of the study.

Abstract: The abstract provides a comprehensive overview of the study. However, the specific hypotheses or

research questions could be more clearly stated.
Thank you for this feedback. We have altered the introduction to specifically state the research question.

Strengths and Limitations: This section is well-done, providing a balanced view of the study's potential strengths and limitations.
Thank you for your comments.

Introduction: The introduction provides a clear and concise background, establishing the need for the study. The opportunity for AI is well presented and the objectives of the study are clearly stated.
Thank you for this positive feedback.

Methods and Analysis: The methods and analysis section is generally well written. However, there could be more detail regarding the statistical analysis. For example, it would be helpful to know what specific statistical tests will be used. Also, the section about the AI device data quality and availability could be more explicit about how the quality of the images will be assessed.
Thank you for this comment. This has prompted us to update our statistics section to include specific statistical tests and also more in depth details around image quality.

Eligibility Criteria: The inclusion and exclusion criteria are clearly defined.
Thank you.

Study Setting: The study setting is well described. However, it could be beneficial to provide some context about why these specific locations were chosen. Additionally, provide some clinical impact to use these locations.
Thank you. The three clusters comprise all of the hospitals within NHS Greater Glasgow and Clyde that have outpatient chest XR imaging. Your feedback has highlighted that the wording was not clear on this point and we have altered it accordingly.

Study Intervention: The intervention is well described. However, it would be beneficial to provide more detail about how the software works and how it was developed.
Thank you, we have provided some supplementary material covering additional details about software features and development.

Technical Evaluation and Health Economic Evaluation: These sections are well written and provide important information about the study. However, more detail on the specific methods of these evaluations would be beneficial.
Thank you for your feedback. We have updated the document with further details on these points in the Health Economic Evaluation section.

Data Collection: The data collection process is well described. However, more information about the data management and quality control procedures would be beneficial.
Thank you for this comment. We will include data flow diagrams as supplementary material in response to your feedback.

Ethics and Dissemination: This section is well written and provides important information about the ethical considerations and dissemination plan of the study.
Thank you for your feedback.

Overall, this is a well-written and thorough study protocol. However, additional details in certain sections would strengthen the manuscript.
Our team are grateful for the time you have put into reviewing this document. Your comments have been enlightening and have resulted in changes to our document which we believe have strengthened it.

Reviewer: 2 Dr. Stéphane Chauvie, Santa Croce e Carle Hospital Comments to the Author:

pag 3 line 24 acronym is not defined for GP
Thank you for your feedback and for your review. We have replaced this with Primary Care clinician, which we agree is more universal.

Reviewer: 3

Dr. Carolyn Davis, Emory University

Comments to the Author:

1. LDCT is used for lung cancer screening in high risk patients. It is established and accepted that chest x-ray is not recommended for lung cancer screening. Given the natural history and pathogenesis of lung cancer, can the authors address why they are proposing using chest xray as a screening tool and how do they account for those high-risk patients who underwent LDCT not CXR for screening?
Thank you for your comment. Unfortunately, the UK does not have widely available access to LDCT through the NHS. Subsequently, CXR is still the most widely used screening tool for initiating investigation of lung cancer. Here is a link to our national institutional guidelines on this matter. Recommendations organised by site of cancer | Suspected cancer: recognition and referral | Guidance | NICE. Please also note that we are not stratifying patients into risk categories, all outpatient and GP x-rays will have a first read by the AI software, regardless of CXR indication. We have now emphasised this in our introduction to avoid confusion

2. CXR is a common imaging modality used to evaluate a range of chest concerns and risk factors for the development of lung cancer are well established. How do the authors account for this.
Thank you for your comment. This intervention covers all outpatient chest x-ray activity rather than focusing on any specific symptomatology. The software will interpret all outpatient and GP requested images, highlighting suspicious images for early reporting by a radiologist. As you have highlighted, there will be other chest pathology on these images that the software will pick up on but not highlight for early reporting. Non-USC pathology will be given a routine urgency reporting priority code. We have now altered the following sections to add clarity: ‘The pressure on Radiology departments’, ‘Study intervention- qXR'.

3. The authors state "Chest X-rays are the primary tool used to detect lung cancer [5] " based on weak evidence, can the authors address the basis of this claim?
Thank you for highlighting this. It would have been more accurate to provide a UK context to this claim and we have updated our manuscript accordingly in the section ‘The pressure on Radiology departments’

4. Please clarify each group in more details including all risk factors and relevant patient factors established for risk assessment for lung cancer screening.
Thank you for your comment. This is a protocol manuscript which is submitted for a study that has not completed. We would like to clarify that these patients have not been selected specifically for lung cancer screening, the AI intervention is interpreting all outpatient chest x-ray images regardless of reason for referral. Once the data is collected, we will be able to describe broad demographic factors but these patients have not been selected specifically for inclusion.
Our inclusion criteria for the prospective component as described in the manuscript:
Over 18 years old
Frontal radiograph collected through usual care via outpatient pathway.

5. Why do the authors focus on time from cxr acquisition to reporting? What is the goal of this and what evidence supports the authors chosen time frame? Why is this time frame different and shorter than even surveillance or follow-up imaging?
Thank you for your question. This time interval allows us to demonstrate the impact that the AI software has on the reporting time of a chest x-ray. This is often referred to as a 'Turn Around Time' and is a metric that radiology support interventions use as an endpoint. There is a shortfall of radiologists in the UK and worldwide relative to workload; this has resulted in a longer than ideal report times for x-rays leading to delays in further investigation and treatment. We hope to demonstrate that this software can assist in prioritising reporting of suspicious x-rays to expedite patient care . This is not a measurement of a follow up or surveillance period, this time period only covers the time elapsing between an x-ray being taken and that same x-ray being reported by the radiologist.

6. Authors mentioned that 2 weeks was considered a delay in reporting-why and what evidence is there that this changes or impacts outcomes?
Thank you for this important question. At the time of writing, UK national targets on cancer waiting times stipulate that a CXR must be acquired and reported within 2 weeks. In our introduction, we have referenced a recent large cohort study in Taiwan that has described the link between delays to treatment and patient outcomes. Please see below.
Tsai C-H, Kung PT, Kuo W-Y, et al. Effect of time interval from diagnosis to treatment for nonsmall cell lung cancer on survival: a national cohort study in Taiwan. BMJ Open 2020;10:e034351. doi:10.1136/ bmjopen-2019-034351

7. Can the authors address the use of chest xray vs LDCT or CTDI for lung cancer screening and provide evidence to support their recommendation to use chest xray for lung cancer screening.
Thank you for highlighting this. We would like to clarify that the intervention is not a recommendation for a lung screening pathway, it is a study into technology implementation to enhance reporting times in an existing pathway that is consistent with national guidelines within the UK and other countries. At a national level, the decision to use CXR rather than LDCT is based on resource constraint rather than clinical outcomes and remains under review, CXR will remain an important tool for identification of suspicion of cancer even with introduction of LDCT in both inpatient and outpatient settings. At the time of writing, the NHS and many other countries are unable to provide a LDCT resource as the primary tool for lung cancer screening.

8. Can the authors explain patients who underwent biopsy or surgery vs CT.
Thank you for your comment. The submitted manuscript is a protocol for a study, we have not formally generated data or patient outcomes. As part of our secondary outcomes we will report on the time to CT and also time to treatment.

9. How does this study propose to address diagnosing the various types of lung cancers (NSCLC vs SCLC etc), cancer of the chest (thymomma, thymic carcinoma etc) vs metastasis vs noncancerous conditions
Thank you for this comment. The AI software is trained and CE marked for CXR abnormalities at a radiological level rather than any histopathological endpoints. The AI is constrained to commenting on radiological features that it has been trained to identify such as 'opacity', 'nodule' and 'mediastinal widening'. Subsequently, it may misinterpret noncancerous findings as suspicious for cancer. We will not be reporting on subtypes of thoracic malignancy specifically. We will report on performance of the AI to detect cancer as described in the manuscript.

10. How does this account for false positive rates and the effects on a patient a false positive screening test could potentially have?
Thank you for this important question. All x-rays in this study will be reviewed by a radiologist, the effect of the intervention is to re-prioritise the reporting work pile with suspicious x-rays being placed at the top of this pile. If the AI flags a patient as positive for USC but the radiologist disagrees (a false positive) then the only impact on

the patient would be that they would have an earlier radiology report. Those that the radiologist disagrees as per figure X will be additional ground truthed by 2 radiologist and when necessary adjudicated.

11. Why have the authors decided to include all patients that are 18 years and older given the pathogenesis of lung cancer?
Thank you for this question. The AI is reviewing all outpatient CXRs to pull out features of USC. The study is evaluating a service level deployment of an AI solution in the general adult population in Glasgow. The software is CE Marked for use in an adult population and regulatory approval is for its use above 18 yearss. We will measure the performance of the AI in this context, this would include the performance of the device in different age groups. If the device performs well, then we hope that it would not have a high false positive rate in a very low risk age group. If it performs poorly then we can report this limitation.

12. Please explain how the authors plan to stratify and identify people at high risk for developing the disease, as this is an essential goal and priority of any screening protocol.
Thank you for this question. We hope to clarify that we are not making any changes to whether patients undergo chest x-ray or not. The AI will act as a rapid triage of x-rays that are being taken for routine care regardless of the indication of this test. For example, the software will analyse an x-ray for pneumothorax follow up in the same way as it would examine a specific USC referral. We are hoping to reduce a delay in the existing diagnostic pathway rather than alter patient selection for lung cancer screening.

13. What is the authors goals for earlier lung cancer screening? How do the authors address patients who are both candidates and non-candidates for curative intent therapy?
Thank you. The goal of this study is to evaluate the effectiveness of a CE marked AI solution at reducing delays in x-ray reporting. This will hopefully reduce the overall time from CXR to formal diagnosis and treatment of lung cancer. The software is unable to indicate whether patients are candidates for radical treatment or not from chest x-ray alone. This will be decided at multidisciplinary team meetings as per existing standard of care.

14. Why are the authors including potentially individuals at low to no risk for screening?
Thank you for this question. We are not selecting patients for x-ray screening however we are letting the AI read all outpatient chest x-rays including those who may be at low risk of cancer. We will record the performance of the software in this setting, as you suggest it may perform poorly in low risk groups and we would wish to report this as part of our outcome. To highlight this is not a screening intervention but use of AI to support prioritisation (detection) of images containing features associated with urgent suspicion of cancer. We agree with reviewer that a translation to LDCT screening will be of benefit to population.

15. How are the authors accounting for patients under lung cancer surveillance?
Thank you for this excellent question. If a patient is already under lung cancer follow up then any X-ray they have is likely to flag as USC. This is because the AI is only able to interpret pixel level, it is unable to contextualise this x-ray into prior imaging or patient history. Importantly, all x-rays in this study will be reported by a radiologist. The above example would be later over-ridden by a radiologist who is able to contextualise the x-ray. Within the study these images will be coded to reflect performance of the AI to detect cancer.

16. How do the authors account for lead-time bias?
Thank you for this question. We have understood by this question that you are concerned about whether we may demonstrate improved survival time through this intervention by virtue of simply diagnosing lung cancer earlier rather than genuinely improving survival. Delays in treatment of lung cancer are associated with worse outcomes, therefore it would seem reasonable to infer that earlier diagnosis will lead to a genuine improvement in survival. Generally, earlier stage lung cancer is more amenable to radical treatment. There may be a lead time bias for patients who are too frail at baseline for any form of treatment, earlier diagnosis in these patients would not affect treatment options. We hope this answers your question but would be happy to discuss further if

we have missed your intention.

17. Can the authors explain how this will improve or change outcomes in lung cancers that may have a different natural history (radiographic vs clinically detected cancers).
Thank you for this question. As described, this is a general intervention that is aimed at tackling a specific delay in an existing pathway. We hope to reduce the time-period between a suspicious x-ray being acquired and reported by a radiologist through triage of x-rays flagged as USC by the AI software. To run the pathway theoretically- In the group of lung cancer patients who present with a detectable chest x-ray abnormality, they will receive a CT scan quicker than previously as the radiologist will see the x-ray earlier and request the CT. An earlier CT will then lead to earlier referral to MDT meetings which will mean an earlier treatment decision and subsequently earlier treatment. Earlier treatment is linked to improved survival outcome as demonstrated by the referenced study in Taiwan alongside expert consensus. If a lung cancer patient has an entirely normal CXR then they will not derive the benefits of this intervention.

18. How do the authors address the issue of overdiagnosis with this proposed model.
Thank you for this important point. The software may result in false positive USC flags but all images are subsequently reported by a radiologist. We will be collecting CT linked data to assess whether this intervention results in an increased volume of CT scans- an important consideration for our health economics model. In a perfect scenario, the intervention should only reduce time intervals rather than change the volume of CT scans. However, we need to be mindful of radiologist behaviours changing in response to the intervention itself. The study design of stepped intervention will enable us to look at trends of requesting during the study and we are monitoring closely CT request rates and its association with AI flags.

19. What about prior comparative images (an important part of lung cancer diagnosis)
Thank you for pointing out this limitation in AI radiology software. At the time of writing, models such as qXR are unable to perform comparative analysis against prior imaging nor place the image within its clinical context; this is a fixed source of repeated false positive flagging by the AI. As described, this is mitigated by the 'human in the loop', a radiologist, who will be able to provide this comparison and over-ride any false positive flags to prevent them progressing to further investigations.

20. Can the authors further explain their decision on selecting the 6 findings on cxr and why they state these 6 are findings to identify "malignancy" and the evidence to support this. How do they account for these findings in non-malignant states and how do they account for potential false negatives?
This is an excellent question, thank you. QXR has a selection of in-built modules for 25 different CXR abnormalities. We have 2 senior radiologists on the study team who have advised on the selection of these 6 findings. As you have correctly identified, each abnormality is also associated with non-malignant findings, the effect of this being a possibility of false positive flagging. USC is a nuanced reporting term which is not amenable to deconstructing precisely into qXR features; hence the selection of these 6 features is a result of radiologist consensus supported by calibration exercises. Many of these problems are also apparent in cross-sectional imaging (albeit at a lower rate)- whereby an identified nodule or pleural effusion may be malignant or non-malignant. False negatives in this study are defined as when the AI reports an x-ray as non-USC and a radiologist subsequently report that this does in fact demonstrate USC. We are reporting these as a secondary outcome connected to safety of intervention.

VERSION 2 – REVIEW

REVIEWER NAME	Ueda, Daiju
REVIEWER AFFILIATION	Osaka City University, Diagnostic and Interventional Radiology
REVIEWER CONFLICT OF INTEREST	None
DATE REVIEW RETURNED	03-Jun-2024

GENERAL COMMENTS	Overall, the revised protocol has been strengthened by the additions and clarifications made in response to the initial comments. The specific research question, statistical methods, study setting context, intervention details, and health economic evaluation methods are now more clearly presented. In my assessment, the authors have comprehensively addressed the points raised in the first review, resulting in an improved manuscript. I have no additional comments or suggestions for further changes at this time.

VERSION 2 – AUTHOR RESPONSE