User and developer views on using Artificial Intelligence technologies to facilitate the early diagnosis of skin cancers in primary care settings: A qualitative study

Owain Tudor Jones, Natalia Calanzani, Suzanne E Scott, Rubeta N Matin, Jon Emery, Fiona M Walter

Submitted to: Journal of Medical Internet Research
on: May 20, 2024

Table of Contents

User and developer views on using Artificial Intelligence technologies to facilitate the early diagnosis of skin cancers in primary care settings: A qualitative study

Owain Tudor Jones[1] BSc, BMBS, MPhil; Natalia Calanzani[2] BSc, MSc, PhD; Suzanne E Scott[3] BSc, MSc, PhD; Rubeta N Matin[4] PhD; Jon Emery[3, 5] DPhil; Fiona M Walter[3, 5] MD

[1]Institute of Applied Health Sciences University of Aberdeen Aberdeen GB
[2]Wolfson Institute of Population Health Queen Mary University of London London GB
[3]Department of Dermatology Churchill Hospital Oxford GB
[4]Centre for Cancer Research, and Department of General Practice and Primary Care University of Melbourne Melbourne AU

Corresponding Author:
Owain Tudor Jones BSc, BMBS, MPhil

Abstract

Background: Skin cancers, including melanoma and keratinocyte cancers, are amongst the commonest cancers worldwide and their incidence is rising in most populations. Earlier detection of skin cancer leads to better outcomes for patients. Artificial Intelligence (AI) technologies have been applied to skin cancer diagnosis, but many technologies lack clinical evidence and/or the appropriate regulatory approvals. There are few qualitative studies examining the views of relevant stakeholders, or evidence about implementation and positioning of AI technologies in the skin cancer diagnostic pathway.

Objective: To understand the views of several stakeholder groups on the use of AI technologies to facilitate the early diagnosis of skin cancer including patients, members of the public, general practitioners, primary care nurse practitioners, dermatologists, and AI researchers.

Methods: A qualitative, semi-structured interview study with 29 stakeholders. Participants were purposively sampled based on age, gender, and geographical location. Interviews were conducted via Zoom between September 2022 and May 2023. Transcribed recordings were analysed using thematic framework analysis. The framework for the non-adoption, abandonment, and challenges to scale-up, spread, and sustainability (NASSS) was used to guide the analysis, to help understand the complexity of implementing diagnostic technologies in clinical settings.

Results: Major themes were ‘the position of AI in the skin cancer diagnostic pathway’ and ‘the aim of the AI technology’; cross-cutting themes included trust, usability and acceptability, generalisability, evaluation and regulation, implementation, and long-term use. There was no clear consensus on where AI should be placed along the skin cancer diagnostic pathway but most participants saw the technology in the hands of either patients or primary care practitioners. Participants were concerned about the quality of the data used to develop and test AI technologies, and the impact this could have on their accuracy in clinical use with patients from a range of demographics and the risk of missing skin cancers. Ease of use and not increasing the workload of, already strained, healthcare services were important considerations for participants. Healthcare professionals and AI researchers reported a lack of established methods of evaluating and regulating AI technologies.

Conclusions: This study is one of the first to examine the views of a wide range of stakeholders on the use of AI technologies to facilitate early diagnosis of skin cancer. The optimal approach and position in the diagnostic pathway for these technologies have not yet been determined. AI technologies need to be developed and implemented carefully and thoughtfully, with attention paid to the quality and representativeness of the data used for development, to achieve their potential.

(JMIR Preprints 20/05/2024:60653)
DOI: https://doi.org/10.2196/preprints.60653

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

User and developer views on using Artificial Intelligence technologies to facilitate the early diagnosis of skin cancers in primary care settings: A qualitative study

Owain T Jones[1], Natalia Calanzani[2], Suzanne E Scott[3], Rubeta N Matin[4], Jon Emery[5,3], Fiona M Walter[3,5]

1. Department of Public Health & Primary Care, University of Cambridge, Cambridge, UK
2. Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK
3. Wolfson Institute of Population Health, Queen Mary University of London, London, UK
4. Department of Dermatology, Churchill Hospital, Oxford, UK
5. Centre for Cancer Research, and Department of General Practice and Primary Care, University of Melbourne, Melbourne, VIC, Australia

Corresponding author: Dr Owain Jones. Department of Public Health & Primary Care, University of Cambridge, Cambridge, UK. otj24@cam.ac.uk

Abstract

Background

Skin cancers, including melanoma and keratinocyte cancers, are amongst the commonest cancers worldwide and their incidence is rising in most populations. Earlier detection of skin cancer leads to better outcomes for patients. Artificial Intelligence (AI) technologies have been applied to skin cancer diagnosis, but many technologies lack clinical evidence and/or the appropriate regulatory approvals. There are few qualitative studies examining the views of relevant stakeholders, or evidence about implementation and positioning of AI technologies in the skin cancer diagnostic pathway.

Objective

To understand the views of several stakeholder groups on the use of AI technologies to facilitate the early diagnosis of skin cancer including patients, members of the public, general practitioners, primary care nurse practitioners, dermatologists, and AI researchers.

Methods

A qualitative, semi-structured interview study with 29 stakeholders. Participants were purposively sampled based on age, gender, and geographical location. Interviews were conducted via Zoom between September 2022 and May 2023. Transcribed recordings were analysed using thematic framework analysis. The framework for the non-adoption, abandonment, and challenges to scale-up, spread, and sustainability (NASSS) was used to guide the analysis, to help understand the complexity of implementing diagnostic technologies in clinical settings.

Results

Major themes were ‘the position of AI in the skin cancer diagnostic pathway’ and ‘the aim of the AI technology’; cross-cutting themes included trust, usability and acceptability, generalisability, evaluation and regulation, implementation, and long-term use. There was no clear consensus on where AI should be placed along the skin cancer diagnostic pathway but most participants saw the technology in the hands of either patients or primary care practitioners. Participants were concerned about the quality of the data used to develop and test AI technologies, and the impact this could have on their accuracy in clinical use with patients from a range of demographics and the risk of missing skin cancers. Ease of use and not increasing the workload of, already strained, healthcare services were important considerations for participants. Healthcare professionals and AI researchers reported

a lack of established methods of evaluating and regulating AI technologies.

Conclusions

This study is one of the first to examine the views of a wide range of stakeholders on the use of AI technologies to facilitate early diagnosis of skin cancer. The optimal approach and position in the diagnostic pathway for these technologies have not yet been determined. AI technologies need to be developed and implemented carefully and thoughtfully, with attention paid to the quality and representativeness of the data used for development, to achieve their potential.

Keywords: Artificial Intelligence, AI, Machine Learning, ML, Primary Care, Skin Cancer, Melanoma, Qualitative Research

Introduction

Skin cancers are amongst the commonest cancers worldwide and their incidence is rising in most populations.[1,2] Melanoma is the most dangerous skin cancer; but the keratinocyte cancers (KCs), which include squamous cell carcinomas (SCCs) and basal cell carcinomas (BCCs), make up the majority of skin cancers.[1-3] Earlier diagnosis of skin cancers is associated with significantly better outcomes.[3]

There has been substantial excitement and interest in applying AI technologies to skin cancer diagnosis. Many AI technologies have been designed for use by patients and a handful of technologies have been developed for use as clinical decision aids by clinicians.[4,5] However, most of these technologies do not have the appropriate regulatory approvals in place to support their safety and efficacy when used in these settings.[6] To date there is limited evidence on the efficacy and accuracy of AI technologies aimed at the diagnosis of skin cancer from clinical trials or from real-life clinical settings.[4,5] Studies on the use of AI technologies in any clinical setting show limited evidence on how users interact with AI technologies and the implications of this for patient safety.[7] There is also a dearth of qualitative studies to gain perspectives of patients and members of the public.[8] Implementation of a diagnostic technology in clinical settings is a complex process and prone to failure - a technology has to pass through several stages of development before implementation is likely to be successful.[9] Several frameworks analyse factors around the implementation of new technologies; the NASSS framework (the non-adoption, abandonment, and challenges to scale-up, spread, and sustainability framework) was developed using complexity principles and allows researchers to identify and explain the manifestations of complexity in technology-supported change projects (figure 1).[10] Some evidence from implementation of AI technologies intended to diagnose skin cancer in clinical settings is emerging.[11,12]

The aim of this study was to consult with several stakeholder groups, including patients, members of the public, general practitioners (GPs), dermatologists, primary care nurse practitioners (NPs), and AI researchers, to identify barriers and facilitators to the implementation of AI technologies to help facilitate the early diagnosis of skin cancers in primary care settings, and provide in-depth, qualitative evidence of their views.

Methods

Design

A qualitative study using semi-structured interviews. Ethical approval was granted by the Cambridge Psychology Research Ethics Committee (Application No: PRE.2021.098).

Recruitment and sampling

Four groups were selected for the study: (1) members of the public; (2) patients previously diagnosed with skin cancer; (3) healthcare professionals (HCPs) (including general practitioners (GPs), primary care nurse practitioners, and dermatologists)[1]; and (4) AI researchers from academic and commercial settings. Members of the public were approached via the Cambridge Biomedical Research Centre Patient and Public Involvement (PPI) group and 'snowballing' invitations to their colleagues. Members of the public with a history of skin cancer were included in the patient group, with additional patients approached via a national charity, Melanoma Focus.[13] GPs and primary care nurse practitioners were identified through the Primary Care Dermatology Society (PCDS),[14] with additional GPs approached via Sermo (medical market research organisation).[15] Dermatologists were identified from the British Association of Dermatologists (BAD),[16] and through snowballing. AI researchers from academic settings were identified through contacts with academic institutions, and AI researchers from commercial settings were recruited from the companies identified in two reviews via email. [4,5]

Participants were sampled to achieve a spread of age, gender, and geographical location within each participant group. Patients and public were recruited to achieve a spread of ages over and under 60 years of age, reflecting the average age of skin cancer diagnosis. HCPs and AI researchers were recruited to achieve a spread of ages over and under 45 years of age, reflecting the mid-career point of these professions and varying experience levels. Patients were also sampled to include a range of

[1] In this paper we use Primary Care Practitioner to denote any medical practitioner that works in a primary care setting and might consult with a patient about a suspicious skin lesion, including general practitioners (GPs), family doctors, nurse practitioners, physician assistants, and paramedic practitioners. Healthcare professional (HCP) is used when we refer to the views of wider healthcare professionals, including secondary care HCPs. Staff who work in primary/community care and do not have clinical training and experience in the diagnosis of skin cancer but could potentially use an AI technology with patients with suspicious skin lesions are referred to as allied healthcare professionals, a wide variety of professions could be included in this group, but it certainly includes practice nurses, healthcare assistants, clinical navigators, pharmacists, and podiatrists.

skin cancer types. GPs were sampled to include a spread of GP roles (GP partner, salaried GP, locum GP, GP with Extended Role [GPwER] in dermatology). We aimed to recruit at least 5-8 participants per group, but more HCPs (10-12) to achieve a broad range of views within each HCP group. To facilitate recruitment, a £20 Apple iTunes or Google Play voucher was offered to all participants.

Data collection

An interview topic guide was developed with input from a PPI group and members of the research team. It aimed to explore views on facilitators and barriers to the use of AI technologies to help diagnose skin cancers in primary care.

All interviews were conducted by OJ, who has a clinical background, with guidance from NC, a health services researcher with expertise in qualitative research. Interviews took place online at a time of the participants choosing, following informed consent. All interviews were completed between 9th September 2022 and 25th May 2023. Interviews continued until we had a rich, multi-faceted dataset.

Analysis

All interviews were digitally recorded and transcribed verbatim by a professional transcription company. Transcripts were checked, anonymised, and analysed inductively and deductively using thematic framework analysis.[17] Two researchers (OTJ and NC) repeatedly read the first 5 transcripts to become familiar with the data and generate initial codes. These initial codes were compared with the NASSS framework (Figure 1)[10] to generate a comprehensive list of codes and place the evidence within the context of implementation of health and care technologies and the complexity associated with this. The remaining transcripts were then read and indexed using this list of codes using NVIVO 14,[18] with codes modified/added as required. Additional codes were created where the data did not fit comfortably into the NASSS framework. Coding was completed by OJ with a sample of transcripts (25%) checked by NC. Consistency of coding was discussed regularly through team meetings, and a third researcher (FMW [academic GP]) was consulted if necessary. Data were charted into Microsoft Excel[19] and the characteristics, similarities, and differences between data were identified. Relationships and connections between categories were then mapped to generate the themes presented in the results. Themes were further refined with guidance from senior team members (FMW and SS [health psychologist]).

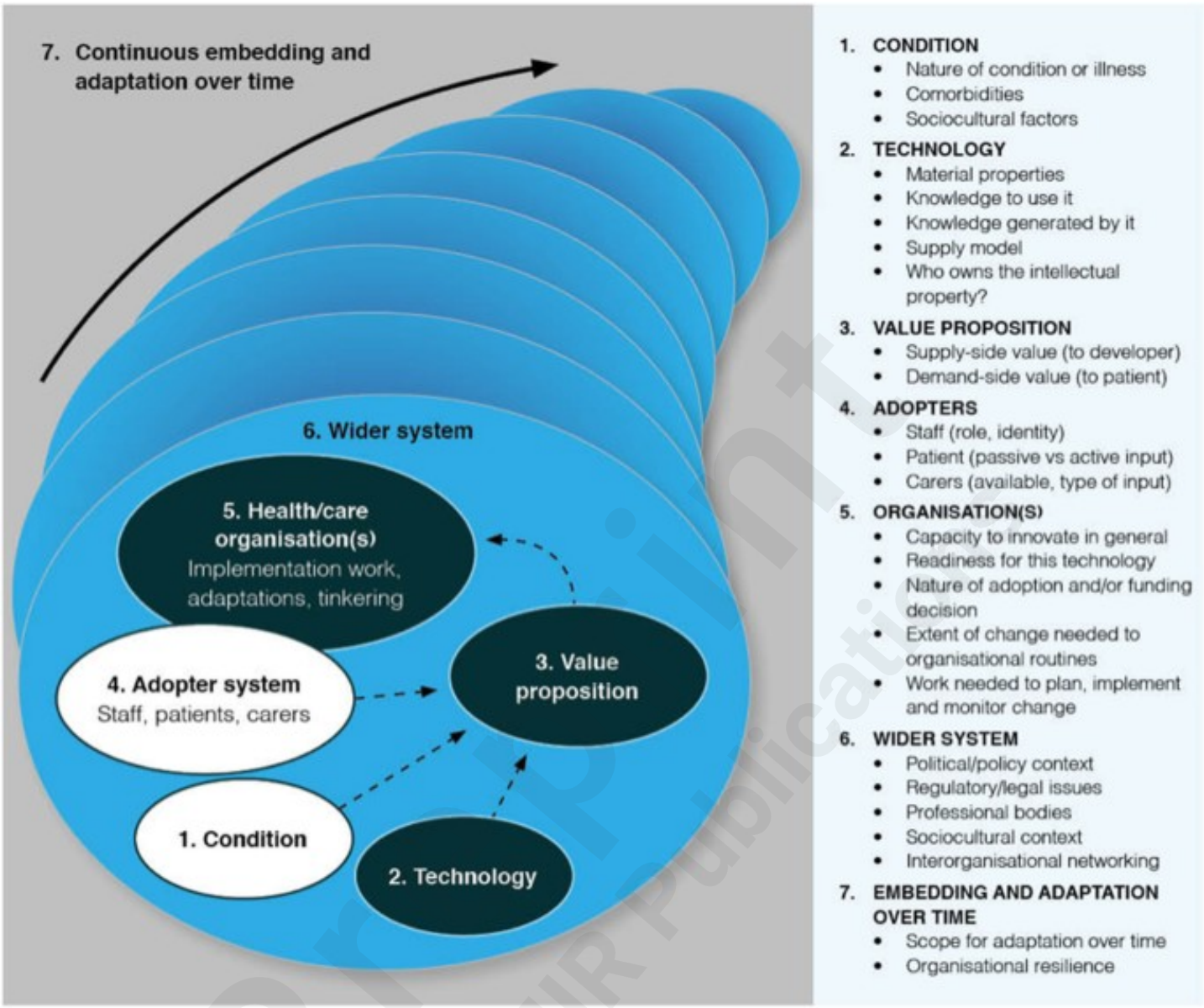


Figure 1: The NASSS Framework[10]

Results

Participants

Twenty-nine interviews were conducted with members of the public (n=6), patients (n=5), healthcare professionals (n=13), and AI researchers (n=5), see Table 1. Participants were recruited via mailing lists and social media, so the denominator is unknown.

Table 1: Participant demographics

	Overall (n=29)	Public* (n=6)	Patients** (n=5)	HCPs *** (n=13)	AI researchers (n=5)
Gender					
Male	10 (34%)	1	1	4	4
Female	19 (66%)	5	4	9	1
Age[1]					
>60		2	2		
<60		4	3		
>45				7	1
<45				6	4
Location					
England	23 (79%)	5	5	11	2
Wales	2 (7%)	1		1	
Scotland	1 (3%)			1	
*Outside UK****	3 (10%)				3
Support for using AI to facilitate the early diagnosis of skin cancers in primary care					
Strongly Approve	10 (34%)	2	1	5	2
Approve	15 (52%)	3	3	6	3
Neutral	2 (7%)		1	1	
Disapprove	1 (3%)	1			
Missing data	1 (3%)			1	
Highest Educational qualification***					
Foundation or intermediate qualifications	0 (0%)				
Advanced qualifications	1 (3%)		1		
Higher qualifications	28 (97%)	6	4	13	5
Ethnic Origin					
White British	18 (62%)	5	3	10	
White European	6 (21%)		1	2	3
Middle Eastern	1 (3%)				1
Black African	1 (3%)	1			
Missing data	3 (10%)		1	1	1
History of Skin Cancer					
Yes			5******		
No		6			
Skin Cancer in a family member or close friend		2	5		

NB: [1]Patients and public were recruited to achieve a spread of participants above and below 60 years of age. HCPs and AI researchers were recruited to achieve a spread of ages above and below 45 years of age
* Occupations: project manager, lawyer, pharmaceutical industry, music industry, biomedical scientist, or missing data (n=1 for each)
** Occupations: fraud services, social researcher, teacher, accountant, research consultant (n=1 for each).
*** HCPs included General Practitioners (GPs, n=6), Primary Care Nurse Practitioners (n=3), and Dermatologists (n=4)
**** Including the Netherlands and North America
***** Educational qualification categories were taken from the UK National Census 2021. Definitions are available from: https://www.ons.gov.uk/census/planningforcensus2021/questiondevelopment/qualificationsquestiondevelopmentforcensus2021
****** basal cell cancers (BCCs) (n=3), melanoma (n=1), Merkel cell cancer (n=1).

The results section is structured around the major themes and sub-themes that were generated from the data. Major themes included the "position of AI in the skin cancer diagnostic pathway" (with sub-themes around whether the AI was positioned before, with, or after the Primary Care Practitioner (PCP)), and the "aim of the AI technology" (with sub-themes on: education, screening and diagnosis; focussing on diagnosing benign or malignant skin lesions; focussing on melanoma alone or all skin cancers; and management of lesions that are close to the threshold for malignancy). There were several cross-cutting themes, including trust, acceptability, generalisability, evaluation, regulation, implementation, and long-term use (see Figure 1). Participant quotations are identified by participant group, gender, and age.

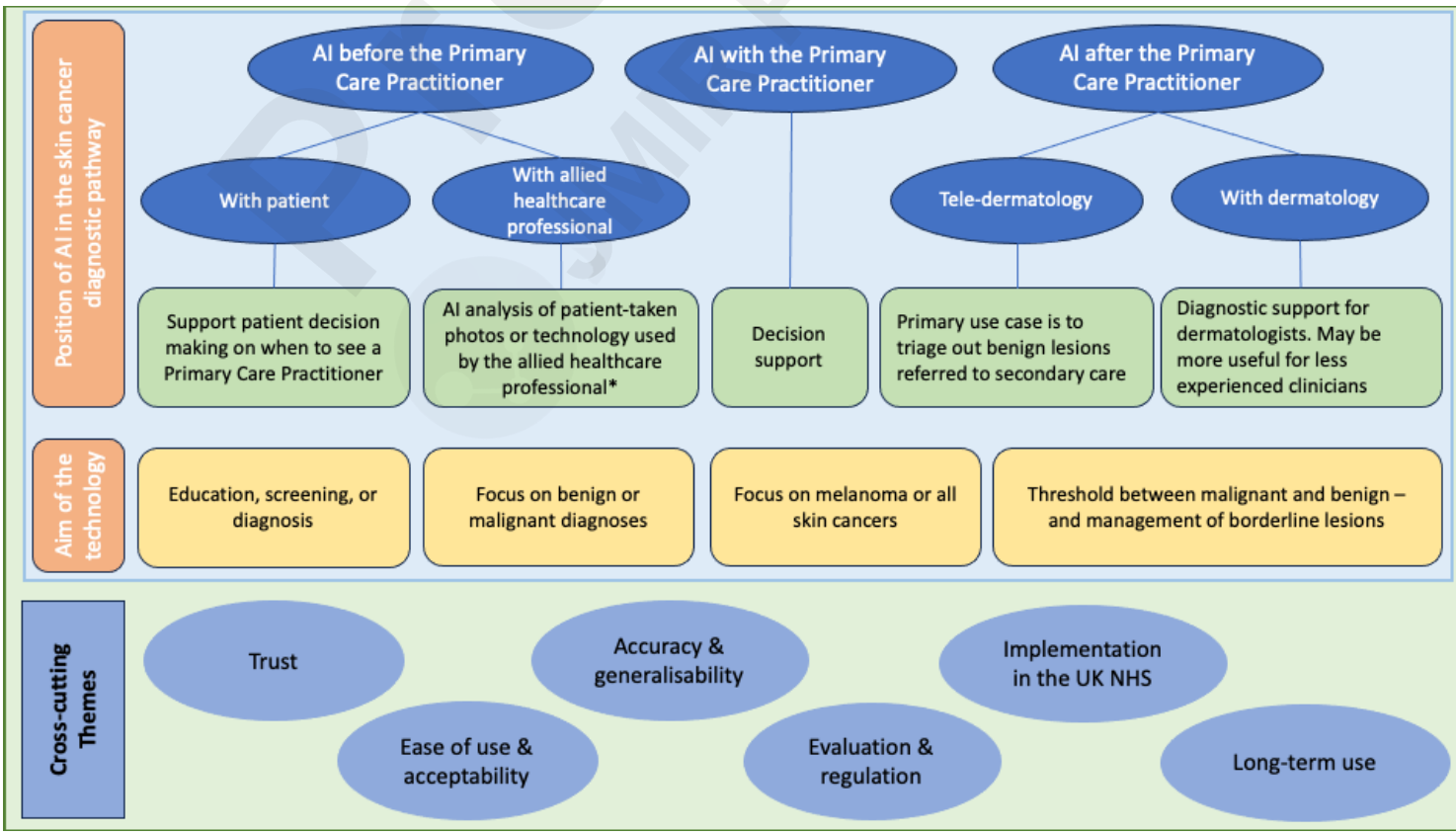


Figure 2: Issues around the design and intended positioning of Artificial Intelligence (AI) technologies for the diagnosis of skin cancer that were identified from the data, and further cross-cutting themes identified

Key: AI = Artificial Intelligence, *allied healthcare professionals that could use the technology include: pharmacists, healthcare assistants, practice nurses, podiatrists, hairdressers, and potentially many more, UK NHS = UK National Health Service

Position of AI in the skin cancer diagnostic pathway (see Table 2)

AI before the primary care practitioner (PCP)

Patients and members of the public felt that, if an AI technology were put in the hands of patients, it would be more accurate than 'random googling' and could help them to decide when to see a doctor. However, they expressed concerns about using the AI technology without clinical input. The loss of the human touch in consultations worried them, including how the diagnosis would be communicated and whether patients would be able to get a PCP appointment to discuss the diagnosis and ask questions. Another concern was whether all patients would be able to use the technology effectively and the potential for it to exclude some groups of patients, including elderly patients. Patients and the public felt that if the AI technology were positioned in a clinic led by a practice nurse or other allied HCP that would reduce the barrier for patients to access skin lesion assessments and could reduce the risk of over-use of the technology by patients.

HCPs could all envision AI technology being used by patients and recognised demand for this service. Some dermatologists felt this could be a useful approach particularly for high-risk patients in between dermatology clinic appointments, as long as the technology was properly evaluated. However, they were concerned about the potential for over-use and subsequent increased workload in primary and secondary care due to the false positives that would inevitably occur. They wondered whether patients would struggle to understand the risks, benefits, and the output of the AI technology, and would find it difficult to take images of good-enough quality for the AI to analyse. Some HCPs commented on the potential psychological benefits of AI technology, in particular reducing patient anxiety associated with referrals on urgent suspected cancer (USC) pathways, although some HCPs also felt that it could increase patient anxiety. Dermatologists and nurse practitioners commented on how an accurate AI technology could enable a wide variety of allied healthcare professionals to employ the technology to triage patients presenting with suspicious skin lesions, which may reduce the barrier for patients to be seen and enable earlier detection of skin cancer. GPs reported mixed views on the potential of positioning AI technologies in the role of triaging patient-submitted photographs before an appointment with a PCP.

AI researchers from academic backgrounds were broadly sceptical of patients having access to AI technology themselves, highlighting concerns about the diagnostic accuracy of the technology, the risk of false reassurance, and that patients would not necessarily understand the context of when it was safe to use. However, AI researchers from commercial backgrounds felt that, if the AI technology was accurate enough, positioning AI technologies with patients had the greatest potential for impact. They highlighted various potential positive effects, including: reducing the barrier for patients to get a skin lesion assessed, helping patients make better decisions about when they needed to see a healthcare professional, and educating patients about skin cancer.

AI with the primary care practitioner (PCP)

Patients commented that GP surgeries are usually more accessible compared to hospitals and so are an ideal place to locate the technology. They felt that if they were consulting about a skin lesion that could be a dangerous skin cancer (i.e. melanoma), they would like to be able to speak to a PCP. Combining the clinical judgement of the PCP with the AI technology was viewed as being more powerful than the PCP alone. This approach could also potentially be more beneficial for PCPs with less experience in skin cancer and reduce the variability in skin cancer triage in primary care. The public discussed whether PCPs would become reliant on the AI technology and lose their clinical judgement, and even whether this was an issue if the AI were more accurate than the PCP in diagnosis.

GPs largely wanted PCPs to have access to the AI technology and were concerned that the implementation of AI technology at other points in the diagnostic pathway could undermine their gate-keeping role. They believed that PCPs are well positioned to triage and monitor skin lesions over time and provide continuity of care, and that it would be more efficient for the technology to be used and a triage decision made in a single consultation by a PCP rather than having multiple appointments before seeing the PCP, for example to image the lesion. One GP emphasised that the positioning of the technology in the diagnostic pathway is probably the most important factor in how impactful it will be. Nurse practitioners also felt that AI technology would be best placed with a PCP.

Dermatologists agreed that some intervention was needed to improve the accuracy of referrals from primary to secondary care but were not sure implementing AI technology in primary care was the best approach. A broad educational program for PCPs was suggested as a better alternative. Whether

AI technology is currently accurate enough to be used by PCPs was a concern, as well as the risk of de-skilling PCPs in skin lesion recognition. Nonetheless, some dermatologists saw benefits in situating an AI technology with PCPs, to help with rarer skin cancer recognition, to enable a PCP to be present to break bad news and answer patient questions, and to give PCPs more confidence in diagnosing a lesion as benign and not referring on to secondary care.

Most AI researchers expressed strongly that a "human-in-the-loop" approach (i.e., where a human clinician is always involved in decision making alongside the AI) is needed. Potential benefits of this approach included: increased safety, allowing for patient interactions, allowing the PCP to focus more on the patient, and increasing knowledge of PCPs. AI researchers also felt that the decision to biopsy or refer a skin lesion involves many biopsychosocial and clinical factors outside of the appearance of the lesion, which cannot be measured or taken into account by the AI, and thus needs input from a PCP.

AI after the primary care practitioner (PCP)

Use of AI technology to triage referrals from primary care to secondary care in a teledermatology setting was only mentioned by one GP and two dermatologists. The dermatologists suggested AI technology could help triage obviously benign skin lesions preventing them being referred to a dermatology clinic.

Few participants commented on positioning an AI technology with dermatologists. Dermatologists believed that it could be a useful training tool. If AI technology is proven to be as good or better than dermatologists in the future, then it could be used to assess and diagnose skin lesions in dermatology clinics. One AI researcher discussed how positioning the AI technology with dermatologists might help their diagnostic accuracy and improve consistency.

Table 2: Positioning of AI in the skin cancer diagnostic pathway
AI before the primary care clinician
So I think people first self-diagnose via Google a lot anyway so having something that's a little bit more accurate than sel diagnosis is probably helpful. Pat4 F<60yrs
I don't think it would be effective giving it directly to patients, least of all because they won't have the training and the righ level of professional knowledge in being able to interact with digital innovations effectively. Pat1 M<60yrs
I would be a little bit concerned about it just being open-ended, because… the potential for over-use by some people an under-use by others would be quite large, I think. So, I think… a nurse-led clinic or something like that might be better. C give people the option. Pub3 F>60yrs
I know from doing online and video consultations during COVID, what was quite evident is that patients have got varyin skills in terms of how they use their mobile devices to show you their skin remotely and taking images, and the quality o

those images. So I think you'd have to think very carefully about who is taking the image, what they understand about wha the clarity of that image needs to be, and what devices they're being used on. Because that's going to affect you interpretation. So I would be hesitant to say as a matter of routine for it to be patient led. NP1 F>45yrs
And seeing this opportunity, for example, hairdressers, podiatrist, pharmacist, they get asked a lot about these things, abou lesions. And from time to time you get referrals because they went to the hairdresser and they spotted a mole… it would be good thing to implement in these areas, expand beyond the GPs. Derm3 F<45yrs
I would really like to then streamline the process. So we have an ability within our practice that the patient doesn't need a appointment for asking about a mole of concern and they can send a photo. That photo would be reviewed by a clinician. would be absolutely fantastic if that photo was also able to be reviewed by an AI process and advise whether there wer concerning features on it or not. It would be useful if that photo could then be reviewed again six weeks later against a new photo to assess for changes. GP5 M<45yrs
Yeah, it [using the app] has especially low barrier to use it so it's way lower than going to a doctor or going to dermatologist. It is really a first step into being more interested or more concerned about skin cancer. And I think it can hel to raise awareness (of skin cancer) and inform people better. AIR4 M>45yrs
AI with the primary care clinician
I don't want to think that a doctor would see results on an app and still not think about it at all… But I think with time the might lose their skills. Pub6 F<60yrs
I'd love it in my hands. Anything that saves primary care queuing people up in dermatology…I think skin is difficult. And s my feeling is that human error will always exist and I guess if you've got something to support you in making the correc diagnosis in serious skin lesions, then it seems to me like a win-win and a sensible option. GP2 M>45yrs
Most patients with a lesion will go straight to their GP. So I think definitely that's where it's going to be best placed an whether that's a specialist nurse within a GP setting or a GP, certainly within that primary care set up I think is most relevant NP1 F>45yrs
I think selfishly, I'd be disappointed if I was told you don't need your dermatoscope anymore, because it's an area I'm reall interested in and I'm really enjoying… because if it's automated and, therefore, you lose the skill set and you lose motivatio around the topic. NP3 M<45yrs
I would be worried about de-skilling our GPs by giving them a tool that tells them what to do. Derm3 F<45yrs
The great thing would be if they were able to pick up those skin cancers that they haven't thought about, for example amelanotic melanomas, or nodular melanomas, that they don't follow the typical a, b, c, d criteria. Derm3 F<45yrs
If AI had a really, really good dataset of benign lesions, that would give a GP confidence to say 'no, that's benign, that doesn need to be referred in' Derm4 F>45yrs
They [Dermatologists] are very experienced and they don't need a tool. But some of those tools might be needed for th people who refer patients, not to make any mistake at that stage for the early detection… So I can think of those two aspects like speed and a better decision for the practitioner. I see really big opportunities in those kinds of things. AIR2 M<45yrs
AI after the primary care clinician
So, … either a patient-generated image, but preferably taken in primary care on a high-quality camera and sent securely t dermatology for triage, then the AI helping with that triage process. GP3 M<45yrs
In secondary care it would be really helpful, I think particularly for juniors starting out, to have a list of the differentia including the rarer things that could possibly be consistent with the appearances that the algorithm has identified. But givin you ranking of likelihood. Derm2 F>45yrs

Aim of the technology (see Table 3)

Education, screening, or diagnosis

Participants mainly discussed AI technology in the role of a diagnostic tool, however, participants from all groups raised alternative approaches. Patients discussed skin self-monitoring including sequential monitoring of skin lesions over time, and patient education about the 'red flags' of skin cancer. All groups discussed the potential of patient-facing AI technologies to raise awareness and educate patients about skin cancer, including how to perform skin self-monitoring, sun-safe

behaviours, and the risks of excessive sun exposure.

GPs and nurse practitioners also discussed the potential for AI technologies to educate PCPs and improve diagnostic skills. They suggested that a potentially potent educational attribute would be if the AI technology could highlight visual features of skin cancer in images of skin lesions, AI researchers commented that they have developed this type of technology. Dermatologists discussed the potential use of a patient-facing AI technology as a screening tool, particularly in high-risk populations.

Focus on benign or malignant diagnoses

Patients and members of the public stated that the primary aim of the AI technology should be a very high accuracy to avoid missing skin cancers and giving false reassurance, to achieve this the technology would need to have a high sensitivity. All HCP groups commented that aiming to diagnose benign skin lesions might be a safer approach with less risk of missing skin cancers. They proposed that focussing on diagnosing common benign skin lesions such as seborrhoeic keratoses and dermatofibromas might help to reduce unnecessary referrals to secondary care.

Focus on melanoma or all skin cancers

Patients were primarily concerned that any AI technology was accurate for melanoma. Some members of the public were aware of morbidity and mortality from keratinocyte skin cancers and felt that AI technologies should also address these. GPs were primarily concerned about melanoma, because of the mortality associated with it and because it affects younger patients – they felt that in the longer term it could develop to include other types of skin cancer. One nurse practitioner agreed with this, but another wanted the AI technology to address all types of skin cancer and benign skin lesions. Dermatologists commented that diagnosing all types of skin lesions, including all skin cancers, was important, especially for technologies that are designed to be used by patients or healthcare professionals with less experience in diagnosing skin cancer. However, some dermatologists thought the diagnosis of BCCs and SCCs relied on history and macroscopic lesion appearance more than dermoscopic appearance, and hence it could be difficult to train AI to accurately diagnose these skin cancers.

Threshold between malignant and benign – and management of borderline lesions

All groups discussed the difficulty in setting the threshold between benign and malignant lesions for

the AI technology, in essence how to translate the continuous risk score generated by the AI into a binary clinical decision of whether to refer a patient or biopsy a skin lesion. A three-layer management strategy was suggested by a number of participants with further assessment of lesions that are close to the threshold, either through further clinical assessment in primary or secondary care, or through sequential monitoring over time (similar to short-term sequential digital dermoscopy imaging (SDDI) models that already exist in some regions [20]). Nurse practitioners and dermatologists reflected on how this issue demonstrates the complexity of clinical practice, and that clear guidelines will be needed about what to do at each risk level.

Table 3: Aim of the AI technology
Education, screening, or diagnosis
I guess you always take a picture at a point in time and… say you noticed it change in a couple of months could you go bac and say, it's changed from this to this, and almost keeping a progression record … I think that would be a really helpfu feature. Pat4 F<60yrs
Give people recommendations of you should not go out unless you're wearing 50 SPF, we think you should not sunbath between 12 and four, just because it's cloudy does not mean you're not going to burn… put a kid in a sun hat so they don have sunstroke.. Pub3 F>60yrs
I say to them (patients) there are mole apps that you can monitor, and that maybe it's not great in terms of getting a diagnosis but they're more aware and they're keeping attention more to a particular lesion. And definitely it helps to diagnose cancer i the early state than what we saw years ago. GP1 F<45yrs
But the way I used AI was to try to train on some features in the image rather than giving me a diagnosis. For instance, t teach a deep learning algorithm to tell me whether that the borders are regular or irregular, 'cause that's something subjectiv sometimes between readers…. So I used AI to reveal the features rather than giving the full diagnosis. AIR2 M<45yrs
So if you could develop one that was validated it might be useful for selected patients, maybe not the entire population bu particularly high risk patients, maybe patients with lots of moles. Derm2 F>45yrs
Focus on benign or malignant diagnoses
The first aim of the app needs to be to go and get it checked or not, or to go and get it biopsied or not. Pat4 F<60yrs
Just screening out the seborrheic keratosis, the pigmented dermatofibromas, the benign…if it could screen all of those ou which are the vast majority of the two-week-wait referrals that we see, that would be incredibly important for providing better, more targeted service… plus potentially triaging things that are so unlikely to be a cancer that they don't need to b referred. So it kind of works both ways, early detection and reducing the massive numbers of 2-week-wait referrals that w get. Derm2 F>45yrs
Focus on melanoma or all skin cancers
I think melanoma is the most important because I'd like to think that because your differential features for a squamous cell ar quite obvious, that would be referred on anyway. I think it's melanoma and other pigmented lesion differentiation that's reall tricky. So my thought would be more melanoma, and certainly that's what I've used it for. NP1 F>45yrs
People just think it's about diagnosing melanoma, where actually it's not. It's about recognising all skin lesions. You've got t diagnose benign lesions, and you've to diagnose malignant lesions, and it's not just melanomas, you've got BCCs, you've go SCCs, you've got AKs, and then you've got all the benign lesions. Derm1 F>45yrs
Threshold between malignant and benign – and management of borderline lesions
I guess one approach you could take with marginal cases, you could say that we suggest you get re-tested in six months' tim or something like that. So, you have a three-layer band. One you definitely need action, one you definitely don't, and then middle level where you come back for a test after a bit. Pub3 F>60yrs
What I suggest would be good...if you're less than 70 per cent sure, it's an arbitrary...whatever number sure, then there clinician involvement. So take a deeper dive into the history and that person then comes in and is looked at. NP3 M<45yrs
So I think it would be really interesting, if when I looked at a lesion I gave the patient a percentage of how right I thought was or even any diagnosis. I mean, I'd love that, that would be really cool to have that honesty. 'I think you've got a vira chest infection, I'm about 50-50, I'm going to hold off on the antibiotics, but here's some strict safety netting'. Yeah, I don

know whether I'd want to give the patient that information. NP3 M<45yrs
So it would come back with a comment like 'this is 80% likely to be a melanoma, this is 20% likely to be a melanoma' an there would have to be some sort of understanding, some sort of cut off, what is the point at which a referral is merited... An I suppose that one of the dangers… where do you cast the net in terms of risk? Derm4 F>45yrs

Cross-cutting themes (See Table 4)

Trust

Patients and members of the public often raised the issue of trust. The newness of the technology, the involvement of private companies and concerns about data privacy, and the diagnostic accuracy being less than 100% were all felt to be reasons for a lack of trust in AI technology. However, there were also participants who felt patients would trust a consultation with a PCP more if it involved AI technology, even if it had made no difference to the assessment.

GPs worried that patients would start to demand the AI technology be used in consultations, and whether they could rely on benign results from the AI technology. Nurse practitioners felt the opposite, that patients fundamentally trust people more than machines. Dermatologists felt that there is not enough evidence that existing AI technologies are safe and accurate enough to be used in clinical practice, and that it is difficult for patients to determine which patient-facing AI technologies they can trust.

AI researchers had concerns about using AI in medical settings. But they also discussed that the 'black box' nature of AI technologies should not be a barrier to trust as we similarly do not understand the mechanism of action of many medications that we use. They felt that if clinicians recommended or adopted an AI technology then patients would be more likely to trust it.

Ease of use and acceptability

Ease of use was a major concern for patients and members of the public, so that the technology did not place a burden on patients and it was easy for them to take high quality images for the AI to analyse. Patient choice was another factor that several members of the public raised, enabling patients to choose the method of consulting that suits them best. GPs and nurse practitioners' primary concern was that it should be easy to integrate the AI technology with their current computer systems and Wi-Fi systems. Dermatologists, like patients, were concerned about how easy it would be to take images of skin lesions that were of sufficient quality for the AI to analyse. AI researchers commented on how practical limitations of a technology can prevent it being used, even if it is potentially very beneficial, and so these 'real world factors' (such as cost, ease of use, and how well it fits into

clinical workflows) need to be considered.

Accuracy and generalisability

Accuracy was raised frequently by all groups and is referenced in many of the sub-themes throughout the results. The primary concerns were the false negative rate and the risk of false reassurance, and whether the diagnostic accuracy is generalisable to other clinical settings, using different camera technologies, and with other populations/demographics. AI researchers commented that currently AI technologies are often developed on small datasets, which are not representative of the general population and may contain errors and biases, and hence do not generalise well to be accurate in all sections of the population. Members of the public and HCPs were concerned this might mean AI technologies would be less accurate in melanin-rich skin or in rarer clinical presentations including rarer skin cancers.

AI researchers discussed how this situation might be improved, with close collaboration between clinical and AI researchers, and to really focus on data quality in a ‘data-centric’ approach. An AI researcher from a commercial background commented that they have found it hard to collect a representative dataset to use for AI development and testing, their app had significantly lower uptake amongst patients with melanin-rich skin and there are fewer publicly available images of skin lesions in melanin-rich skin.

Evaluation and regulation

Patients felt that AI technologies should be evaluated by a mix of professionals before it can be adopted, including confirmation of any diagnostic accuracy claims. Patients were concerned that AI technology would be adopted based on novelty and hype or because it is cheaper than clinicians’ time, when it is not in the best interests of patient and their clinical care. HCPs felt that a large amount of data from clinical trials would be needed to evaluate AI technologies but recognised that this might take time. AI researchers were concerned that we currently lack good measures to evaluate AI algorithms for use in clinical settings. They stated that the current practice of using simple diagnostic accuracy measures (e.g. sensitivity, specificity) is probably not comprehensive enough to demonstrate the accuracy, benefits and risks of AI technologies. AI researchers added that, whilst clinical studies, due diligence, and understanding the biases and flaws in an AI system are all important aspects of evaluation, there are other factors including business sustainability that need to be considered before a decision can be made to adopt an AI technology. Many groups commented on

the need for health-economic evaluation as part of any evaluation programme.

Patients and members of the public often disclosed that they did not understand regulatory processes. Some felt it was important that AI technologies had a CE marking and was evaluated by a national body. Healthcare professionals were concerned about regulatory processes for AI technologies not being as robust as for medicines and treatments. GPs raised concerns about where the medicolegal responsibility for errors related to the use of AI technologies would lie. Dermatologists wanted the use case for AI technologies to be made clear, and for the regulator to assess the technology based on this use case, but acknowledged that regulating AI technologies is challenging. An AI researcher commented that regulatory processes for AI technologies are becoming more complex.

Implementation in the UK NHS

Members of the public and GPs discussed variation in NHS infrastructure in different regions of the UK, including wireless connectivity, and how this could make implementation difficult.

Most participant groups commented on the capacity of the healthcare system to implement a new technology in the current climate where resources are strained. Current pressures may mean that PCPs are unable to take the time to understand and implement a new technology. Some HCP participants also highlighted that an AI technology with a low specificity might lead to a significant increase in referrals to secondary care and worsen workload pressures. Conversely, some nurse practitioners and dermatologists commented that use of AI technology could help to reduce the number of referrals to secondary care and AI researchers hoped AI technologies could be used alongside PCPs and/or dermatologists to ease workload pressures.

All groups commented on the importance of professional bodies in the implementation of new technologies. Patients and HCPs felt that implementation would be greatly helped if professional bodies such as the National Institute for health and Care Excellence (NICE), NHS England, Integrated Care Boards (ICBs), the British Association of Dermatologists (BAD), or the Medicines and Health products Regulatory Agency (MHRA) had recommended or evaluated it. GPs and nurse practitioners believed any decision to adopt and fund AI technology will need to come from a higher body than individual GP practices. AI researchers commented on the importance of the views of professional bodies for the adoption of AI technologies, and that these are variable and country specific.

Nurse practitioners and Dermatologists highlighted the need for adequate training in how to use an AI technology as part of the implementation process and clear guidelines on how the AI technology should be used and interpreted.

Long-term use

The unique potential for AI technologies to continue learning after implementation and for diagnostic performance to improve over time was raised by many participant groups. GPs and patients assumed that this feature would be standard practice and that regular updates would sequentially improve the AI technologies performance as it learnt from new data. Few participants commented on how this process could be managed in practice.

A major concern of all groups was how to check that the AI technology was performing accurately. AI researchers were particularly worried about the potential for the performance of the AI technology to deteriorate over time, which they referred to as 'drift in performance'. They commented that this could occur because of a change in the way the AI technology is used, a change in the population (for example employing the technology in a population with lower skin cancer prevalence or different demographics compared to the development and testing datasets), a change in the hardware, or a change in the accuracy of the technology over time. All groups suggested 'sanity checks' that could be used regularly to help pick up if an AI technology was not performing accurately or if there has been a deterioration in its performance. These 'sanity checks' included: expert systems monitoring image metrics over time and comparing the prediction of the new AI technology to the previous AI gold standard; a clinician reviewing all cases that the AI diagnoses as "likely skin cancer" or where the technology had low confidence in the diagnosis; or limiting the use of AI technologies as with a 'human-in-the-loop' approach. AI researchers added that there needs to be an incentive of the makers of the AI technology to maintain and provide ongoing support long-term.

Table 4: Cross-cutting themes
Trust
The fact that it hasn't been the norm within the healthcare setting, I think will make quite a lot of people fee uncomfortable. So the fact that it's just so up and coming and new will naturally spark a bit of anxiety in terms of people an they think about if it's safe, if it's got the same principles and standards in place in terms of care, safety, confidentiality e cetera. Pat1 M<60yrs
I think there is also a section of the population who are very suspicious of data not being used correctly. So, you would nee to have some sort of reassurance about correct data use as well. Pub4 F>60yrs
I think people fundamentally put trust in people and we're still wary about putting trust in machines, or computers, becaus

we've been fed sci fi for years that makes us worry, and also I do think that we still think we know best, even though th algorithms that will be in the computer will be absolute gold standard, and they won't be tired, hung over, jaded, they'll b right every time. NP3 M<45yrs
I think the concern at the moment is that there isn't enough evidence that any of these machines, or machine learning is u to speed to be able to make diagnoses without missing any skin cancer, and that includes rare skin cancers, skin cancers wit rare presentations, or in different ethnicities. Derm3 F<45yrs
Yeah. I'm just really unsure in how far, to be honest, machine learning can or will take over a medical setting, or should. Th more I see and read, the more I'm getting also suspicious. AIR3 F<45yrs
I would question the premise that being a black box is actually a terrible thing or even a novel thing, because clinicians ar black boxes. As far as I understand for most medicines we have no idea why they work, and we still use them becaus statistically it works. And as long as we regularly check that our models work statistically then who cares whether that's chemical or a computational black box. Obviously, it's nicer if you can also present the clinician with some data that th clinician then can use to make further judgement calls.. AIR1 M<45yrs
As a field we're only beginning to scratch the surface of what that means, to trust technology. If say my physician tells me this is a good way to understand more about your skin conditions, there's some element where I trust he or she as professional and the information relayed is therefore accurate, and therefore that trust extends to the thing they ha suggested. AIR5 M<45yrs
Ease of use and acceptability
So, being accurate, being honest, easy to use, easy to understand. And easy for the GPs to use as well, the other healthcar professionals who've got to get the information through that as well, easy for them. Because they're going to have to look a whatever it's come up with and try and make some sense of that before they actually sit down in front of us. Pub1 F>60yrs
For me patient choice is so important – so both have a machine in the GP surgery for those who would prefer to do that an have an app for the people who are quite comfortable using those. Pub1 F>60yrs
I suppose if it was a lengthy process. So if there was dodgy Wi-Fi or it's hard to get your image or it's taking time to upload... suppose it's just the practicalities and the ease of the software and hardware that you're using, those will be barriers, if didn't work. NP1 F>45yrs
How easy is it to take the photograph? I think that's really important. Because I get loads of referrals with photographs, bu the photographs are completely useless, they're blurred and a complete waste of time. Somebody's ticked the box and said sent a photograph, but they might as well not be there. Derm1 F>45yrs
If you have a technology that's extremely beneficial for a certain disease but it's really expensive, it takes a really long time and it's hard to use then no one will ever use it. So there's real world factors there. AIR5 M<45yrs
Accuracy and generalisability
I think the key issue is generalisability, because algorithms are developed on a very small set which have very specifi properties and there is biases. And then of course, they do not work on a wider range of other images from other scanner from other countries … this is a huge issue, the generalisability. You have all these algorithms that achieve higher numbers i one setting but it doesn't mean that they will work well on new data they have not seen. AIR3 F<45yrs
They need to make sure that there is a proper diversity of people in there. Because AI is only ever going to be as good a what you use to train it. So I think they need to be particularly careful about diversity of skin colour and capturing the wid variety that's needed. Pub3 F>60yrs
For us it's a kind of a chicken and egg problem. So, there's not a lot of data for darker skin available, so training on that...an especially proving the accuracy on dark skin it's almost impossible... We don't have the users, we don't get the data, w don't get the proof of how good we are on darker skin – so that definitely is a loop that we need to break at some point AIR4 M>45yrs
Evaluation and regulation
I have come across many issues because algorithms have been evaluated with measures that are actually not measurin what you would want or need in a clinical setting... So I actually think the most important thing right now would be to set u proper evaluation schemes and to think about how deep learning models should be or can be properly evaluated... The rea question is, how do we evaluate them to make sure they will work in a medical clinical setting. AIR3 F<45yrs
So this is definitely a concern, not to send too many users into healthcare. The first concern's accuracy, we don't want t miss too many skin cancers, for sure, but the second concern is definitely also the health economic case. So, we can c

course send a huge amount of people into healthcare. We find more skin cancer, we reach our goals of finding more ski cancer, that's fine. But we don't solve a real problem, we make it worse for the healthcare system as well. AIR4 M>45yrs
I worry a little bit that if it's something that is a manageable cost it might be overused. It's a way to shift a number o patients who you might otherwise see… They're always on an efficiency drive, they're always under pressure to sav money… I don't know whether, if the technology was affordable for GP surgeries, they would just think, oh great… we'r short of doctors we can just process people through the hands-off routine…you stop thinking about it in clinical terms an you start thinking about it in financial terms. Pat2 F<60yrs
I think that's really important that it has CE marking; because I see a lot of apps and crap in the digital space that a lot o people are buying and spending money on and they're not evidence-based medicine. Pub2 F<60yrs
I know it [regulation] is fairly patchwork. Obviously, when it comes to medication and prescribing, there's fairly robus regulatory systems… When it comes to certain equipment, some patient devices it's a bit piecemeal. But I would be fairl reassured if the MHRA covered this equipment. GP3 M<45yrs
I've been indoctrinated by the British Association of Dermatologists. And what they explained. They gave a positio statement, I think it was a year or two ago, and we were've told these are medical devices and therefore they need to g through these medical regulatory agencies. I think before Brexit it obviously would have been Europe as well and now I thin it's MHRA should be responsible for this. Which makes sense because obviously it's a very important thing for patients an for doctors, medicolegally as well. So it needs to work for its purpose, for what they say it's going to be doing. Derm F<45yrs
And also where does it stand legally? If you come to me and you show me a mole and I do my AI thing and I say to yo computer says no, and off you go and you keep doing the things that you do, and then two years later you come back to m with a large black blobule, sentinal lymph node biopsy positive…is that my fault? Is that the AI's fault? And if it's the AI fault, who are you suing? GP6 F>45yrs
Implementation in the UK NHS
I think when you're an external looking in at the NHS it's just a monster beast to try and understand different parts of it. Overall I think integration of digitals is great, but it's not uniform, and also there are massive connectivity issues still in som areas… Some parts of the country are doing amazing things, and other parts are so backward. Pub2 F<60yrs
I'm just very much aware that the basics of general practice is under so much pressure that adding a new technology an complication is not everybody's first priority… I appreciate a lot of these new technologies can save time and resources i the long term, but certainly, if I'm thinking now of the start of a difficult winter, it's going to be a tough sell right now. GP M<45yrs
One of the things I came across is that there was a shortage of dermatologists in general. This means that the practitioner i expected to see a lot of patients in the same day… the practitioner is a human, maybe some stress, or he had a bad day, h might not spot some feature in that image. So those [AI] tools can be something that fixes this gap if in that day he had a ba day and didn't notice some features. AIR2 M<45yrs
So it's going to cost something to have this kind of level of equipment, but there's not an infinite amount of money in G land, if they're spending on that then it's coming out of somewhere else, which might mean we won't have the money o equipment for something that may benefit patients. NP2 F<45yrs
It's got to come with the support and the training along with it, it's not just about getting a new piece of kit… I think th important thing is that that cost incorporates training and a clear understanding of the device and what it's used for, an why, and how to use it. It's silly to give a piece of kit and then not have the support to know how to use it best. NP1 F>45yrs
Long-term use
In some ways the accuracy would build as the technology was used. So the longer it's been in play the more you can depen on it probably… I assume that you don't just test the heck out of it and say it's fine now and then stop refining it. Pat F<60yrs
Let's say the model was trained in a way it works really well when you are taking dermatoscope photos, and it gives you really good sense of the malignancy potential there and the studies support this. Then a patient sends you a photo… and thi photo was taken using a smart phone with different kinds of lighting conditions, in a setting where it was actually neve designed for use but that fact is not obvious to anyone who hasn't been thinking about this for a really long time. So wit the best of intentions, I think you could still have a situation where the performance starts drifting away from how it wa designed for use. AIR5 M<45yrs

Maybe scanners will change… It is not clear that if an algorithm works well on the scanner of the last generation, even fron the same company, [if] you get a new one that it would work on that as well… They [scanners] will not stay the same, bu then we really should think about how do we integrate new modalities into the machine learning models, because it's no sustainable to develop something on a fixed dataset. AIR3 F<45yrs
It should be tested before it gets to the GP surgery, and then it should be checked… at regular intervals, to make sure it's sti working correctly. Because if you start falsely diagnosing people, that could be a complete waste of time. Pat3 F>60yrs
If we look at a top-level view across all AI it's really hard to summarise in a single sentence whether this will help or hurt. Bu if developed well, if used well I still do believe there's lots of great potential here. AIR5 M<45yrs

Discussion

Summary

In this interview study a large proportion of the discussions centred around two simple questions: who is going to use the AI technology and what is it going to do. Most participants commented on positioning AI technology before the PCP, with patients or allied healthcare professionals, or with the PCP as a decision support tool. Few participants proposed positioning the technology in secondary care, either to triage referrals to secondary care or as a decision aid for dermatologists. Participants highlighted several overarching topics as important to them, including trust, acceptability, generalisability, evaluation, regulation, implementation, and long-term use.

The risk of false negatives results from the AI technology (i.e. a benign result from the AI technology when the lesion is actually malignant) was a major concern for all participant groups. Poor sensitivity of an AI technology could lead to false negatives and false reassurance of patients and PCPs. Missed skin cancers are likely to lead to late diagnosis and increased morbidity and mortality for patients. One potential benefit of AI technology mentioned by all groups was reduction in the workload of healthcare services, primarily through effective triage of patients that need to see PCPs and/or specialist clinicians. This could be the case if the AI technology has a good specificity with few false positive diagnoses. However, false positives will inevitably occur with a diagnostic technology, and there is a risk that an AI technology with a low specificity would result in high levels of false positives and could significantly increase the workload of healthcare services. This was a concern raised by healthcare professionals, but few other participants groups commented on this risk. Poor specificity of an AI technology could also contribute to overdiagnosis of melanoma. It has been suggested that overdiagnosis of melanoma is rising due to increased rates of skin examination and decreased thresholds for biopsying skin lesions and for labelling morphological changes on histopathological examination as malignant.[21] Implementation of an AI technology with low

specificity could increase both rates of skin examination and biopsy rates, both of which could contribute to overdiagnosis.

AI researchers were the most pragmatic, expressing concerns about the generalisability of many AI technologies and the relevance of current testing approaches in preparation for clinical implementation. They were also concerned about the robustness of the datasets underlying AI technologies, including the skin cancer prevalence and the breadth of patient demographics included in the datasets. They felt that, at the current time, these technologies need to be implemented with a human-in-the-loop (PCP/specialist clinician). Many HCPs were aware of the lack of evidence and the potential risk to patients that have been publicly commented on by the BAD.[6] Patients and public were aware of potential improvements in patient access, diagnostic accuracy, and reduced workload AI/ML technologies could offer, but were also concerned about the risk of missing cancers, and about losing the opportunity for human interaction and to ask questions.

Complexity underpins most of the generated themes. Developing, implementing, evaluating, regulating, and maintaining an AI technology in healthcare settings are all multi-faceted, complex processes containing many opportunities for error. It is therefore unsurprising that so many diagnostic technologies fail at the implementation stage.[22] The complexity of implementing diagnostic tests and the multiple stages involved are laid out in the CanTest framework.[9] Several studies have discussed the complexity of implementing digital, AI and ML interventions into healthcare settings, and recommended a whole-of-system approach with particular focus on how users interact with devices and user training.[7,23]

Comparison with existing literature

The NASSS framework [10] was chosen to guide this study because it includes a wide range of domains which help capture the complexity of implementing healthcare interventions.[24] Most of the cross-cutting themes link closely to a few NASSS domains or sub-domains, specifically ‘knowledge to use the technology’, ‘demand-side value [to patient]’, ‘adopters’, and ‘organisations’. There were several themes raised in discussions which did not fit comfortably into the NASSS framework (such as the potential for continued learning with AI technologies), as well as themes (such as trust and acceptability) which seemed to lose some of their breadth and nuance by being

contained within the NASSS domains. Our choice to adopt both an inductive and deductive approach to coding allowed us to capture participants' views, and to build knowledge about application of the NASSS framework in different settings and with novel technologies.

A recent Swedish study used an AI-based melanoma decision support aid for clinicians as an example technology to generate discussions with participants about implementation of decision aids. In keeping with their findings many participants in our study discussed the issues of accuracy, safety, data security, liability, ease of use and integration.[25] Our findings also support those of a recent systematic review on the use of ML-based risk prediction models in healthcare settings[8] in which participants demonstrated largely positive views of AI/ML technologies, but identified many barriers. We found that AI researchers generally had more negative views compared to other participant groups, mostly based on concerns about the generalisability of models and hence assuring their safety in clinical practice. We identified concerns about diagnostic accuracy, risk of patient harm, ease and speed of use, HCP over-reliance on the technology, legal liability, data protections, data quality, impersonality, and positioning in the diagnostic pathway; which echoed findings from several recent studies.[26-29] In particular, we identified aspects of the consultation participants felt that AI technologies could not replicate, namely the need for human touch, and clinical experience and judgement. Participants also commented on the risk AI technologies will not be effective in minority populations who are inadequately represented in training and testing datasets.

The wide variety in positioning and approach of existing AI technologies[4,5] indicate that the optimal position and approach has yet to be determined. Several participant groups commented on this and highlighted that they wanted to see more evidence of the accuracy of technologies in real-life clinical use. This fits with recent reviews of AI technologies aimed at detection of skin cancer, [4,5,29] although increasing evidence from clinical trials is emerging.[11] PCPs were keen to have access to an AI technology which could support their diagnostic decision making, in keeping with findings from a previous study.[30] AI researchers highlighted a growing body of research in AI technology development for healthcare settings, including the 'human-in-the-loop' approach and the 'data-centric AI movement'.[7,31,32] Regulation was a topic raised by several participant groups. Established regulatory frameworks are trying to catch up with the rapid developments in AI technologies, there are also emerging national and international regulatory policies including the European AI Act, the USA AI Bill of Rights, and the developing UK policy on AI.[33-35]

Strengths and limitations

To the authors' knowledge, this is the first study to report the views of patients and members of the public, in addition to a variety of healthcare professionals and AI researchers about the use of AI technologies to facilitate the earlier diagnosis of skin cancer in primary care settings. We consulted a wide variety of stakeholders and had good variation among interviewees in terms of background, age, sex, and geographical location. The study benefitted from patient and public involvement at every stage, and a strong conceptual framework was used to develop the framing of the interview schedule and the initial data coding stages.

Aiming to recruit a wide range of stakeholders was a conscious choice as we felt it was important to recruit a wide range of participant groups to achieve a breadth of opinions; the trade-off for this was that time and resources meant we were only able to include limited numbers in each participant group. The aim was to achieve breadth of opinion without necessarily achieving saturation. AI and clinical implementation are complex subjects, which meant we were more likely to recruit participants with higher educational attainment, and who are engaged with healthcare research and/or AI: both these aspects may have affected the balance of views we obtained in this study. Participants were also mostly white British or white European. Whilst skin cancer is more common in melanin-poor skin, a key limitation of current AI technologies is their lack of training and testing on skin lesions in patients with melanin-rich skin. Data on participant views on this issue may be incomplete due to the limited numbers of non-white participants. In contrast to other participants AI researchers were largely non-UK: this was inevitable as the majority of commercial companies developing these technologies are based outside of the UK. Nonetheless, it meant their knowledge of UK clinical practice and diagnostic pathways was more limited compared to other participant groups.

Implications for clinical practice, research, adoption, and policy

Healthcare services are currently working under extreme pressures in primary and secondary care. [36] AI technologies aimed at the diagnosis and triage of skin cancer could facilitate the early diagnosis of skin cancer to improve outcomes for patients and potentially help ease some of these pressures. However, before this can happen, research is required to prove their efficacy with real clinical populations and to address the questions that remain about the most effective positioning of the technologies in the diagnostic pathway and the optimal approach for them to take. Better measures of clinical performance are required to inform these studies, that consider not only diagnostic accuracy but also provide a measure of generalisability and dataset quality. Diagnostic

technologies that are employed in populations that are different to those they were developed and tested in are prone to spectrum bias.[37]

There are aspects of our findings which could be used to further develop the NASSS framework, for example to consider how users interact with the AI technology in more depth and the potential for continued development after implementation. When developing an AI technology aimed at the diagnosis or triage of skin cancer, developers need to consider carefully and be specific about the intended use case of the technology, including where it will fit into the diagnostic pathway for skin cancer and the approach that it is going to take. Developers should also consider the quality and representativeness of the data they use to develop the AI.

The decision to adopt an AI technology is complex and multi-faceted. Clear regulatory processes need to be established which consider unique features of AI technologies, including continued learning, to ensure AI technologies are safe and effective when used in clinical settings. Adopters should also consider what safety nets are in place to identify poor performance and prevent false negative results, such as expert systems and regular ‘sense checks’. Policy is needed to support development, regulatory processes, and adoption.

Conclusions

AI technologies are currently being designed with a wide variety of approaches, and the optimal approach and position in the skin cancer diagnostic pathway for these technologies have not yet been determined. AI technologies have great potential to help detect and diagnose skin cancer, to improve patient experience and outcomes, and to reduce workload of over-stretched healthcare systems. However, we have identified important concerns surrounding trust, acceptability/usability, generalisability, evaluation, regulation, implementation, and long-term use. These technologies need to be developed carefully and thoughtfully to achieve their potential, guided by evidence-based approaches and appropriate implementation, taking into consideration long-term sustainability and safety.

Acknowledgements

The research was supported by a Cancer Research UK Cambridge Centre Clinical Research Fellowship for OTJ. JE is supported by an NHMRC Investigator Award APP1195302. RNM is Chair of the AI Working Party Group at the British Association of Dermatologists. The funding sources had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit for publication. The authors would like to thank Professor Steve Morris who contributed to the development of the study.

Conflicts of Interest

We declare no competing interests.

Abbreviations

AI – Artificial Intelligence

ML – Machine Learning

NASSS - The framework for the non-adoption, abandonment, and challenges to scale-up, spread, and sustainability

KC – Keratinocyte cancer

SCC – Squamous Cell Carcinoma

BCC – Basal Cell Carcinoma

GP – General Practitioner

NP – Nurse Practitioner

HCP – Healthcare Professional

PPI – Patient and Public Involvement

PCDS – Primary Care Dermatology Society

GPwER – General Practitioner with Extended Role

PCP – Primary Care Practitioner

NHS – National Health Service

USC – Urgent Suspected Cancer (referral pathways)

SDDI - short-term sequential digital dermoscopy imaging

CE marking – Conformite Europeenne marking

NICE - National Institute for health and Care Excellence

ICB – Integrated Care Board

BAD – British Association of Dermatologists

MHRA - Medicines and Health products Regulatory Agency

Multimedia appendices

Appendix 1: Interview schedule for patients and members of the public

Appendix 2: Interview schedule for healthcare professionals and AI researchers

References

1. C Garbe, U Keim, S Gandini, *et al.* Epidemiology of cutaneous melanoma and keratinocyte cancer in white populations 1943–2036. Eur J Cancer 2021; 152: 18–25.
2. Cancer Research UK. Non-melanoma skin cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/non-melanoma-skin-cancer (accessed 25th April 2024)
3. Cancer Research UK. Melanoma skin cancer survival statistics. https://www.cancerresearchuk.org/about-cancer/melanoma/survival (accessed 25th April 2024)
4. D Wen, A Coombe, M Charalambides, O Jones, N Wietek, J Dinnes and R Matin. DS03 A scoping review of commercially available artificial intelligence-based technologies for diagnosis or risk assessment of skin cancer. 102nd Annual BAD Meeting 2022 187(Suppl. 1), pp158–172
5. OT Jones, RN Matin, M van der Schaar, K Prathivadi Bhayankaram, CKI Ranmuthu, *et al.* Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review. Lancet Digital Health 2022; 4:e466-76
6. British Association of Dermatologists Position Statement on Artificial Intelligence (AI) Interventions. BAD Position Statement v2 Feb 2022. https://www.bad.org.uk/clinical-services/artificial-intelligence/ (accessed 25th April 2024)
7. D Lyell, Y Wang, E Coiera, F Magrabi. More than algorithms: an analysis of safety events involving ML-enabled medical devices reported to the FDA, *Journal of the American Medical Informatics Association*, Volume 30, Issue 7, July 2023, Pages 1227–1236, https://doi.org/10.1093/jamia/ocad065
8. R Giddings, A Joseph, T Callender, SM Janes, M van der Schaar, *et al.* Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review. The Lancet Digital Health (2024): 6;2:E1313-E144
9. FM Walter, MJ Thompson, I Wellwood, *et al.* Evaluating diagnostic strategies for early detection of cancer: the CanTest framework. BMC Cancer 2019; 19: 586.
10. T Greenhalgh, J Wherton, C Papoutsi, J Lynch, G Hughes, *et al.* Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *J Med Internet Res 2017;19(11):e367.* DOI: 10.2196/jmir.8775
11. SW Menzies, C Sinz, M Menzies, SN Lo, W Yolland, *et al.* Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial. Lancet Digit Health 2023; 5: e679–91
12. National Institute for Health & Care Excellence (NICE) Early Value Assessement (EVA): Artificial Intelligence

technologies for assessing skin lesions selected for referral on the urgent suspected cancer pathway to detect benign lesions and reduce secondary care specialist appointments. Expected Publication Date 10th April 2024. https://www.nice.org.uk/guidance/indevelopment/gid-dg10086 (accessed 25th April 2024)

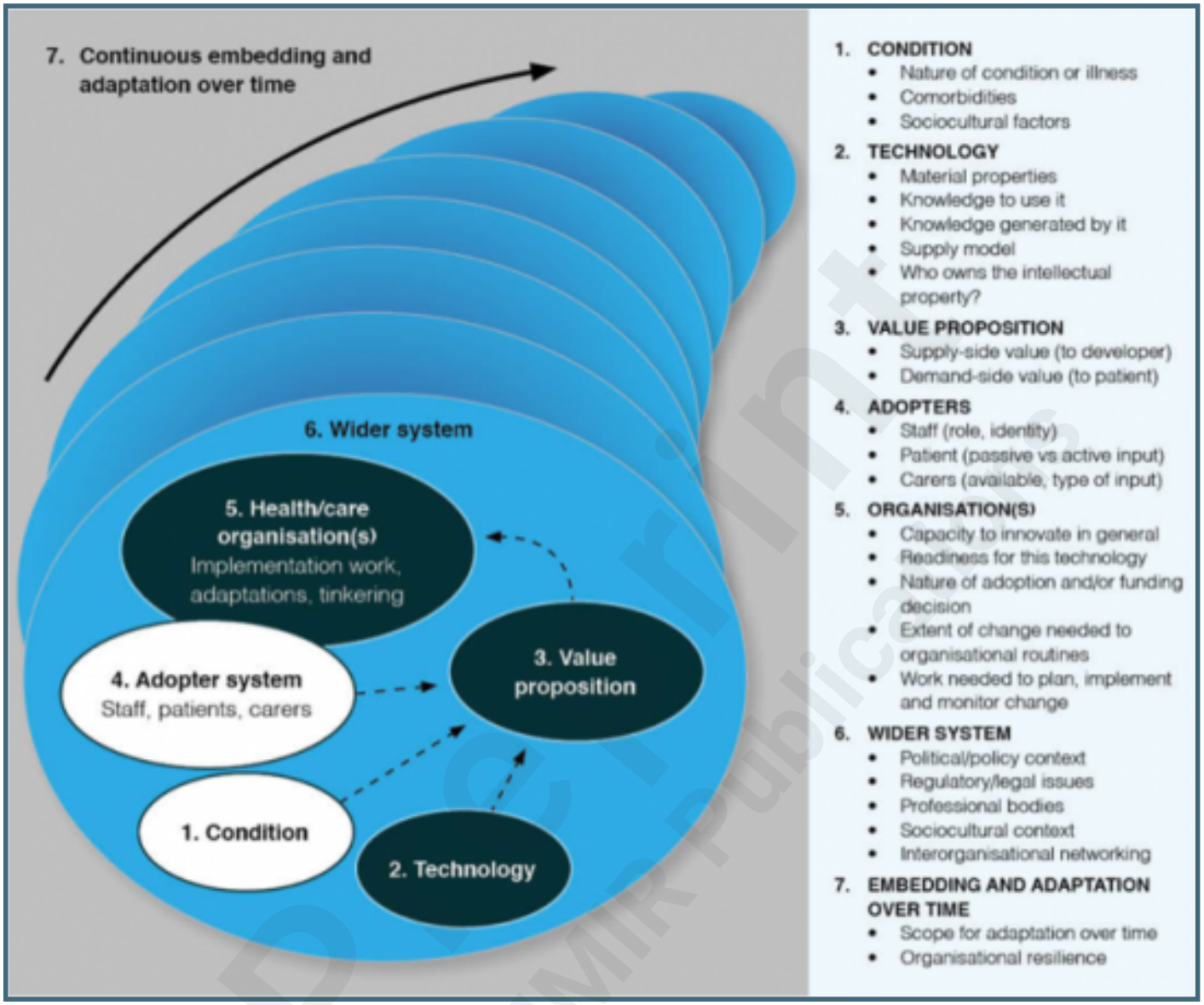
13. Melanoma Focus. https://melanomafocus.org/ (accessed 25th April 2024)
14. Primary Care Dermatology Society (PCDS). https://www.pcds.org.uk/ (Accessed 25th April 2024)
15. Sermo Market Research. https://www.sermo.com/resources/business-market-research/ (Accessed 25th April 2024)
16. British Association of Dermatologists (BAD). https://www.bad.org.uk/ (Accessed 25th April 2024)
17. NK Gale, G Heath, E Cameron, *et al*. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* **13**, 117 (2013). https://doi.org/10.1186/1471-2288-13-117
18. Lumivero (2023) *NVivo* (Version 14) www.lumivero.com
19. Microsoft Corporation. Microsoft Excel [Internet]. 2018. Available from: https://office.microsoft.com/excel
20. SW Menzies, J Emery, M Staples, S Davies, B McAvoy, *et al*. Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial. 2009 BJD 2009;161;6:1270-1277. https://doi.org/10.1111/j.1365-2133.2009.09374.x
21. H Gilbert Welch, BL Mazer, AS Adamson. The Rapid Rise in Cutaneous Melanoma Diagnoses. N Engl J Med 2021;384:72-79)
22. JY Verbakel, PJ Turner, MJ Thompson,*et al*. Common evidence gaps in point-of-care diagnostictest evaluation: a review of horizon scan reports. BMJ Open 2017;**7**:e015760. doi:10.1136/ bmjopen-2016-015760
23. J Nadav, A-M Kaihlanen, S Kujala, E Laukka, P Hilama, *et al*. How to Implement Digital Services in a Way That They Integrate Into Routine Work: Qualitative Interview Study Among Health and Social Care Professionals. J Med Internet Res 2021 | vol. 23 | iss. 12 | e31668
24. PE Plsek, T Greenhalgh. Complexity science: The challenge of complexity in health care. Br Med J 2001 Sep 15;323(7313):625-628 [FREE Full text] [doi: 10.1136/bmj.323.7313.625] [Medline: 11557716]
25. A Frisinger, P Papachristou. The voice of healthcare: introducing digital decision support systems into clinical practice - a qualitative study. *BMC Prim. Care* **24**, 67 (2023). https://doi.org/10.1186/s12875-023-02024-6
26. C Buck, E Doctor, J Heinrich, J Buck C, Doctor E, *et al*. General Practitioners' Attitudes Toward Artificial Intelligence–Enabled Systems: Interview Study. J Med Internet Res 2022;24(1):e28916. doi: 10.2196/28916 PMID: 35084342 PMCID: 8832268
27. S Haggenmuller, E Krieghoff-Henning, T Juts, N Trapp, L Kiehl, *et al*. Digital Natives' Preferences on Mobile Artificial Intelligence Apps for Skin Cancer Diagnostics: Survey Study. JMIR Mhealth Uhealth. 2021;9;8:e22909

28. TB Jutzi, EI Krieghoff-Henning, T Holland-Letz, JS Utikal, A Hauschild, *et al.* (2020) Artificial Intelligence in Skin Cancer Diagnostics: The Patients' Perspective. Front. Med. 7:233. doi: 10.3389/fmed.2020.00233
29. P Venkatesan. Artificial intelligence and cancer diagnosis: caution needed. Lancet Oncol 2021. Published Online September 9, 2021 https://doi.org/10.1016/S1470-2045(21)00533-7
30. R Samaran, JM L'Orphelin, B Dreno, C Rat, A Dompmartin. Interest in artificial intelligence for the diagnosis of non-melanoma skin cancer: a survey among French general practitioners. *Eur J Dermatol* 2021; 31(4): 457-62 doi:10.1684/ejd.2021.4090
31. Suzanne Bakken, AI in health: keeping the human in the loop, Journal of the American Medical Informatics Association, Volume 30, Issue 7, July 2023, Pages 1225–1226, https://doi.org/10.1093/jamia/ocad091
32. D Zha, ZP Bhat, K-H Lai, F Yang, X Hu. Data-centric AI: Perspectives and Challenges. arXiv:2301.04819. https://doi.org/10.48550/arXiv.2301.04819
33. Proposal for a regulation of the European parliament and of the council: laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. COM/2021/206 final. European Commission, Directorate-General for Communications Networks, Content and Technology. Published 21st April 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 (accessed 25th April 2024)
34. Blueprint for an AI Bill of Rights: Making Automated Systems Work ForThe American People. White House Office of Science and Technology Policy. October 2022. https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (accessed 25th April 2024)
35. AI regulation: a pro-innovation approach. Policy paper. Department for Science, Innovation and Technology and Office for Artificial Intelligence. Published 29 March 2023. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach (accessed 25th April 2024)
36. British Medical Association. An NHS under pressure. https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/an-nhs-under-pressure (accessed 25th April 2024)
37. JA Usher-Smith, SJ Sharp, SJ Griffin. The spectrum effect in tests for risk prediction, screening, and diagnosis BMJ 2016; 353 :i3139 doi:10.1136/bmj.i3139

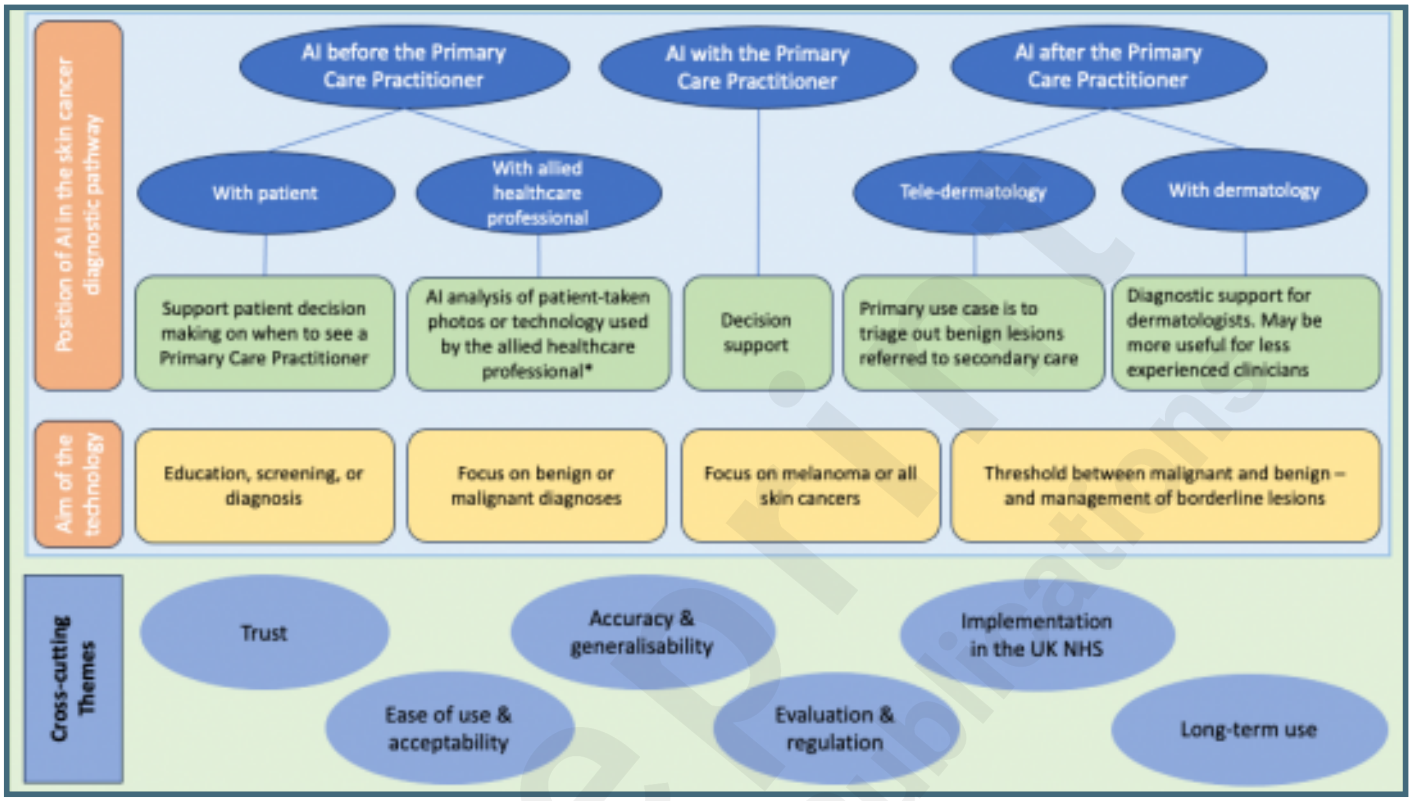
Supplementary Files

Figures

The NASSS Framework[10].



Issues around the design and intended positioning of Artificial Intelligence (AI) technologies for the diagnosis of skin cancer that were identified from the data, and further cross-cutting themes identified Key: AI = Artificial Intelligence, *allied healthcare professionals that could use the technology include: pharmacists, healthcare assistants, practice nurses, podiatrists, hairdressers, and potentially many more, UK NHS = UK National Health Service.



Multimedia Appendixes

Interview Schedule for patients & members of the public.
URL: http://asset.jmir.pub/assets/a4f40d84c4ae1bd80553bab0d8d2338c.doc

Interview Schedule for HCPs & AI researchers.
URL: http://asset.jmir.pub/assets/023dd3957ddb61cc74fc99a11010dec5.doc