UK National Screening Committee

UK National Screening Committee (UK NSC)

Evidence summary on the use of artificial intelligence for mammographic image analysis in breast cancer screening

Date: 04 November 2021

Contents

Aim

To summarise the evidence summary and the responses received during the three-month public consultation on the use of artificial intelligence (AI) for mammographic image analysis in breast cancer screening, and to ask the UK NSC to approve that the evidence summary, consultation comments and responses are ready to be published on the UK NSC website.

Current Recommendation

The UK NSC recommends screening for breast cancer. National screening programmes are in place in each of the four countries of the UK.

No prior review has been conducted on the use of AI for mammographic image analysis in breast cancer screening by the UK NSC.

Evidence Summary

The 2020 evidence summary was an exploratory review rather than a recommendation making review. The purpose of this evidence summary was to prepare the UK NSC for the receipt of a proposal to modify the Breast Screening Programme in the UK by implementing AI for mammographic image analysis. The methods in this evidence summary may be used as a baseline to build upon in the future review.

The 2020 evidence summary was undertaken by the University of Warwick, in accordance with the triennial review process: https://www.gov.uk/government/publications/uk-nsc-evidence-review-process/uk-nsc-evidence-review-process

The 2020 evidence summary covered relevant literature since 2012 and addressed 2 key questions:

1. What is the accuracy of AI algorithms to detect breast cancer in women attending screening mammography?

2. What is the clinical impact of the use of AI algorithms to detect breast cancer in mammograms compared to current practice in breast screening programmes?

The conclusion of the 2020 evidence summary is that, based on the current evidence, the UK NSC does not recommend using AI in the Breast Screening Programme. This is because:

- the intervention of AI in the breast screening pathway for image analysis has the potential to overall improve or worsen the current breast screening programme as it can have unintended consequences

- AI might reduce the workload of staff, the number of cancers missed at screening, and the number of women called back for further tests when they do not have cancer, however, the quality of evidence is very low. AI could also increase the workload of staff, and/or the numbers called back, and/or reduce the numbers of cancers detected

- therefore, we will need strong evidence on how accurate AI is in breast screening clinical practice as well as its effect on outcomes in the whole pathway before changing it

- the performance of AI systems varies in different settings but there are no good quality studies in the UK

- it is not clear how good AI is at finding different types of breast cancer or at finding breast cancers in different groups of women (for example, different ethnic groups)

- based on the current evidence, we do not know how human readers will behave when interacting with the AI and what are the outcomes from the whole pathway as there are no prospective studies comparing a breast screening pathway integrating AI versus the pathway in current practice

The evidence summary recommended that a review in 1-3 years' time may be necessary as the evidence base is expected to develop in the next few years.

Consultation

A three-month consultation was hosted on the UK NSC website, which closed on 13 August 2021. Direct emails were sent to 58 stakeholders (please note that multiple individuals from the same organisation were invited, Appendix A)

Comments were received from the following 8 stakeholders:

1. National Co-ordinating Committee for Breast Pathology
2. The Royal Marsden
3. Hologic
4. The Chartered Institute for IT
5. The Royal College of Pathologists' Digital Pathology Committee
6. Kheiron Medical Technologies
7. Gloucester hospitals NHS FT
8. Royal Society of Biology

The consultation comments are presented below in Appendix B.

Overall, two stakeholders agreed with the conclusions of the UK NSC review, and remaining stakeholders did not provide a direct statement. Stakeholders were in agreement with the majority of the methodological considerations covered in the evidence summary and noted that they can be used as baseline. Any disagreements are discussed below. Also, similar methodological issues relating to published studies were highlighted in the evidence summary and by stakeholders.

Several key themes emerged from this consultation: new evidence, methodological considerations, national test sets, use-case of AI in the mammogram reading pathway, representativeness of the population, harms to environment.

New evidence

Currently, this topic is an active research area. Stakeholders brought to the UK NSC's attention one study by Sharma et al., 2021[1].

Response: the study by Sharma et al., 2021 was published after the search date of this review therefore was not included. The committee decided to not extend the

search dates because this study has not been published in a peer-review journal yet and is only available as a pre-print. The study was informally examined by the reviewers and they concluded that it was a two-part study: a retrospective test accuracy study and simulation study and the evidence would not have altered the conclusions of the evidence summary. Also, reviewers were concerned that assumptions in a simulation study were inappropriate. A short summary of this study has been added to the discussion of the report.

References

1 Sharma N, Ng AY, James JJ, et al. Large-scale evaluation of an AI system as an independent reader for double reading in breast cancer screening. 2021:2021.02.26.21252537. doi: 10.1101/2021.02.26.21252537 %J medRxiv

Methodological considerations

Some stakeholders suggested that in the evidence summary, the considerations around temporal/ geographical validation and differential verification bias in retrospective studies should be reconsidered. Specific to the types of validation, the UK NSC evidence summary suggested that geographical validation is the preferred method, however, some stakeholders disagreed. They suggested that both types of validation (temporal/ geographical) should be included in the future review as geographical validation does not completely eliminate the risk of the same women appearing in both datasets (i.e. training and test tests) as people tend to relocate. Specifically to differential verification bias, the UK NSC review noted that retrospective test accuracy studies suffer from differential verification bias more than prospective test accuracy studies, however, stakeholders disagreed. They presented the opinion that retrospective test accuracy studies can effectively mitigate differential verification bias with a sufficiently long follow-up.

Response: the committee noted that in general, due to deidentification, it may be difficult to ensure that the same women are not being included in the training and test sets. While geographical validation within the same country cannot ensure that there would not be an overlap between training and test sets, it may be lower than with temporal validation due to repeat screens. Also, images from the same screening sites are less likely to capture variations in image acquisition parameters and personnel. The review described the additional issues with temporal validation associated with the same machines and readers.

The committee noted that in retrospective studies differential verification bias cannot be eliminated. Differential verification involves the use of different reference standards to verify positive and negative index test results. Bias usually arises because one of the reference standards is less reliable than the other. In the case of breast screening studies follow up to clinical presentation is less reliable than triple assessment which may include biopsy.

In retrospective studies AI positive / test set negative women cannot be characterised (for example false positive / true positive, clinically significant or overdiagnosed, stage at time of screening) through assessment and if indicated biopsy. Thus, while test sets can incorporate clinically presenting cases into the data over time it is still prone to uncertainty.

Differential verification bias cannot be completely eliminated from prospective studies either because screen negative women cannot undergo a biopsy. However, in prospective studies this source of bias can be reduced. This is because women who are screen positive in both the AI pathway and the standard screening pathway can be recalled for assessment which may include biopsy. Thus, while differential verification bias is common to both retrospective and prospective studies, the problem can be mitigated in prospective studies.

National test set

Stakeholders agreed that AI algorithms are 'short lived' and noted that developers will continue improving their AI algorithms therefore it is important to have mechanisms in place that would allow monitoring the performance of AI algorithms post deployment. They also noted that this mandates establishing a national dataset that could be used to test AI algorithms.

Response: the committee supported the idea of establishing a national dataset, approved by regulators, however, they acknowledged that a particular test set can only be used a limited number of times on different versions of the same AI system. This is because running multiple versions of the same AI system with different hyperparameters and then testing them all against the same test set, and retrospectively choosing the parameters that achieve the highest diagnostic accuracy may inflate accuracy compared to real-world practice.

Use-case of AI in the mammogram reading pathway

Stakeholders provided suggestions on the implementation strategies of AI within the mammogram reading pathway. A stakeholder suggested that the committee should consider the approach of gradually introducing AI in the mammogram reading pathway i.e. firstly using AI as a pre-screening tool followed by the use of AI as a second reader, once the required evidence becomes available.

A stakeholder noted that in addition to the implementation strategies listed in the evidence summary, the fourth option should be added, where AI is being used after manual grading to pick up cases that were missed by human readers.

Response: the committee noted that the gradual introduction approach is unlikely to be feasible because each use-case of AI is associated with different challenges that should be addressed in studies prior implementation to ensure that the use of AI does more good than harm at a reasonable cost.

The reviewers noted that this evidence summary did not intend to provide an exhaustive list of potential implementation strategies and only focused on the options that were commonly mentioned in the literature. They noted that the implementation strategy, where AI is being used after manual grading, was used in one simulation study and included in the evidence summary.

Representativeness of the population

Stakeholders were concerned that none of the studies include transgender men and cisgender men.

Response: it was noted by the Breast Screening Programme manager that only people who are registered as women or with indeterminate sex are invited to attend the Breast Screening Programme. Transgender men are invited for screening if they are still registered as a woman with their GP and if they still have breast tissue (more detailed information is available here: https://www.gov.uk/government/publications/nhs-population-screening-information-for-transgender-people/nhs-population-screening-information-for-trans-people#breast-screening). However, the Breast Screening Programme does not record such information. The reviewers investigated included studies further and found that studies use the term 'women' (except, two studies where terms 'patients'/ 'cases' were used). Also, all mammograms came from routine screening centres. None of the studies reported that they excluded men and / or transgender people but only one study sampled consecutively. Given this, the committee agreed that transgender people would not form a subgroup of non-standard images that requires further research and consideration for AI.

Harms to environment

Stakeholders suggested that the development and use of AI should take into account harms to the environment.

Response: the committee acknowledged that this is an important issue. However, it was noted that normally in the cost-benefit analysis, the UK NSC looks at the benefits and harms of screening from healthcare and social care perspective.

Action

The Committee is asked to approve that the review, consultation comments and responses are published on the website.

Appendix A: List of Organisations Contacted

1. Breast Cancer Care
2. Breast Cancer Now
3. British Association of Surgical Oncology
4. Cancer Research UK
5. Faculty of Public Health
6. Macmillan
7. Northern Ireland Cancer Network
8. Royal College of General Practitioners
9. Royal College of Nursing
10. Royal College of Pathologists
11. Royal College of Physicians
12. Royal College of Physicians and Surgeons of Glasgow
13. Royal College of Physicians of Edinburgh
14. Royal College of Radiologists
15. Royal College of Surgeons
16. Royal College of Surgeons of Edinburgh
17. Society and College of Radiographers
18. The British Association for Cancer Research
19. NHS
20. Accelerated Access Collaborative
21. Faculty.ai
22. Google Health
23. IBM

24. Imperial College London
25. Kheiron Medical Technologies
26. NICE
27. Optos
28. Researcher with interest in AI (Queen Marry University of London)
29. Researcher with interest in AI (University of Manchester)
30. Northgateps

UK National Screening Committee

Appendix B: Consultation Responses

Note: Personally identifiable information has been redacted from certain comments, where individuals have chosen not to have personal details made public

Use of artificial intelligence for image analysis in breast cancer screening

Consultation comments

1. **National Co-ordinating Committee for Breast Pathology**

Name:	Sarah Pinder	**Email address:**	xxxx xxxx
Organisation (if appropriate):	National Co-ordinating Committee for Breast Pathology		
Role:			
Do you consent to your name being published on the UK NSC website alongside your response? **Yes**			

Section and / or page number	Text or issue to which comments relate	Comment *Please use a new row for each comment and add extra rows as required.*

1	**Title = Use of artificial intelligence for image analysis in breast cancer screening**	The title (and areas in the main document) imply that artificial intelligence for analysis of images is only relevant to mammography. In the future similar reviews for other imaging modalities and histopathology are likely to be required. As a small (and pedantic) point, perhaps the title could make it clearer that this particular report relates specifically to the present evaluation of AI for mammographic images?

UK National Screening Committee

2. The Royal Mardsen

Name:	Richard Sidebottom	**Email address:**	xxxx xxxx
Organisation (if appropriate):	The Royal Marsden		
Role:	**Consultant radiologist, artificial intelligence imaging hub**		

Do you consent to your name being published on the UK NSC website alongside your response?

Yes **No**

Section and / or page number	Text or issue to which comments relate	Comment *Please use a new row for each comment and add extra rows as required.*
Page 6,8,10,18,34, 67,68	Only accepting geographic validation not temporal validation testing should be reconsidered.	We think it is vital that whilst the report discusses temporal and geographic validation, the most important factor is that data from an individual woman included in training should not be used in testing. Whilst geographic validation will ensure this the vast majority of the time there will undoubtedly be instances where women have moved and could appear in the

		training and test sets if this is controlled only by geographic validation. Temporal validation using the same site as the training set should be considered providing that data from individual women included in training are not be used in testing (for example if she has attended for a different screening attendance). In fact depending on the methods of deidentification used, it may be possible to be more confident that an individual is not present in both the training and testing data if temporal validation from the same site is used. It would be reassuring to see examples of both types of testing used. This is discussed to some extent on page 67,68.
Page 38-40, 70-72	Incorporation bias and differential verification bias. We think that some of these studies mitigate this effectively.	The 'gatekeeper effect' of only finding screen detected cancers is mitigated by several of these studies by using a longer period of follow up to confirm that a case is normal. The length of time this needs to be is debateable. Some excluded studies using UK data have defined that if a cancer presents within 3 years as an interval cancer or at the next screening round that index mammogram is not considered normal. Probably if a cancer is overlooked that is so slowly progressive as to

		not present within this time then it is of a reduced biological significance. Indeed the use of retrospective data allows for this determination of more reliable ground truth and offers an advantage over prospective studies which might not capture this unless they employ sufficiently long follow up (which should also be done eventually). This is discussed further on Page 70-72.
	Prospective vs retrospective.	We agree that prospective studies are required because these should be able to demonstrate the real impact of the intervention on the workflow. We think that retrospective studies have greater value in demonstrating system performance in a stand alone capacity. This is because large volumes of data and the determination of a more reliable ground truth for normal cases as discussed above. We agree with the comment about laboratory reader studies. They probably have a role in early work but we are sceptical about the value and applicability of laboratory type reader studies where enriched datasets are used with readers compared to AI outside of a normal clinical reading setting.

Page 73.	‘Algorithms are short lived’. Therefore we think that large scale testing and monitoring is needed.	We agree. This observation suggests we need some new ways of ensuring safety. As this type of technology rapidly improves, there is currently ample opportunity for AI developers to refine and improve (change) these algorithms. We suggest that large national datasets should be used to benchmark performance on retrospective data. In addition active monitoring of AI system performance and overall screening system performance is feasible to ensure safety is maintained with prospective deployment.
Other	Technical recalls, non standard images / patients. Use outside population screening.	There are mundane but important issues which have not been addressed in the report or the studies such as handling technically inadequate images (technical recalls). More than the standard 4 mammography images (larger breast sizes). Post treatment images. Implants. We will ultimately require evidence on these issues. Hopefully these factors will be considered when prospective studies appear. Use in symptomatic clinics has not been addressed in this review and is perhaps outside the scope of this review by the screening committee however use in the symptomatic clinic may be beneficial to patients and

		evidence of this, particularly in a UK setting would be valuable to appraise if available in future.

UK National Screening Committee

3. Hologic

Executive Summary

Hologic recognises that, currently, lack of UK data prevents the immediate application of artificial intelligence (AI) within breast cancer screening.

However, the considerable pressures faced by the breast cancer screening programme, most significantly an acute shortage of radiologists and mammographers, means that we cannot rely on 'business as usual', and must find ways of unlocking the benefits of AI in a safe way.

In light of this, we urge the Committee to adopt an approach to AI that is centred on establishing **how incremental benefits can be achieved by the gradual introduction of these technologies**. The Committee should work with industry and academia to assess how elements of AI can be safely introduced in their current form, for specific tasks, and how a pathway to wider adoption can be established in the longer term.

Already there are clear indications of AI's potential to support clinicians. The area in which there is most near-term potential and fewer obstacles to adoption, is with regard to **AI assisting in the 'triage' of breast cancer screening patients**. As an

adjunct to standard mammography, AI based software solutions could be utilised to initially review a mammogram, a risk assessment could be made and an indication given to radiologists about which patients may warrant further investigation and which patients can be returned to the normal screening programme.

In the longer term, as AI technology matures, **AI guided imaging** could remove the need for a second reader to review images during breast cancer screening, increasing the UK's breast screening capacity when combined with additional screening resources.

In parallel to establishing a roadmap for the safe and efficient introduction of AI to support screening, all stakeholders in this process must consider how this can be achieved whilst gaining and maintaining clinician and patient confidence.

Introduction

Hologic recognises that, currently, lack of UK data prevents the immediate application of artificial intelligence (AI) within breast cancer screening. However, there are also indications of the technology's potential to help clinicians detect disease quicker and more accurately. These early indications are compelling enough to warrant further exploration, and a recalibration from 'does AI work now?' to 'how can we unlock the benefits AI will bring?'.

We urge the Committee to adopt an approach to AI that is centred on establishing how incremental benefits can be achieved by the gradual introduction of these technologies. The Committee should work with industry and academia to assess how elements of AI

can be safely introduced in their current form, for specific tasks, and how a pathway to wider adoption can be established in the longer term.

The imperative for action on AI

The draft report has been published at a pivotal point for breast cancer screening. The COVID-19 pandemic disrupted the programme, leading to a backlog of patients waiting for an essential preventative service. The health service is also experiencing a significant shortage in radiologists and mammographers.

Against this backdrop, there is an urgent need to embrace advanced technologies to help clinicians tackle immediate challenges and implement longer term strategies.

Hologic recognises that AI is not yet capable of taking on a clinician's role in breast cancer screening. Its near-term application is more about helping clinicians decide where their finite amount of time is best spent. Compelling evidence suggests that AI could expediate the scan reading process and assist in triaging patients, helping to ensure those most at risk are given the greatest focus.

Hologic encourages the Committee to focus its attention on these near-term applications, the incremental benefits they can bring, and what will be required from industry, academia and the health service to bring them to fruition safely, quickly and cost effectively. The Committee should also assess how an environment can best be created in which innovation, such as AI, can be easily adopted once demonstrated to be of clinical value.

In the remainder of this submission, we highlight a key area in which AI could bring significant improvements to the breast cancer screening process, for both clinicians and patients, in the near-term.

Helping clinicians prioritise their time

Currently, every image captured during routine breast screening is reviewed by two independent readers. This established practice is resource intensive – in terms of the number of personnel required and the costs associated with this.

As an adjunct to standard mammography, AI based software solutions could be utilised to initially review a mammogram, a risk assessment could be made and an indication given to radiologists about which patients may warrant further investigation and which patients can be returned to the normal screening programme.

There is also the potential to use AI in conjunction with other advanced technologies to even greater effect in this regard. Tomosynthesis – more commonly known as 3D Mammography – detects up to 65% more invasive breast cancers, and reduces patient recalls by up to 40%, when compared to traditional 2D mammography alone.[1,2]

AI can be used to analyse tomosynthesis images and highlight areas of interest to clinicians, directing their attention to where it is most needed. Based on a comparison with the average time taken to read an image without AI, a time saving of up to 13% may be achieved using this technology,[3] improving the efficiency with which images are reviewed.

In these examples, all mammograms would still be reviewed by two radiologists, but the time spent reading images would be reduced and potentially a more robust assessment of the potential risk of developing breast cancer for any given patient, could be obtained, opening up the possibility for more personalised screening and treatment programmes.

In the long-term, and as more evidence becomes available, the use of AI to review screening mammograms could mean that only one reader would be required, effectively freeing up the time and reducing the costs associated with the second reader. This technology has the potential to significantly improve existing breast screening capabilities and capacity in the UK.

Building confidence in AI

[1] Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA*. 2014 Jun 25;311(24):2499-507.

[2] Rafferty E, Park J, Philpotts L, Poplack SP, Sumkin JH, Halpern EF, et al., Assessing Radiologist Performance Using Combined Digital Mammography and Breast Tomosynthesis Compared with Digital Mammography alone: Results of a Multicenter, Multireader trial. Radiology, 2013 Jan; 266(1):104-13. Epub 2012 Nov 20.

[3] Keller B, Kshirsagar A, Smith A. 3DQuorum™ Imaging Technology. Improving radiologist performance through Artificial Intelligence and SmartSlices. WP-00152-EUR-EN Rev 001 (10/19) US/International © 2019 Hologic, Inc.

An essential part of any pathway to the wider adoption of AI in breast cancer screening, once sufficient data and evidence exists to prove its utility and safety, will be building confidence in its application amongst patients and clinicians.

Patient acceptance

Education on AI and its positive impact on care is vital for patients to accept its role in their diagnosis. It will be the responsibility of government and health officials, working with healthcare providers, to work in collaboration with clinicians to communicate this. For example, educating patients about how AI-guided technologies have the potential to deliver greater accuracy of results, and to speed up diagnosis and, ultimately, treatment.

Clinical confidence

Another area in which government and health officials can play a crucial role in unlocking the potential of AI for breast cancer screening, is by fostering a sense of collaboration with and between radiologists and industry partners, to improve confidence in the medical image analysis capabilities of AI. Clinicians play a critical role in evaluating cases where this technology is used, becoming more confident in the results over time as more data are collected and audited.

Regulatory clarity

There are now some CE marked algorithms for use in medical imaging (including breast cancers) to assist prioritisation and risk stratification. This is a positive step, but we need further consensus.

There also needs to be greater clarity and consensus from governments worldwide to decide on a regulatory approach to AI. One of the biggest and most recent breakthroughs has been in the United States, where the FDA is changing regulations on how it approaches AI, giving more guidance on how AI systems can be trained. This is a welcome change that Hologic would like to see replicated in the UK, to provide more confidence around the use of AI.

A safeguarding framework

It is a reality that humans make mistakes in diagnosis, but we also need to consider what happens should AI be involved in misdiagnosis and navigate questions of accountability. The Government must take a leading role in this process by convening a multidisciplinary taskforce to work through these types of questions so that we build a more holistic approach to safeguard AI systems. Asking radiologists, experts in ethics and IT, and clinicians to work with health officials and politicians to map out a way forward on this issue would help develop broader perspectives and build effective frameworks.

Conclusion

Hologic recognises that, currently, lack of UK data prevents the immediate application of artificial intelligence (AI) within breast cancer screening. However, the challenges faced by our health system, including the breast cancer screening programme, demand the adoption of new approaches to integrating cutting edge technology into preventative care. AI has the potential to improve patient outcomes and support clinical process at a critical time. It is, therefore, imperative that a process and roadmap for the safe and widespread use of AI is set out.

We urge the Committee to focus its work on how to secure incremental benefits from the gradual introduction of AI in the near-term, and develop a pathway to wider adoption in the long-term.

About Hologic

Hologic is an innovative medical technology company primarily focused on improving women's health and well-being through early detection and treatment. Hologic enables people to live healthier lives, everywhere, every day through early detection and treatment.

Hologic provides the technology that underpins breast cancer, cervical cancer, and sexually transmitted infection screening in the UK.

UK National Screening Committee

4. The Chartered Institute for IT

Name:	Arnoldis Nyamande	**Email address:**	XXXX XXXX
Organisation (if appropriate):	BCS, The Chartered Institute for IT		
Role:	**Policy Manager**		
Do you consent to your name being published on the UK NSC website alongside your response? **Yes**			

Section and / or page number	Text or issue to which comments relate	Comment *Please use a new row for each comment and add extra rows as required.*
p.11	A range of risk factors for breast cancer have been identified, including sex, age, breast density, family history of breast cancer, genetic mutations, reproductive history, BMI, inactivity, and the use of hormone replacement therapy.	The comment identifies sex as being a risk factor for breast cancer development, but the study appears to exclusively target women. If AI breast screening is going to be the future of breast cancer screening, there needs to be a diverse and inclusive sample size. This must include transgender men, cisgender men and broad samples of people within the categories mentioned. Including a diverse range of people and ensuring categories are inclusive of this diversity will ensure the AI is enabled to aid in the treatment of all people without discrimination. This

		will help it avoid the 'poor generalisation' mentioned on page 16. A comprehensive equality impact assessment should be undertaken and published with provision for periodic iteration.
18	The primary drivers for AI in medical imaging have been cited as the desire for greater efficacy and efficiency in clinical care.	While we appreciate the need for efficacy and efficiency, we must ensure that AI in medical imaging doesn't come at the expense of high calibre service. This includes ensuring high professional standards are met at all stages of the development and implementation of the AI. An accelerated adoption of AI must be led by tried, tested, and verified data. To ensure that there aren't barriers to access, we strongly encourage uploading the data to an open source website.
19	Secondly, an algorithm is unaffected by fatigue or subjective diagnosis.	We champion any effort that seeks to ease the strain on NHS workers. In helping NHS workers, however, we must be sure not to harm the environment. There should be efforts to ensure that energy efficiency and environmental considerations should be factors in the development of screening algorithms.
19	Biases may develop through features of the mammogram, or different demographics of the women screened. This speaks to the importance of understanding the validity of studies involving AI and algorithms' transferability to other settings, but also	The probability of this happening can be lowered by implementing the suggestions covered in page 11 around ensuring diversity.

	the crucial problem of interpretability…Carter et al. argue that AI systems will inevitably encode values, and that those values may be in turn difficult to discern.	A rigorous equality impact assessment would cover most of the concerns here, however a further safeguard to avoid coder bias would be to identify and use a diverse cohort of coders to design and work with the data and the algorithms. Data samples used must also represent the diversity of the population so the AI can learn with less risk of bias. By making diversity and inclusion an integral part of the AI from the beginning, the risk of it inheriting biases with the potential to cost lives or provide suboptimal care is reduced.
20	The sharing of data has significant monetary implications, and governmental release of data to private providers without consent raises significant ethical questions.	Failing to be clear with the public about what patient health data will be used, and what it will be used for, has negative implications for public perception, trust and willingness to provide vital information to clinicians and healthcare professionals. Patients need clear communication about the nature of the AI programme, in as much detail as they require, to preserve and protect public trust. Failure to do so risks eroding public trust in the NHS Breast Screening Programme and wider NHS programmes. A lack of effective communication and an erosion of public trust may have serious implications for public health as people pause for thought before engaging with the Programme.

		Trust can be gained as the public become increasingly aware of the positive benefits of technology such as AI on their lives. To ensure this increase in trust continues to develop, we must support the professional training and development of the analysts and data scientists working with this data. We must also establish clear ethical and professional standards for use across sensitive public data. This is something BCS, The Chartered Institute for IT, champions as one of its core values – as highlighted in the recent 'Priorities for the National AI Strategy - policy discussion document'. BCS is also working with xxxx xxxx, xxxx xxxx, xxxx xxxx, xxxx xxxx, xxxx xxxx, to collaboratively shape and develop the data science profession. As part of this programme, industry-wide professional standards are to be established for data science to ensure an ethical and well-governed approach so the public can have confidence in how their data is being used[4].
20	The possibility of AI replacing radiologists is already leading to a significant proportion of medical students discounting the speciality as a career choice	Automation shouldn't necessarily lead to unemployment or make a profession obsolete. This decline in medical students choosing radiology could be the result of a lack of clear communication between prospective radiologists, the NHS

[4] https://www.bcs.org/more/about-us/press-office/press-releases/professional-standards-to-be-set-for-data-science/

		and software manufacturers about the role of the AI; it needs to be made clear that it is there to assist, not replace, them.

BCS, The Chartered Institute for IT

The purpose of BCS as defined by its Royal Charter is to promote and advance the education and practice of computing for the benefit of the public. We bring together industry, academics, practitioners, and government to share knowledge, promote new thinking, inform the design of new curricula, shape public policy and inform the public. As the professional membership and accreditation body for IT, we serve nearly 60,000 members including practitioners, businesses, academics, and students, in UK and internationally. We accredit the computing degree courses in ninety-eight universities around the UK. As a leading IT qualification body, we offer a range of widely recognised professional and end-user qualifications.

BCS is the largest professional body in the Federation for Informatics Professionals (FEDIP), the awarding body for the only UK professional register dedicated to health and social care. In this sector we collaborate with the UK Government and devolved administrations, over 40 NHS Trusts and health organisations and thousands of members to support the development of IT, digital and information professionals; driving professional-development and lifelong learning to improve professional skills, competence and public trust.

Summary of the BCS position

BCS is supportive of the role AI has to play in improving the health and care of the population and supports the NSC's position outlined in the consultation document: 'Use of artificial intelligence for image analysis in breast cancer screening – Rapid review and evidence map[1]' not to endorse the implementation of AI for Breast Cancer Screening in the UK at present.

Taking an iterative approach where we walk before we can run in such critically important areas of public health is vital. There are a number of critical questions which need to be explored before a full roll out of AI in the screening of breast cancer; such as how we inform the public and maintain public trust, how we minimise and watch for bias, how we establish a mainstreamed culture of data ethics amongst those governing, collecting and using the data that will inform the AI, and vitally, establish clear evidence on AI's impact in increasing accurate diagnosis.

BCS is keen to support the initiative in the future once it is evident that more research has been done to ensure that AI Breast Screening will be safe for all who use it.

UK National Screening Committee

5. The Royal College of Pathologists' Digital Pathology Committee

Name:	Janine Aldridge	**Email address:**	xxxx xxxx
Organisation (if appropriate):	The Royal College of Pathologists' Digital Pathology Committee		
Role:	**Public Affairs Officer**		

Do you consent to your name being published on the UK NSC website alongside your response?

Yes X **No**

Section and / or page number	Text or issue to which comments relate	Comment *Please use a new row for each comment and add extra rows as required.*
	The College welcomes the evidence-based approach to the adoption of AI in the NHS. The main mention of pathology in the text is in relation to the use of histopathology as a gold	The College is aware that histopathology is one of the areas in which AI is likely to be used in the near future. And pathology is an area in which there is clinical need for AI (both in terms of clinical capacity and opportunities to improve diagnosis).

	standard in the development or evaluation of radiology AI tools.	Several AI products have already been developed in this area for pathology and are in trial or early clinical evaluation stages. The College recommends that the National Screening Committee initiate work to include the consideration of the use of AI in breast pathology, which is of moderate urgency given the clinical need and growing use of AI in this area. The College would be pleased to support and advise in this area.
P3-4	Several potential places of AI in the breast screening pathway have been envisaged – 3 options given.	A member of the committee has suggested a fourth option which would be to use AI after all the normal screening processes had been followed as a tool to capture lesions overlooked by human screeners. This will require some resource as someone has to decide on whether any additional lesions picked up by AI are clinically relevant. Perhaps this would be best done as a study.

UK National Screening Committee

6. Kheiron Medical Technologies

Name:	Simon Harris	**Email address:**	xxxx xxxx
Organisation (if appropriate):	Kheiron Medical Technologies		
Role:	**Senior Project Manager – Leading the NHSx AI Award**		

Do you consent to your name being published on the UK NSC website alongside your response?

Yes **No**

Section and / or page number	Text or issue to which comments relate	Comment *Please use a new row for each comment and add extra rows as required.*
(Executive Summary) Page 4	The aim of this review was to synthesise the evidence on the use of deep learning AI algorithms to read mammograms (as reader aid or stand-alone) of women attending routine breast screening for digital (full field digital mammography, FFDM)	Due to the timing of this rapid review and that it looked at studies published between January 2010 and September 2020 – it did not include Kheiron's retrospective study results from 2020, which provides significant evidence on the efficacy of at least one AI algorithm for breast screening. The official names of the studies conducted were: AUX-07-2018-KMT 1.8 (UK) and AUX-07-2018-KMT 1.2 (HU)

	mammograms. The evidence is presented in the form of a rapid review (question 1) and an evidence map (question 2). The review included studies published between January 2010 and September 2020 and aimed to address the two questions answering the UK NSC criteria as outlined.	The pre-print of them can be found here: https://www.medrxiv.org/content/10.1101/2021.02.26.21252537v1.full This has now been submitted for peer-review. This retrospective study evaluated the performance of the Mia™ version 2.0.1 AI system from Kheiron Medical Technologies on an unenriched sample (275,900 cases from 177,882 participants) collected across seven screening sites in two countries and four hardware vendors and is representative of a real-world screening population over 10 years. Performance was determined for standalone AI and double reading to assess non-inferiority and superiority on relevant screening metrics. Crucially to the NSC review, this included data from 3 different UK NHS Breast Screening sites. The results demonstrate that the evaluated AI system can be an effective solution acting as an independent reader in the double reading workflow. The results show that when Mia is in used in double reading as an independent reader, we can expect the standard of care at least preserved on all relevant screening metrics and to be improved on a subset (i.e. superiority or non-inferiority in each of RR, CDR, PPV, SEN, SPEC between double reading with Mia and double reading without in our study). The scale and diversity of samples

		support that the findings are generalisable to many screening programmes and the use of practical metrics ensures that the impact of introducing AI into everyday screening is reliably estimated and of clinical relevance. This work was subject to an independent CRO analysis and review.
(Recommendations on screening) Page 7	**Recommendations on screening** There is insufficient evidence in quality and quantity to recommend implementation of AI into clinical practice of the NHS breast screening programme. Overall, the evidence on the test accuracy of AI algorithms to detect breast cancer in women attending screening mammography using geographical validation test sets was sparse and lacked applicability to the UK context (no	Please see the results contained within Kheiron's pre-print which has now been submitted for peer-review: https://www.medrxiv.org/content/10.1101/2021.02.26.21252537v1.full This retrospective study evaluated the performance of the Mia™ version 2.0.1 AI system from Kheiron Medical Technologies on an unenriched sample (275,900 cases from 177,882 participants) collected across seven screening sites in two countries and four hardware vendors and is representative of a real-world screening population over 10 years. Performance was

	study used a UK dataset). Except for one study, study populations were small with a cancer prevalence atypical of the screening context.	determined for standalone AI and double reading to assess non-inferiority and superiority on relevant screening metrics. Crucially to the NSC review, this included data from 3 different UK NHS Breast Screening sites.
(Summary of Findings Relevant to Criteria 4 and 5) Page 55	There were no studies that described accuracy of AI integrated into any breast screening pathway, and no prospective studies of test accuracy in clinical practice. Therefore, there is no direct evidence on how AI may affect accuracy if integrated into UK breast screening practice. There were three enriched test set MRMC laboratory studies reporting test accuracy for a single read of AI as a reader aid, but these will be subject to the laboratory effect bias where radiologists act differently in test sets than clinical practice. There were four studies examining AI accuracy in test sets, of which only one was a consecutive or random sample of women attending breast cancer screening, and this study did not use an AI algorithm with a pre-set threshold. There is some evidence from early-stage evaluation	Please see the results contained within Kheiron's pre-print which has now been submitted for peer-review: https://www.medrxiv.org/content/10.1101/2021.02.26.21252537v1.full The study results compared the performance of screening with double reading with and without Mia as an independent reader in the workflow. Further information can be provided and will be published from this study on how various workflow integrations impact performance. We believe that this study presents strong evidence on how AI may affect accuracy if integrated into UK breast screening practice. With regards to the current lack of evidence of prospective studies of AI within the breast screening pathway – this is exactly what Kheiron is undertaking as part of the Phase 4 AAC/NIHR/NHSx AI in health & care award. Kheiron was one of the successful recipients in the first round of the awards. We believe that we have conclusive evidence

	studies that AI has the potential to be an accurate tool to detect cancer in breast screening mammograms. However, the current evidence is a long way from the quality and quantity required for implementation into clinical practice.	but will complete confirmatory prospective studies within the next few months as part of the AI Award programme. Kheiron would be delighted to engage with the NSC, provide our current evidence and help shape future recommendations.
Page 67	In summary, at present there is an insufficient volume of evidence on clinical utility related to the use of AI in the NHSBSP or analogous populations to justify commissioning an evidence review. No evidence from high quality randomised controlled trials or prospective cohort studies was identified that compared the benefit of a breast cancer screening programme using AI to a screening programme without AI on clinical outcomes and patient	It is helpful to have this summary from the NSC. We hope that in the comments above we have provided sufficient evidence as to why Kheiron have already or are addressing these challenges head-on. In addition to this, we believe it is important to point out that we have also developed other versions of the AI product which may have much lower evidentiary requirements due to lower impact. We strongly believe that performance and evidence requirements are dependent on intended use. For instance, since most of the workflow configurations that Mia is intended for are strictly back-end and not interacting with the human readers, we believe that 'the influence that the knowledge of AI scores has on radiologists' is not relevant.

	management and practical implication outcomes. The limited evidence currently available from retrospective simulation studies, retrospective cohort / case-control or enriched test set MRMC laboratory reader studies show potential for AI to reduce radiologist workload without compromising performance. However, these studies do not allow evaluation of the influence that the knowledge of AI scores has on radiologists in a prospective clinical setting, making the quality of the evidence unsuitable for drawing conclusions on the effectiveness of AI use in screening practice.	

UK National Screening Committee

7. Gloucester hospitals NHS FT

Name:	xxxx xxxx	**Email address:**	xxxx xxxx
Organisation (if appropriate):	Gloucester hospitals NHS FT		
Role:	**Consultant radiologist**		

Do you consent to your name being published on the UK NSC website alongside your response?

Yes **No**

Section and / or page number	Text or issue to which comments relate	Comment *Please use a new row for each comment and add extra rows as required.*
Multiple pages especially 67, 68	Strategy for validation needs rethinking	The key point is that training and test sets should not overlap at. HOWEVER, whilst geographic validation will ensure this in most instances, in urban areas in particular, like London, as many as 30% of women invited to screening may move into

		different screening catchment areas between screening rounds! In fact, temporal validation using the same site as the training set will often be valid.
Page 27	AI as full single read	Agree helpful for direct head to head comparison of different AI systems but only if there is a database of curated mammograms that have not been used by ANY of the algorithms being compared in development/validation – as could be held by OPTIMAM for example.
Page 36	Retrospective studies using validation test sets	True status of AI pos/human reader neg cases can be inferred if follow-up is long enough e.g. 1 or 2 subsequent screens, amounting to 3 or 6 years of follow-up. See comment below.
Page 38-40, 70-72	Incorporation bias and differential verification bias.	The use of long term follow-up, as described above, can mitigate

		this bias. This is one of the advantages of using retrospective data with a long follow-up period for diagnostic accuracy studies.
	Prospective and retrospective studies	It is true that prospective studies are very important in demonstration of the true impact of AI on workflows in real world scenarios. However, retrospective studies are still very valid to demonstrate standalone performance of a given system in a given screening context where the ground truth is known. We fully concur with the statement about laboratory reader studies and share concern about the overemphasis is results from multireader studies of enriched datasets.
Page 37	The review identified 4 studies (1 non-enriched[81] and 3 enriched[76 78 80]) reporting the test accuracy for a single read of AI as a stand-alone system, of which 3 used a retrsopective cohort design (retrospective test accuracy study)[81 73 77] and one an enriched test set MRMC laboratory study design.[75] Furthermore, the review identified	Am I alone in being unable to make sense of this? It states 4 studies were identified but 7 different references are given…

Page 73.	‘Algorithms are short lived’	Very true. There needs to be some mechanism whereby continued safety is checked and performance is regularly monitored prospectively. AI developers will continue to tinker with and improve their algorithms. This mandates establishment of a ‘quarantined’ national dataset to which AI developers can apply to use for testing of their algorithm.
Other	Non standard images	Many if not all algorithms cannot handle more that the regulation 4 views (very large breasted women), technical repeats or recalls, implants etc. etc. This will need to be addressed.

UK National Screening Committee

8. Royal Society of Biology

Name:	Asari E. Inyang	**Email address:**	xxxx xxxx
Organisation (if appropriate):	Royal Society of Biology		
Role:	Associate Member		

Do you consent to your name being published on the UK NSC website alongside your response?

Yes **No**

Section and / or page number	Text or issue to which comments relate	Comment *Please use a new row for each comment and add extra rows as required.*
Page 3	Several potential places of AI in the breast screening pathway have been envisaged.	It would be great to have a hybrid of AI and human readers for the purpose of accuracy, however a review in 1-3 years' time would be essential and the methods to be used as a baseline for future review.