NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Early value assessment guidance consultation document

Artificial intelligence (AI) technologies for assessing and triaging skin lesions within the urgent suspected skin cancer pathway: early value assessment

Guidance development process

Early value assessment (EVA) guidance rapidly provides recommendations on promising health technologies that have the potential to address national unmet need. NICE has assessed early evidence on these technologies to determine if earlier patient and system access in the NHS is appropriate while further evidence is generated.

The diagnostics advisory committee has considered the evidence and the views of clinical and patient experts.

This document has been prepared for public consultation. It summarises the evidence and views that have been considered, and sets out the recommendations made by the committee. NICE invites comments from registered stakeholders, healthcare professionals and the public. This document should be read along with the evidence (an EVA report).

The advisory committee is interested in receiving comments on the following:

- Has all of the relevant evidence been taken into account?
- Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- Are the recommendations sound, and a suitable basis for guidance to the NHS?

Equality issues

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the recommendations may need changing to meet these aims. In particular, please tell us if the recommendations:

- could have a different effect on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology
- could have any adverse effect on disabled people.

Please provide any relevant information or data you have about such effects and how they could be avoided or reduced.

Note that this document is not NICE's final guidance on artificial intelligence (AI) technologies. The recommendations in section 1 may change after consultation.

After consultation, NICE will consider the comments received. The final recommendations will be the basis for NICE's early value guidance.

Key dates:

Closing date for comments: 25 September 2024

1 Recommendations

1.1 More research is needed on using 2 artificial intelligence (AI) technologies in teledermatology services to assess and triage skin lesions in people within the urgent suspected skin cancer pathway, before they can be used in the NHS. The technologies are:

- Deep Ensemble for Recognition of Malignancy (DERM)
- Moleanalyzer pro.

1.2 Access to the technology should be through company, research, or non-core NHS funding, and clinical and financial risks should be appropriately managed.

What research is needed

1.3 More research is needed on:

- how accurate AI technologies used in teledermatology services are at detecting cancer and non-cancer skin lesions compared with teledermatology services alone
- how accurate AI technologies are at detecting non-cancer and cancer skin lesions in people with black or brown skin
- the effect of using AI technologies in teledermatology services on the number of referrals for face-to-face dermatology appointments compared with teledermatology services alone
- the proportion of lesions referred from primary care that would be eligible for assessment by automated AI technologies used in teledermatology services and by teledermatology services alone.

Research should compare use of the AI technologies with and without healthcare professional review.

The research plan gives further information on the prioritised evidence

gaps and outcomes, ongoing studies and potential real-world data sources. It includes how the evidence gaps could be resolved through real-world evidence studies.

Why the committee made these recommendations

Staff shortages in dermatology services and an increasing number of urgent referrals for suspected skin cancer are causing delays in diagnosing and treating all skin conditions in secondary care. AI technologies (DERM and Moleanalyzer pro) could be used within a teledermatology service (a secondary care service that uses digital images to remotely assess skin conditions) to identify non-cancer skin lesions that do not need a further face-to-face dermatologist assessment. DERM can be used alone (automated) or with a virtual review by the company's dermatologists (a second read). Moleanalyzer pro is designed to be used only with dermatologist review.

Some of the clinical evidence on using automated DERM in teledermatology services raises concerns about the risk of missed or delayed cancer diagnoses. Using DERM with a second read could reduce this risk, but it is uncertain if this approach would help increase staff capacity in dermatology services. It is unclear whether DERM (either automated use or with a second read) is more accurate at detecting non-cancer skin lesions than using teledermatology alone.

The clinical evidence on the benefits and safety of using Moleanalyzer pro in teledermatology services is uncertain. It is also uncertain if its use could lead to fewer face-to-face referrals and biopsies compared with using teledermatology alone. Studies on Moleanalyzer pro did not include non-melanoma skin cancers and were not based in the NHS, so it was unclear how a melanoma-only tool would be used in NHS practice. So, the potential benefits from adding it to an established teledermatology service are uncertain.

The clinical evidence suggests that automated DERM is likely to be diagnostically accurate in people with black or brown skin. But, the evidence for both AI technologies is mostly for skin lesions in people with white skin. The amount of data in people with black or brown skin remains small, at around 3% of all data collected,

so more data is needed to be certain that these technologies do not incorrectly detect or miss skin cancer in people with black or brown skin.

More research is needed on DERM and Moleanalyzer pro to better understand their clinical and cost effectiveness and whether their use could benefit teledermatology services and lead to more timely care for people with all skin conditions.

2 The technologies

Deep Ensemble for Recognition of Malignancy (DERM; Skin Analytics)

2.1 DERM is an artificial intelligence (AI)-based skin lesion analysis software technology intended for use in the screening, triage and assessment of suspected skin cancer lesions, in people aged 18 or over. It is intended to be used as an automated tool or with a second read, to decide if further assessment by a dermatologist is needed. A smartphone is used to capture images of skin lesions using a dermoscopic lens attachment, and the images are uploaded to the online platform. The DERM platform uses an AI-based fixed algorithm (it does not update itself automatically) to analyse the dermoscopic images and provide a suspected diagnosis of the lesion. If DERM labels the lesion as benign, the person is discharged from the urgent suspected skin cancer pathway and is told the results with safety netting advice. If DERM labels the lesion as pre-cancer or cancer, an NHS dermatologist reviews the case virtually and decides on a management plan for the person. DERM can classify lesions as: melanoma, squamous cell carcinoma, basal cell carcinoma, intra-epidermal carcinoma, actinic keratosis, atypical nevus or benign lesions (this includes benign vascular lesion, seborrheic keratosis, dermatofibroma, solar lentigo and melanocytic benign nevus). If a lesion has features of more than 1 lesion type, DERM uses a risk hierarchy to diagnose the lesion as the more severe suspected lesion type. DERM is used within teledermatology services after referral from primary care. The cost of using the online platform for a DERM assessment is £30 per

referral. There is an extra cost of £8.20 per referral if NHS teledermatology staff virtually review a case to decide on the most appropriate outcome. The total price can be discounted to £35.90 if the subsequent biopsy results from the lesions that have been assessed by the technology are shared with the company. It costs an extra £17 to have a case reviewed by the company's second-read dermatologist. The company state that these costs include training and data storage. DERM is UKCA marked (class IIa) for AI used as a medical device.

Moleanalyzer pro (FotoFinder Systems)

2.2 Moleanalyzer pro is a software, intended to be used by a healthcare professional for assessing single skin lesions to help to recognise melanoma lesions. The technology is not intended to be used to confirm a clinical diagnosis of melanoma and can be used for any age group. The target population is people with skin lesions, moles or multiple nevus syndrome. Moleanalyzer pro is used with the FotoFinder Universe software platform which includes the FotoFinder AI scoring assistant. FotoFinder provides 2 options: online AI in which the algorithm is updated continuously and offline AI in which the algorithm can be updated annually. The software needs a dermoscopic image of the lesion for the AI score analysis. The AI score is based on comparisons with images of cancer skin tumours, such as melanoma, basal cell carcinoma, lentigo maligna, squamous cell carcinoma, actinic keratosis and many others. The score indicates how similar a lesion is to these comparison images, so it only provides a statistical estimate of the similarity of the person's lesion to the cancer lesion images. A score of 0 to 0.2 indicates the lesion is inconspicuous, 0.21 to 0.49 indicates further clarification is needed, and 0.50 to 1.0 indicates a conspicuous lesion which should be observed with great attention. The cost of the FotoFinder AI scoring assistant is offered at a flat fee of £1,210 plus VAT per year for single-user access. The Moleanalyzer pro including AI scoring assistant's offline package is £1,750 plus VAT per year for single-user access. There is no cost for training and there is a discount available for multi-user access.

Moleanalyzer pro is CE-marked (class IIa) for AI used as a medical device.

Care pathway

2.3 Skin cancer is an abnormal growth of skin cells and most often develops on skin that has been exposed to the sun. There are 3 major types of skin cancer: melanoma, squamous cell carcinoma (SCC), and basal cell carcinoma (BCC). There are also other rare skin cancers. The first assessment of a skin condition is done by a GP in primary care, to determine the appropriate referral pathway. Section 1.7 on skin cancers in NICE's guideline on suspected cancer describes the criteria for urgent referral of suspected skin cancer. People whose lesions are referred on this pathway should be given a diagnosis or ruling out of cancer within 28 days of being referred by their GP.

2.4 Historically, GPs directly referred everyone with suspicious skin lesions for a face-to-face appointment with a consultant dermatologist, using the urgent suspected skin cancer pathway. Face-to-face appointments are still used for people with multiple suspicious lesions, a history of skin cancer or other risk factors, or when other clinical pathways are unavailable in the local area or are unsuitable.

2.5 NHS dermatology services need to reduce backlogs and delays in providing face-to-face appointments because of limited staff, including dermatologists. The demand from an increasing number of urgent referrals for suspected skin cancers can mean that face-to-face appointments for people with other non-cancer skin conditions (including painful and debilitating inflammatory skin conditions) are delayed. The NHS Plan introduced teledermatology services for triaging, diagnosing, and managing skin conditions without the person being physically present. These services are intended to help manage demand and reduce the number of face-to-face appointments offered to people with

low-risk non-cancer lesions. This evaluation assesses the benefits of using AI technologies within teledermatology services.

The comparator

2.6 The 2 comparators are teledermatology alone and face-to-face dermatology assessment.

3 Committee discussion

NICE's diagnostics advisory committee considered evidence on Deep Ensemble for Recognition of Malignancy (DERM) and Moleanalyzer pro to assess and triage skin lesions within the urgent suspected skin cancer pathway from several sources, including an early value assessment (EVA) report by the external assessment group (EAG), and an overview of that report. Full details are in the project documents for this guidance on the NICE website.

Unmet need

3.1 In the UK, dermatology services receive 1.2 million referrals each year from primary care. About 60% are urgent referrals for suspected skin cancer. Of these, only 6% are confirmed to be skin cancer and the remaining 94% are either non-urgent or non-cancer cases. The high number of urgent referrals combined with staff shortages have resulted in delays in diagnosis and care for people with non-cancer, non-urgent inflammatory skin conditions that need face-to-face assessment. The committee heard about the effect this can have on the quality of life and health outcomes of people with non-cancer dermatological conditions, such as psoriasis. Depending on the local services, urgent suspected skin cancer lesions are seen either in a face-to-face dermatology appointment or through teledermatology. NHS England's (NHSE) teledermatology roadmap supports local NHS systems to accelerate the roll out of teledermatology to help manage demand and reduce face-to-face appointments. Artificial intelligence (AI) technologies used within a

teledermatology service could increase staff capacity to help address the unmet need.

Patient considerations

3.2 The EAG noted that people who were offered an assessment using DERM, were generally supportive of AI technologies being used in some form as part of their assessment (such as a decision support tool). But many would prefer to also have a face-to-face dermatology appointment. The lay members of the committee expressed their preference for a face-to-face assessment of suspicious lesions because they perceived it to be a more comprehensive assessment. They expressed concern about the early use of AI technologies, particularly if they are to be used without a second read (see section 3.14). They were particularly concerned about the potential for misdiagnosis because skin cancer can be life-threatening, meaning there are high risks associated with missed or delayed diagnoses. They were concerned that people who had a skin lesion identified as non-cancer by an AI technology alone may not trust the decision and may re-present in primary care. People may also be concerned and unsure about monitoring their suspicious lesions if they are discharged with safety netting advice, especially if they are older or have multiple lesions.

Healthcare professional considerations

3.3 There was limited data on healthcare professional's opinions of AI technologies. One published study of a staff survey with 6 respondents reported that healthcare professionals expressed mixed opinions about their confidence in automated use of AI technologies to reliably distinguish between non-cancer and cancer lesions.

Automated DERM diagnostic accuracy

3.4 Company data from NHS services which are already using DERM (collected from April 2020 to November 2023) shows automated DERM has a 97% sensitivity for detecting cancer lesions and a 95% sensitivity

for detecting melanoma. This data included 72,390 people (with 85,955 lesions), but only 27,747 of these lesions were assessed in secondary care using a recent version of DERM and had final outcomes that could be used to calculate sensitivity. The committee highlighted that this data suggests that 1 in 20 melanomas could be misdiagnosed using automated DERM, and the person discharged incorrectly. The sensitivity of automated DERM to detect malignant lesions ranged from 91.0% to 100% across 3 published studies (DERM-003 Marsden et al. 2023; DERM-005 Marsden et. al 2024; and Thomas et al. 2023). The committee had some concerns around the risk of bias for the reference standard in DERM-003 because 1 dermatologist provided the clinical diagnosis used as the ground truth for non-biopsied lesions. The committee acknowledged that using DERM with a second read could reduce the risk of missing skin cancers, but it is uncertain if this approach would increase staff capacity (see section 3.14).

3.5 It is unclear whether automated DERM is as sensitive in detecting malignant lesions as current teledermatology alone. A recent study (Marsden et al. 2024) reported sensitivities for detecting cancer lesions of 94.0% (95% confidence intervals [CI]: 84.7 to 98.1) for automated DERM and 97.0% (95% CI: 88.7 to 99.5) for teledermatology, and noted that the confidence intervals overlapped. The EAG did not systematically review the evidence on the diagnostic accuracy of teledermatology alone but noted that the sensitivity of teledermatology to detect cancer lesions is uncertain. Clinical experts noted that teledermatology has become more widespread since the pandemic and greater use may impact on the accuracy seen in practice. The committee noted that if the sensitivity of teledermatology is high, then the potential benefit of improved diagnostic outcomes from adding automated AI technologies may be limited. The committee concluded that more research is needed on the sensitivity of automated AI technologies to detect malignant lesions used within a well-established teledermatology service compared with the sensitivity of a well-established teledermatology service alone.

Moleanalyzer pro diagnostic accuracy

3.6 The committee noted that more research is needed on the diagnostic accuracy of Moleanalyzer pro in non-melanoma skin cancers. The evidence suggests that Moleanalyzer pro has lower sensitivity but higher specificity for detecting melanoma than face-to-face assessment with a dermatologist. There are no prospective studies that report the diagnostic accuracy of Moleanalyzer pro to detect non-melanoma skin cancers. The committee noted that Moleanalyzer pro studies were not explicitly based within teledermatology services nor based within the NHS, so it was unclear how a melanoma-only tool would be used in NHS practice. There was also a lack of evidence on the proportion of people the technology is unsuitable for, how Moleanalyzer pro would affect the number of referrals and biopsies, and the cost effectiveness of using Moleanalyzer pro.

Diagnostic accuracy in people with black or brown skin

3.7 The committee was concerned about the diagnostic accuracy of using automated AI technologies to detect skin cancer in people with black or brown skin. There is limited data to validate AI technologies for people with black or brown skin because there is a low incidence of skin cancers among people from Black, Black Caribbean, Black African and Asian ethnic groups. The committee noted that high risk cancers (squamous cell carcinomas [SCCs] and melanoma) are 20 to 30 times more likely to occur in people from White ethnic groups. But people from Black, Black Caribbean, Black African and Asian ethnic groups are more likely to have a worse prognosis because lesions may be detected later. They are also more likely to have acral lesions (lesions on palms of hands and soles of feet) which have a higher risk of cancer. AI assessment is not suitable for assessing acral lesions and these are referred directly for dermatologist assessment. Even when skin cancer is diagnosed at the same stage, people from Black, Black Caribbean, Black African and Asian ethnic groups have a greater risk of mortality than people from White ethnic groups. Automated DERM has primarily been evaluated in people with

white skin (Fitzpatrick skin types 1 to 3). Similarly, most people in the Moleanalyzer pro studies had white skin (Fitzpatrick types 2 to 3). Most studies did not report the proportion of participants with different Fitzpatrick skin types, but DERM-003 reported that 0% of participants had black skin and DERM-005 reported that 1% of participants had black skin. The EAG noted that recent company data on using automated DERM in people with brown or black skin (Fitzpatrick skin types 5 and 6) showed that no cancer lesions were missed, which suggests that automated DERM is as diagnostically accurate in people with black or brown skin as it is in people with white skin. But only 3% of lesions assessed by DERM with confirmed diagnoses were in Fitzpatrick skin types 5 and 6. The committee emphasised that because the amount of data remains small, more research should be done on the performance of automated DERM in people with black or brown skin to ensure AI technologies are not incorrectly detecting (false positive) or missing (false negative) skin cancer. The clinical experts also advised that studies should measure skin tone with spectrophotometry rather than using the Fitzpatrick scale because spectrophotometry is a more accurate way of measuring total melanin content in skin.

Eligibility for assessment with AI technologies

3.8 The committee noted that a large proportion of skin lesions are not eligible for assessment by AI technologies and would need face-to-face appointments, for example, those obscured by hair, tattoos or scars. The EAG reported that the proportion of participants that were excluded from studies because of ineligible lesions ranged between 15.6% and 27.4%, where reported. The clinical experts noted that similar exclusion criteria also apply with teledermatology assessment, but the company's economic model assumed that fewer people were eligible for assessment by automated DERM than teledermatology (81% compared with 90%). This would have an impact on the cost of the service with AI technologies. The committee concluded that more research is needed to

understand the proportion of skin lesions that are eligible for assessment by automated AI technologies and teledermatology alone.

Impact on referral rates

3.9 An analysis by the EAG suggested that, of eligible lesions, automated use of DERM could approximately halve the number of referrals to a dermatologist within the urgent skin cancer pathway. The EAG's analysis also suggested that automated use of DERM could result in more lesions being correctly identified as non-cancer without a biopsy. So fewer biopsies would be needed, and people would be correctly discharged from the service. The committee noted that a well-established teledermatology service could also reduce the number of referrals to face-to-face dermatologist appointments. It is uncertain whether DERM used with a second read would reduce the number of referrals and biopsies compared with a well-established teledermatology service.

Potential cost effectiveness of automated DERM

3.10 Early modelling done by the company suggested that automated DERM used for assessing suspicious skin lesions within a well-established teledermatology service has the potential to be cost effective compared with face-to-face assessment. It is less certain if automated DERM used within a teledermatology service would be cost effective compared with a well-established teledermatology service alone. The EAG noted that in the company's economic model the specificity of teledermatology is a key driver in determining cost effectiveness. A low specificity to detect cancer lesions would result in a high number of lesions referred for further assessment and would increase costs. Specificity of teledermatology to detect cancer lesions is uncertain, with estimates ranging from 35% (taken from real-world data from DERM pilot studies) to 84.3% (taken from a Cochrane review). The model assumes that automated DERM has a specificity of 42% based on real world performance data. The clinical experts noted that the Cochrane review was published before the COVID 19 pandemic and that it is not generalisable to the current UK skin cancer

pathway. So, the cost effectiveness of automated DERM within a teledermatology service compared with teledermatology alone is uncertain. The committee concluded that more research is needed on the specificity of a well-established teledermatology service, to help ascertain the cost effectiveness of using automated DERM within teledermatology services.

Infrastructure costs

3.11 AI technologies can only be used after primary care referral, in local areas where a teledermatology service is available. This is because a dedicated service for taking high quality medical photographs of the suspicious lesion by a trained medical photographer is essential for an accurate AI assessment. There are costs associated with setting up this infrastructure and for training medical photographers. The committee noted that although there are teledermatology services in many areas, there is variation across the UK and many areas still refer all suspected skin cancer lesions for an urgent face-to-face appointment. With the wider roll out of teledermatology services, these costs will likely be incurred regardless of whether AI technologies are used or not.

Conceptual model

3.12 The committee thought that the conceptual model proposed by the EAG was appropriate. It captured the costs and long-term health consequences associated with the misdiagnosis of BCCs. The committee suggested comparing costs of using AI technologies (see section 2.1 and section 2.2) with the costs incurred by the NHS for outpatient referrals, and that these should be included in the EAG's cost-effectiveness modelling. It also noted that it would be important to consider how increases in staff capacity could be captured in the model, to meaningfully quantify the impact of reducing demand on dermatology services.

Equality considerations

3.13 The technologies may not be suitable for everyone. Skin cancer is known to be more difficult to accurately detect in people with black or brown skin, which has led to poorer outcomes associated with later diagnosis. There is less data in people with black or brown skin because of their lower incidence of skin cancer. So, the committee recommended more research on the performance of automated DERM in people with black or brown skin to ensure AI technologies are not incorrectly detecting (false positive) or missing skin cancer (false negative, see section 3.8). AI technologies may not be suitable for people with more than 3 lesions and older people. This is because in people who are older or who already have several skin lesions, a whole body skin examination by a dermatologist is more likely to find more skin lesions than those originally presented with (see section 3.9).

4 Committee members and NICE project team

Committee members

This topic was considered by NICE's diagnostics advisory committee, which is a standing advisory committee of NICE.

Committee members are asked to declare any interests in the technologies to be evaluated. If it is considered there is a conflict of interest, the member is excluded from participating further in that evaluation.

The minutes of the diagnostics advisory committee meetings, which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Additional specialist committee members took part in the discussions and provided expert advice for this topic:

Specialist committee members

Eugene Healy

Department chair & head of dermatology, University of Southampton, and honorary consultant dermatologist, University Hospital Southampton NHS Foundation Trust

Kate Hawley

Lay member

Nalayini Kumaralingam

Nurse consultant, Kent Oncology Centre, Maidstone Hospital

Roger Aldridge

Consultant plastic surgeon & consultant dermatological surgeon, NHS Lanarkshire & NHS Lothian

Rubeta Matin

Consultant dermatologist, Oxford University Hospitals NHS Foundation Trust

Stephanie Gallard

Dermatology GP with special interest (GPSI), Liverpool Intermediate Community Assessment and Treatment Service (ICATS) & Liverpool University Foundation Trust

Stephen McKenna

Professor, research cluster lead (computer sciences & informatics), University of Dundee

Susan Mountain

Lay member

NICE project team

Each early value assessment topic is assigned to a team consisting of 1 or more health technology assessment analysts (who act as technical leads for the topic), a health technology assessment adviser and a project manager.

Anam Khan

Health technology assessment analyst

Frances Nixon

Health technology assessment adviser

Toni Gasse

Project manager

ISBN: