NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 30 • Issue 10 • February 2026

ISSN 2046-4924

Artificial Intelligence technologies for assessing skin lesions for referral on the urgent suspected cancer pathway to detect benign lesions and reduce secondary care specialist appointments: early value assessment

Matthew Walton, Alexis Llewellyn, Eleonora Uphoff, Joseph Lord, Melissa Harden, Robert Hodgson and Mark Simmonds

DOI 10.3310/GJMS0317

NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Artificial Intelligence technologies for assessing skin lesions for referral on the urgent suspected cancer pathway to detect benign lesions and reduce secondary care specialist appointments: early value assessment

Matthew Walton[1], Alexis Llewellyn[1], Eleonora Uphoff[1], Joseph Lord[1], Melissa Harden[1], Robert Hodgson[1] and Mark Simmonds[1*]

[1]Centre for Reviews and Dissemination, University of York, Heslington, York, UK

[*]**Corresponding author** mark.simmonds@york.ac.uk

Published February 2026
DOI: 10.3310/GJMS0317

This report should be referenced as follows:

Walton M, Llewellyn A, Uphoff E, Lord J, Harden M, Hodgson R, Simmonds M. Artificial Intelligence technologies for assessing skin lesions for referral on the urgent suspected cancer pathway to detect benign lesions and reduce secondary care specialist appointments: early value assessment. *Health Technol Assess* 2026;**30**(10). https://doi.org/10.3310/GJMS0317

Health Technology Assessment

NICE TAR and DAR

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal
Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was commissioned and funded by the Evidence Synthesis Programme on behalf of NICE as award number NIHR136014. The protocol was agreed in October 2023. The draft manuscript began editorial review in March 2024 and was accepted for publication in April 2025. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

Background: Skin cancers are some of the most common types of cancer. Dermatology services receive about 1.2 million referrals a year, but only a small minority are confirmed skin cancer. Artificial intelligence may be helpful in the diagnosis of skin cancer by identifying lesions that are or are not cancerous.

Objectives: To investigate the clinical and cost-effectiveness of two artificial intelligence technologies: DERM (Deep Ensemble for Recognition of Malignancy, Skin Analytics) and Moleanalyzer Pro (FotoFinder), as decision aids following a primary care referral.

Methods: A rapid systematic review of evidence on the two technologies was conducted. A narrative synthesis was performed, with a meta-analysis of diagnostic accuracy data.

Published and unpublished cost-effectiveness evidence on the named technologies, as well as other diagnostic technologies were reviewed. A conceptual model was developed that could form the basis of a full economic evaluation.

Results: Four studies of DERM and two of Moleanalyzer Pro were subject to full synthesis. DERM had a sensitivity of 96.1% to detect any malignant lesion (95% confidence interval 95.4 to 96.8); at a specificity of 65.4% (95% confidence interval 64.7 to 66.1). For detecting benign lesions, the sensitivity was 71.5% (95% confidence interval 70.7 to 72.3) for a specificity of 86.2% (95% confidence interval 85.4 to 87.0). Moleanalyzer Pro had lower sensitivity, but higher specificity for detecting melanoma than face-to-face dermatologists.

DERM might lead to around half of all patients being discharged without assessment by a dermatologist, but a small number of malignant lesions would be missed. Patient and clinical opinions showed substantial resistance to using artificial intelligence without any assessment of lesions by a dermatologist.

No published assessments of the cost-effectiveness of the technologies were identified; three assessments related to skin cancer more broadly in a National Health Service setting were identified. These studies employed similar model structures, but the mechanism by which diagnostic accuracy influenced costs and health outcomes differed. An unpublished cost–utility model was provided by Skin Analytics. Several issues with the modelling approach were identified, particularly the mechanisms by which value is driven and how diagnostic accuracy evidence was used.

The conceptual model presents an alternative approach, which aligns more closely with the National Institute for Health and Care Excellence reference case and which more appropriately characterises the long-term consequences of basal cell carcinoma.

Limitations: The rapid review approach meant that some relevant material may have been missed, and capacity for synthesis was limited. The proposed conceptual model does not capture non-cash benefits associated with demand on dermatologist time. An assessment of the likely budget impact and resource use could not be provided.

Conclusions: DERM shows promising diagnostic accuracy for triage and diagnosis of suspicious cancer lesions in selected patients referred from primary care. Its impact on the diagnostic pathway and patient care is, however, uncertain. Moleanalyzer Pro shows promising accuracy for diagnosing melanoma, but its evidence base is limited.

Future work: While artificial intelligence has the potential to be cost-effective for the identification of benign lesions, further research addressing the limitations in the diagnostic accuracy evidence is necessary. Without comparative evidence on the diagnostic accuracy of artificial intelligence technologies, their value will remain uncertain.

Study registration: This study is registered as PROSPERO CRD42023475705.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR136014) and is published in full in *Health Technology Assessment*; Vol. 30, No. 10. See the NIHR Funding and Awards website for further award information.

Contents

List of tables

List of figures

List of abbreviations

2WW	2-week wait
ACM	Association for Computing Machinery
AI	artificial intelligence
AJCC	American Joint Committee on Cancer
AK	actinic keratosis
BCC	basal cell carcinoma
CDC	community diagnostic centre
CENTRAL	Cochrane Central Register of Controlled Trials
CP	clinical photographer
DERM	Deep Ensemble for Recognition of Malignancy
EAG	Evidence Assessment Group
EQ-5D-3L	EuroQol-5 Dimensions, three-level version
EQ-5D-5L	EuroQol-5 Dimensions, five-level version
FN	false negative
FP	false positive
GP	general practitioner
HCA	healthcare assistant
HRQoL	health-related quality of life
ICER	incremental cost-effectiveness ratio
IEC	intra-epidermal carcinoma
LM	lentigo maligna
MDT	multidisciplinary team meetings
NICE	National Institute for Health and Care Excellence
NPV	negative predictive value
PPV	positive predictive value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
prosp.	prospective
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life-year
QUADAS-2	quality assessment of diagnostic accuracy studies-2
QUADAS-C	quality assessment of diagnostic accuracy studies-comparative
RCT	randomised controlled trial
ROC	receiver operating characteristic
SCC	squamous cell carcinoma
SLNB	sentinel lymph node biopsy
TN	true negative
TP	true positive
UHBFT	University Hospital Birmingham Foundation Trust
UHL	University Hospital Leicestershire
WSFT	West Suffolk Foundation Trust

Note

This monograph is based on the Diagnostic Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed confidential. The full report was used by the Diagnostic Advisory Committee at NICE in their deliberations. The full report with each piece of confidential data removed and replaced by the statement 'confidential information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Plain language summary

Skin cancers and suspicious skin lesions are very common. People with moles or lesions that might be cancerous are referred to a skin cancer specialist (a dermatologist) to make a diagnosis. This places a very high burden on dermatology clinics and, as a result, there can be delays in seeing a dermatologist and getting a diagnosis. Artificial intelligence systems could potentially use a high-quality photograph to identify which lesions do not need to be seen by a specialist. This could be done by the artificial intelligence system alone, or in combination with remote review by a dermatologist.

This project investigated whether two artificial intelligence technologies: DERM (Skin Analytics) and Moleanalyzer Pro (FotoFinder) could be useful in reducing the burden on dermatology services while helping to identify skin cancer. The evidence was reviewed to investigate whether the technologies can accurately identify skin cancer cases, and whether their use might improve the diagnosis process for patients. We also designed a theoretical model in which the economic value of artificial intelligence technologies for the diagnosis of skin cancer could be assessed. As part of this process, we sought to outline what further evidence would be needed to implement a full assessment.

The evidence we reviewed suggests DERM could potentially reduce by half the number of patients that would be referred to specialist dermatologists, while still identifying 95% of all skin cancers. Moleanalyzer Pro could identify about 85% of malignant melanomas. This appears to be a similar accuracy to that achieved by using a remote view of the lesions by dermatologists alone. How DERM or Moleanalyzer Pro use would impact diagnosis and treatment for patients in practice, and the burden on clinicians, is currently unclear.

Because of limitations in the evidence on the diagnostic accuracy of artificial intelligence technologies, a full assessment of their economic value is not possible at this time. Further research should focus on better establishing the diagnostic accuracy of both artificial intelligence technologies and current service provision.

Scientific summary

Background

Skin cancers are some of the most common types of cancer. Over 16,000 cases of melanoma, and more than 210,000 cases of non-melanoma skin cancer are diagnosed every year in the UK. In current practice, patients with suspicious skin lesions are referred to secondary care through the urgent suspected skin cancer referral pathway, where people attend a secondary care dermatology department for a face-to-face appointment with a consultant dermatologist. As benign skin lesions and skin cancer are so common, this places a very high burden on dermatology clinics, which may lead to a reduction in capacity to handle other skin conditions.

Artificial intelligence (AI) may be helpful in the diagnosis of skin cancer. An AI system could potentially identify which referred lesions are not cancerous using a high-quality photograph. An AI system could be used alone, or in combination with a dermatologist looking at the photograph. People judged not to have cancer could then be quickly discharged prior to secondary care consultation, while people whose lesion may be cancerous may be seen by a specialist in person. AI systems could therefore potentially speed up the diagnostic process and reduce the burden on the health service. AI systems are already used in the NHS in a research context, but there is a need to evaluate their clinical impact and value.

This project investigated whether two such AI technologies – Deep Ensemble for Recognition of Malignancy (DERM; Skin Analytics) and Moleanalyzer Pro (FotoFinder) –can produce clinically meaningful benefits for skin cancer diagnosis, and whether they have the potential to be cost-effective.

Objectives

The aim of the project was to investigate the clinical and cost-effectiveness of the two AI technologies, DERM and Moleanalyzer Pro, as decision aids to triage and diagnose suspicious skin lesions following a referral on the urgent suspected skin cancer pathway. To achieve this, the following objectives were proposed:

- To perform a rapid systematic review, narrative synthesis, and, where feasible, a meta-analysis, of the diagnostic accuracy, clinical impact and practical implementation of the included AI technologies.
- To perform a rapid systematic review of published cost-effectiveness studies of diagnostic strategies used to aid the diagnosis of skin cancer.
- To develop a conceptual model that will identify likely drivers of health benefit, harms and costs associated with implementing the included AI technologies in the NHS and identify areas for further research.

Methods

Data sources

MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials and the Association for Computing Machinery Digital Library were searched in November 2023. Clinical trial registries were searched. Unpublished material supplied by the included companies was also assessed.

Inclusion criteria

Any clinical study evaluating DERM or Moleanalyzer Pro in people with skin lesions suspicious of cancer, presenting in primary care, rapid diagnostic clinic, teledermatology or secondary care settings were eligible for inclusion. Included studies must report diagnostic accuracy, clinical outcomes, or evidence on implementation. The comparator was clinical judgement by dermatologists, but this did not need to be reported for a study to be eligible. The preferred reference standard for diagnosis was histology, but for unbiopsied lesions, clinical confirmation of non-malignancy was accepted.

The cost-effectiveness review included any economic evaluation including budget impact models, return on investment analysis, and other cost-only analyses of either DERM or Moleanalyzer Pro in the above population and setting. It was anticipated that no relevant studies would be identified for the named technologies; therefore, additional searches were also conducted to identify cost-effectiveness studies looking at any technology used to aid diagnosis of skin cancer in an NHS setting.

Data extraction

An initial scoping of studies was performed by extracting data on intervention, study location, size, setting, type of outcomes reported, and design and key quality indicators. Only studies with prospective recruitment of patients were taken forward for full data extraction and synthesis. For those studies, full data on the intervention, patient characteristics and all reported outcomes were extracted. Risk of bias was assessed using quality assessment of diagnostic accuracy studies-2 and quality assessment of diagnostic accuracy studies-comparative.

Identified economic evaluations were reviewed and discussed in detail, with the aim of informing the design and parameterisation of conceptual model. Material provided by submitting stakeholders pertaining to the value case for their product was also reviewed.

Synthesis

A scoping process was used to classify identified studies for relevance to the decision problem, based on study quality, setting, outcomes reported and relevance to the NHS. For studies taken forward from the scoping phase for full synthesis, a narrative synthesis was performed. Results are presented in structured tables and figures as appropriate, with a text summary. Random-effects meta-analyses of sensitivity and specificity were performed to pool diagnostic accuracy estimates across studies.

Evidence related to cost-effectiveness studies was reviewed and synthesised narratively.

Modelling

The conceptual model described sought to provide an overview of the structure of a cost–utility model and key evidence required for the assessment of AI technologies for the identification of benign lesions among suspected cancer cases referred on the urgent referral pathway. The structure of the conceptual model was designed considering the strengths and limitations of previously published diagnostic models for skin cancer in an NHS setting, and evidence submitted by stakeholders.

Results

Diagnostic accuracy and clinical impact of DERM

Six studies of DERM were identified, of which four were considered for full synthesis. Those four studies were all conducted in the UK. All studies excluded a substantial proportion of participants from assessment, which may produce biased results.

Meta-analysis of diagnostic accuracy data supplied by the company suggested that DERM has a high sensitivity of 96.1% to detect any malignant lesion [95% confidence interval (CI) 95.4 to 96.8], at a specificity of 65.4% (95% CI 64.7 to 66.1). The diagnostic accuracy for detecting melanoma or squamous cell carcinoma specifically was similar. For the detection of benign lesions, the sensitivity was 71.5% (95% CI 70.7 to 72.3) for a specificity of 86.2% (95% CI 85.4 to 87.0). This appears to be comparable in diagnostic accuracy to that achieved by dermatologists without the use of DERM. The diagnostic accuracy of combining DERM with assessment by a dermatologist could not be assessed.

Data on the clinical impact of using DERM were limited, and mostly unpublished. Some trial data suggested that autonomous use of DERM would lead to approximately half of patients being referred to a dermatologist for further assessment, and half being discharged. However, around 1% of people would be discharged with malignant lesions [mostly basal cell carcinomas (BCCs)]. DERM could potentially be used as part of a teledermatology service. However, use of DERM may slow progress to diagnosis.

Patient and clinical opinions of DERM were generally favourable towards accepting its use as part of the diagnostic pathway. However, there was very substantial resistance, particularly among clinicians, to using DERM without any assessment of lesions by a dermatologist.

Diagnostic accuracy and clinical impact of Moleanalyzer Pro

Seventeen publications of Moleanalyzer Pro were identified, but these were mostly retrospective reviews, and two prospective studies were eligible for full data extraction. The applicability of the evidence for Moleanalyzer Pro to practice is limited, notably due to the lack of studies from the UK and the lack of data for non-melanocytic lesions.

When pooled, these studies found that Moleanalyzer Pro alone had a sensitivity of 84.4% (95% CI 73.9 to 91.0) and a specificity of 84.5% (95% CI 72.0 to 92.1) to detect melanoma. Moleanalyzer Pro had a lower sensitivity and higher specificity to detect melanoma when compared with face-to-face dermatologist and remote teledermatology. There was no evidence on the diagnostic accuracy of Moleanalyzer Pro to detect other skin cancers, and no evidence was found on its clinical impact.

Economic evidence

No economic studies relating to the named technologies were identified from searches of the literature. Broader searches for any technology used to aid diagnosis of skin cancer in an NHS setting identified three studies. Although relevant to this review, none related to the use of AI for the detection of skin cancer and considered populations which were not relevant to the decision problem. While all identified studies adopted similar model structures, the mechanisms by which diagnostic accuracy generated value (in terms of either cost savings or quality-adjusted life-year gain) differed across these models. For instance, diagnostic sensitivity had less value in some models with value instead generated by the avoidance of unnecessary referral and diagnostic procedures.

Economic evidence on the cost-effectiveness of DERM was submitted by Skin Analytics and NHS England. This evidence was preliminary and did not include an executable model. The most relevant analysis was a cost–utility model developed by the Exeter Test Group and Skin Analytics. The Evidence Assessment Group (EAG) considered the model structure largely appropriate to capture important direct cost and health implications of AI technologies for directing discharge in a post-referral setting. However, a lack of key comparative data meant the relative clinical and cost-effectiveness of alternative pathways was necessarily based on often optimistic assumptions. The model suggested DERM could be highly cost-effective in the NHS, but we note that results may be very sensitive to the use of alternative sources of diagnostic accuracy data. We also noted several issues which may mean that the main value drivers were not appropriately characterised. Namely, the model imposed disincentives for the correct diagnosis and treatment of BCC; structurally imposed assumed sensitivity benefits for any strategy incorporating a triage step; used costs associated with biopsy and treatment which were inconsistent with sources generally used in National Institute for Health and Care Excellence appraisals, and may overvalue specificity in terms of generating cost savings.

No economic evidence related to Moleanalyzer Pro was identified.

Conceptual model

We developed a conceptual model aimed at providing an alternative to that presented in the Skin Analytics submission. While the proposed model retained the structure reported by Skin Analytics, the EAG propose an alternative structure for patients with BCC, aimed at better capturing the cost and health consequences of BCC, particularly with reference to disease recurrence.

We consider the current evidence inadequate to fully address the decision problem. Current evidence for both DERM and Moleanalyzer Pro is lacking with regard to the diagnostic accuracy of the whole diagnostic pathway (i.e. inclusive of subsequent steps). Availability of these data is essential to understanding the likelihood of missed cases which cannot be inferred from the partial data currently available. Similarly, comparable diagnostic accuracy data describing current service provision is lacking, particularly for the teledermatology pathway.

Conclusions

Impact on practice

The diagnostic accuracy of DERM suggests that it has potential for use within a post-primary care referral setting. This could be either alongside assessment by dermatologists or as an autonomous tool within the post-referral pathway within a subset of patients. However, the practical impact and clinical benefit of using DERM in a post-referral setting is currently unclear. In particular, the impact on referrals and secondary care appointments, the burden on clinicians and the subsequent clinical impact on patients are largely unclear. Although Moleanalyzer Pro shows promising accuracy for diagnosing melanoma, its evidence base is currently too limited to fully assess its clinical value.

Evidence on the diagnostic accuracy and clinical value of AI in people with darker skin tones or with lesions that are more difficult to assess (such as when versions are large, or obscured by scarring, tattooing or hair) was largely absent. Only a small number of people with darker skin tones were recruited to the included studies, and people with hard-to-assess lesions were often excluded. This raises concerns as to whether AI could be used in these people.

Current economic evidence supporting the cost-effectiveness of DERM is limited, and it is unclear whether the plausible advantages of DERM represent value for money relative to other strategies. Company-sponsored analyses suggested that DERM used autonomously and with a second read could be highly cost-effective compared to current 2-week wait diagnostic models. However, much of this value is generated through potentially optimistic assumptions around the diagnostic accuracy of comparators, and of the surrounding pathway (confidential information has been removed). Notably, the magnitude of uncaptured non-cash-releasing benefits remains unquantified.

There is currently no economic evidence supporting the use of Moleanalyzer Pro, but assuming a similar use case to DERM and appropriate data collection, the value of Moleanalyzer Pro could be assessed using the conceptual framework presented by the EAG.

Future research needs

The diagnostic accuracy of AI in a post-primary care referral pathway is uncertain and requires further evaluation. A lack of key comparative data on diagnostic accuracy means the relative clinical and cost-effectiveness of pathways incorporating AI technologies and teledermatology remains highly uncertain. Assessments of diagnostic accuracy of AI in people with darker skin tones or with hard-to-assess lesions are urgently needed.

Directly comparable evidence on the diagnostic accuracy of AI technologies and teledermatology in a post-referral setting compared with unassisted teledermatology is required to assess the potential value of AI technologies. This would require studies comparing AI with dermatologists' assessments, recruiting a representative population and case-mix, use of up-to-date versions of AI and dermoscopy, and with a robust independent reference standard for all patients.

A better understanding of the clinical benefits and resource implications associated with the implementation of AI technologies will also require further research to set up AI and teledermatology services in the NHS. Further research must also be undertaken to quantify the benefits to population health within skin cancer and other dermatological indications associated with any release of NHS consultant dermatologist resource, and understand the effects of these technologies on waiting times for final diagnosis.

This could potentially be achieved through continuations and extensions of existing ongoing pilot studies of DERM, but truly comparative evidence may also be required. Moleanalyzer Pro requires evaluation within a UK teledermatology setting.

The substantial resistance from both patients and clinicians to using AI without any human dermatological assessment means that if AI is to be used to direct discharge autonomously, more evidence is needed to demonstrate that it has clear benefits to patients, without sacrificing accuracy.

Study registration

This study is registered as PROSPERO CRD42023475705.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR136014) and is published in full in *Health Technology Assessment*; Vol. 30, No. 10. See the NIHR Funding and Awards website for further award information.

Chapter 1 Background and definition of the decision problem

Purpose of the decision to be made

The purpose of this assessment was to investigate the use of artificial intelligence (AI) technologies for the analysis of skin lesions suspicious of cancer following a referral on the urgent suspected skin cancer pathway. The assessment considered the use of two technologies: Deep Ensemble for Recognition of Malignancy (DERM) (Skin Analytics) and Moleanalyzer Pro (FotoFinder systems). The assessment considered existing evidence and identified potential evidence gaps on whether these technologies have the potential to be clinically useful and cost-effective to the NHS.

Interventions

The Evidence Assessment Group (EAG) evaluated whether two AI technologies, DERM and Moleanalyzer Pro, represent an effective and reliable means of triaging cancer from benign skin lesions, alongside current clinical practice.

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

DERM (Skin Analytics) is a UKCA class IIa AI-based skin lesion analysis technology intended for screening, triage and assessment of suspicious skin lesions. It is indicated for use on dermoscopic images of skin lesions where skin cancer is suspected in patients aged 18 years or over.

DERM uses AI-based algorithms to provide a suspected diagnosis of a given lesion and, where applicable, a referral recommendation (e.g. discharge and give safety netting advice or urgent referral for suspected cancer). DERM can classify lesions as: melanoma, squamous cell carcinoma (SCC), basal cell carcinoma (BCC), intra-epidermal carcinoma (IEC), actinic keratosis (AK), atypical nevus or benign lesions (this includes benign vascular lesion, seborrheic keratosis, dermatofibroma, solar lentigo and melanocytic benign nevus). If a lesion exhibits features of more than one lesion type, DERM uses a risk hierarchy to return the more severe suspected diagnosis. The algorithm was trained on both historical (retrospectively) and prospectively collected images from populations in the UK, USA and Italy. DERM uses a fixed algorithm and does not update itself automatically.

The technology has been deployed in the NHS since April 2020, including as a triage tool following a primary care referral. Over 51,000 patients have been assessed following a general practitioner (GP) referral on the urgent suspected skin cancer pathway, to identify patients with benign lesions who can be discharged from the pathway without requiring specialist input from secondary care. People with suspicious lesions after DERM assessment have then been referred to a teledermatology review by a secondary care specialist.

Moleanalyzer Pro (FotoFinder systems)

Moleanalyzer Pro (FotoFinder systems) is a class IIa CE marked AI-based technology intended to be used by a medical professional for non-invasive visual documentation of skin lesions and aims to help the recognition of melanoma lesions. The technology is not intended to be used to confirm a clinical diagnosis of melanoma, and it can be used for any age group. The target population is people with skin lesions, moles or multiple nevus syndrome. Lesions can be between 2 mm and 20 mm and should be on intact skin without additional psoriasis, eczema, acute sunburn or on hair-covered parts of the body.

Moleanalyzer Pro is used with the FotoFinder Universe software platform. The system requires a dermoscopic image for the AI score analysis. The software can only be used with the FotoFinder dermatoscopes: Dermlite Handyscope (this is compatible with any smartphone or tablet) and with Medicam 1000.

FotoFinder provides two options: online AI where the algorithm is updated continuously and offline AI in which the algorithm can be updated annually. This AI score is based on comparisons with images of malignant skin tumours such as: melanoma, BCC, lentigo maligna (LM), SCC, AK, and many others. The score indicates how similar a lesion is to these comparison images; therefore, it is only meant to provide a statistical estimate of the similarity to the malignant lesion images. A score between 0 and 0.2 indicates the lesion is inconspicuous, 0.21–0.49 indicates further clarification is necessary, and 0.50–1.0 indicates a conspicuous lesion which should be observed with great attention. Moleanalyzer Pro is already in use in some NHS centres.

Populations and relevant subgroups

The population of interest was people with skin lesions suspicious of cancer, who have been referred from primary care for further evaluation. The particular setting of interest was patients undergoing teledermatology assessments, but all settings after primary care referral were considered.

Subgroups relevant to this appraisal were according to skin colour and type, and socioeconomic status.

Place of the intervention in the treatment pathway

In the UK, dermatology services receive about 1.2 million referrals a year and about 60% of these are suspected skin cancer pathway referrals, but only about 6% are converted to a confirmed case of skin cancer.[1] A significant proportion of people referred by GPs may not require face-to-face appointments in dermatology departments. The Getting It Right First Time report on dermatology highlighted that there are shortages in the workforce leading to delays in the diagnosis and treatment of skin cancer.[2] Furthermore, experts in dermatology mentioned there is a low threshold for referral because GPs do not receive in-depth dermatology training and many do not have access to dermatoscopes, which are essential for confidently identifying both benign skin lesions and skin cancer.

Types of skin cancer

This assessment covers all types of skin cancer. This includes three main types of skin cancer: melanoma, SCC and BCC, as well as other, rarer, forms of skin cancer.

Melanoma

A melanoma is a malignant tumour arising from melanocytes in the skin and is usually seen as a pigmented lesion on the skin. Melanoma is the fifth most common cancer in the UK, accounting for around 4% of all new cancer cases and more cancer deaths than all other skin cancers combined. On average, between 2016 and 2018, 16,744 new cases of melanoma were diagnosed each year in the UK [National Institute for Health and Care Excellence (NICE), 2022].[3,4] The incidence of melanoma is projected to increase by 7% in the UK between 2014 and 2035.

Prognosis is highly dependent on the stage at diagnosis. For people with stage 1 melanoma (thickness is 2 mm or less, no sign that it has spread), the 5-year survival rate is almost 100%, compared with 30% for people with stage 4 melanoma (spread to distant lymph nodes or other parts of the body).

A weighted 7-point checklist is used to assess pigmented skin lesions and determine the need for referral. A pigmented lesion scoring of 3 or more on the weighted 7-point checklist is referred to the suspected cancer referral pathway.[5]

Weighted 7-point checklist:

- Major features of the lesions (scoring 2 points each):
  - change in size
  - irregular shape
  - irregular colour.
- Minor features of the lesion (scoring 1 point each):
  - largest diameter 7 mm or more

- inflammation
- oozing
- change in sensation.

Squamous cell carcinoma

Squamous cell carcinoma is the second most common type of non-melanoma skin cancer. It starts in the cells lining the top of the epidermis (outermost layer of the skin) and accounts for about 20 in every 100 skin cancers (NHS, 2020).[6] Approximately 28,000 SCCs of the skin are diagnosed in England each year.[4]

There is a small risk (up to 5%) of SCC spreading to other parts of the body, such as the lymph nodes (NHS, 2020). The risk of spread with SCCs is greater than with BCCs, especially for people who are immunosuppressed. Death from SCC is rare.

Actinic keratoses are dry, scaly patches of skin caused by damage from sun exposure. There is a small risk that the patches could develop into SCC if untreated (NHS, 2020).[6]

Basal cell carcinoma

Basal cell carcinoma is the most common form of skin cancer and accounts for about 75 in every 100 skin cancers. Approximately 92,000 BCC of the skin are diagnosed in England each year.[4,7]

Basal cell carcinoma does not usually spread to other parts of the body, but if left untreated for a long time, they may get larger and grow deep into the skin and destroy skin, tissue and bone. In rare cases, BCC can spread to other parts of the body and sometimes become life-threatening (NHS, 2020).[6] Death from BCC is exceptionally rare.

Other rare skin cancers

There are 45 other types of non-melanoma skin cancers. Merkel cell carcinoma is rarer and more aggressive than melanoma cancer. It is usually found in the head and neck region. Other types of rare non-melanoma skin cancers can be found in *Appendix 1* of the NICE CSG8 guideline.

Current diagnostic pathway

The initial assessment of a person presenting with a skin condition occurs at the primary care level to determine the appropriate referral pathway. Traditionally, GPs directly referred everyone with suspicious skin lesions to secondary care through the urgent suspected skin cancer referral pathway where all referrals required people to attend a secondary care dermatology department for a face-to-face appointment with a consultant dermatologist. This pathway continues to exist where other clinical pathways are unsuitable or unavailable and is particularly well suited for people with multiple suspicious lesions, a history of skin cancer and other risk factors. *Figure 1* summarises the diagnostic pathway for suspected lesions from the NICE scope.

Urgent suspected skin cancer referral pathway

A person on the urgent suspected skin cancer referral pathway should receive a diagnosis or ruling out of cancer within 28 days of being referred urgently by their GP. For further details, see NHS England's web page on faster diagnosis of cancer.[8] Section 1.7 of the NG12 guideline describes the criteria for an urgent referral for skin cancers (melanoma, SCC and BCC) to the urgent suspected skin cancer referral pathway.[9] These are summarised below.

Sections 1.7.1–1.7.3 of NICE guideline NG12 recommend that urgent referral using a suspected cancer pathway for melanoma should be arranged for people if:

- they have a suspicious pigmented lesion with a weighted 7-point checklist score of 3 or more
- dermoscopy suggests melanoma
- they have a pigmented or non-pigmented skin lesion that suggests nodular melanoma.

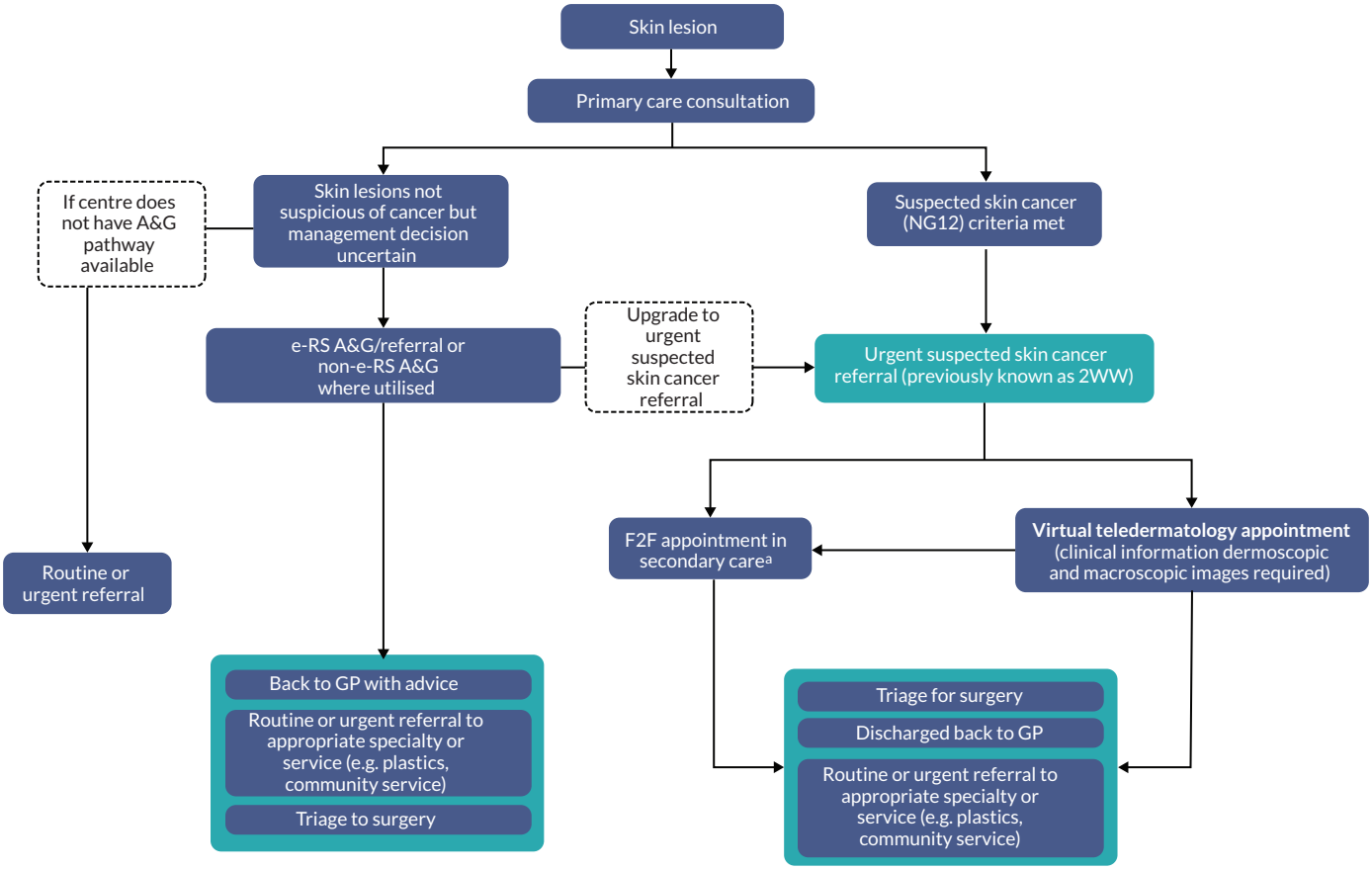


FIGURE 1 Current diagnostic pathway for suspect skin lesions (from NICE scope). A&G, advice and guidance; F2F, face-to-face.

Additional criteria[10–12] also recommend urgent referral if:

- any new persistent skin lesion, especially if growing, pigmented, or vascular in appearance and the diagnosis is unclear
- a new pigmented line in the nail (especially if there is associated damage to the nail), or a lesion growing under the nail
- there is any doubt about the lesion, or there is a history of recent change
- a biopsy has confirmed the diagnosis of malignant melanoma. Note: if a lesion is suspected to be melanoma, an urgent referral to a dermatologist or other suitable specialist with experience of melanoma diagnosis should be made, and excision in primary care should be avoided
- a pigmented or non-pigmented skin lesion suggests nodular melanoma
- any major features in the 7-point checklist, or any features of the ABCDE system.

Section 1.7.4 of NICE guidelines NG12 recommends a person is referred to an urgent suspected cancer pathway if they present with a skin lesion that raises the suspicion of SCC. Section 1.7.5–1.7.6 of NICE guidelines NG12 recommend a routine referral for people if they have a skin lesion that raises the suspicion of a BCC. An urgent suspected cancer pathway referral should only be considered for a lesion that raises suspicion of BCC if there is a particular concern that a delay may have a significant impact, because of factors such as lesion site or size.

As shown in *Figure 1*, a referral to the urgent suspected skin cancer pathway results in either an urgent virtual teledermatology review or an urgent face-to-face appointment in secondary care. If a primary care centre does not have a virtual teledermatology pathway available, the urgent face-to-face appointment pathway is used.

Teledermatology service

Teledermatology refers to the use of static digital images and relevant patient information to triage, diagnose, monitor or assess skin conditions remotely. If a person is referred through the urgent suspected skin cancer referral pathway, clinical information along with a high-quality macroscopic image and dermoscopic images of the skin lesion are required. Images should be taken by a healthcare professional trained in medical photography. Images can be taken:

- in a GP surgery
- at a community diagnostic centre (CDC) close to a person's home
- at a teledermatology clinic based at a hospital.

Images are sent to be assessed by a consultant dermatologist using the teledermatology service and stored in the person's record. The person can be:

- booked directly for surgery
- discharged back to their GP
- referred for a routine or urgent referral to the appropriate specialty or service.

Virtual teledermatology cannot be used for lesions on difficult sites such as palms, soles, scalp and intimate areas, or for people with multiple lesions. Virtual teledermatology is not used for children.

Teledermatology hubs, also referred to as Community Hubs, have been rolled out in a minority of Trusts in the UK. Patients with a GP referral for a suspicious skin lesions are sent to attend a centre in the community where a clinical photographer or healthcare assistant (CP/HCA) captures standardised photographic images of their lesion(s).

Potential positioning of artificial intelligence technologies in the pathway

Artificial intelligence technologies to detect skin cancer could be used at various points in the diagnostic pathway:

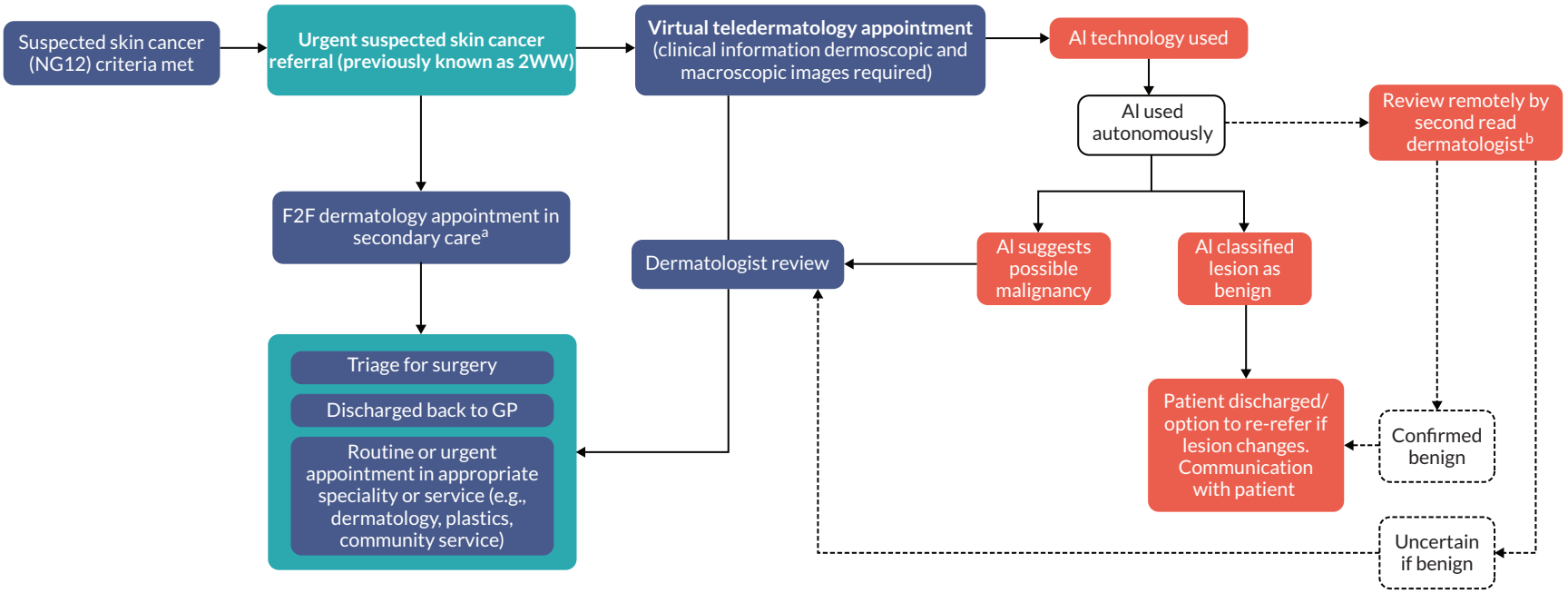
1. By individuals concerned about suspect lesions, prior to consulting a GP.
2. As an adjunctive diagnostic in primary care settings (e.g. by a GP or nurse), to identify lesions that need referral.
3. As an autonomous post-referral assessment between primary and secondary care settings.
4. As an adjunctive diagnostic between primary and secondary care settings (e.g. teledermatology triage settings).
5. As an adjunctive diagnostic in a secondary care setting (e.g. by specialist dermatologists at face-to-face consultations).

This report focuses on settings 3 and 4 but considered evidence from other settings where it informed understanding. This aligns with where DERM is currently being used in a pilot to triage suspicious skin lesions after they have been referred by their GP on the urgent suspected skin cancer referral pathway. *Figure 2* shows a possible pathway for AI use in post-referral that aligns with the current use of DERM. This post-referral assessment is used to identify those with benign lesions to be discharged from the urgent suspected skin cancer referral pathway. People identified with suspicious lesions (cases that contain at least one atypical, pre-malignant or malignant classification) from an AI assessment go on to a review by a specialist in secondary care.

Adjunctive use of artificial intelligence with dermatologist assessment

Artificial intelligence technologies could be used with a dermatologist review. After the AI assessment, a dermatologist will review the results. This is done through virtual teledermatology with the aim of minimising false-negative (FN) results (i.e. cancerous lesions missed by the AI technology).

If the lesion is confirmed to be benign by the dermatologist, the patient is discharged from the pathway. The results are communicated to the patient and primary care referring clinician with safety net information to seek further medical advice if the lesion changes. If the dermatologist is uncertain about the diagnosis or if the AI suggests possible malignancy (whether the AI is used autonomously or not), the images are reviewed by a Trust dermatologist and triaged appropriately.



a, Includes community based 2-week wait (2WW) F2F 'spot clinics'

b, Second read dermatology review currently in place for evaluation and safety; to be removed once evidence safe to do so

FIGURE 2 Proposed positioning of AI technologies in post-referral setting (from NICE scope). F2F, face-to-face.

This 'second read' dermatology review is currently in place for evaluation and safety, but the long-term plan is to remove this and for AI technologies to work autonomously, maximising efficient use of specialist dermatologist's time (see below).

Autonomous use of artificial intelligence

If AI technologies are used autonomously, a lesion classified as benign by the AI technology can be discharged without review by a dermatologist. The patient is discharged from the pathway and the results are communicated to the patient and primary care referring clinician with safety net information to seek further medical advice if the lesion changes. Lesions with suspected malignancy will be transferred to a dermatologist for teledermatology or face-to-face review.

Treatment of confirmed skin cancer

Treatment of skin cancer follows NICE guidance and British Association of Dermatologists guidelines.[12–14] In brief, early-stage melanoma is usually treated by surgical excision; later-stage melanoma may also require lymph node resection, radiotherapy or chemotherapy. SCC and BCC are usually treated by surgical excision, but other treatments, including radiotherapy or chemotherapy, may occasionally be used.

Relevant comparators

The comparator for this assessment was clinical assessment and triage of suspicious lesions through the existing diagnostic pathway without use of AI. This can include assessment by specialist dermatologists either remotely or in person.

Key outcomes addressed as part of the assessment

Outcomes fall into four main areas:

- Diagnostic accuracy.
- Implementation, resource use, and practicality.
- Clinical impact and patient benefit.
- Costs.

Diagnostic accuracy

Diagnostic outcomes are:

- Diagnostic test accuracy [sensitivity and specificity, area under receiver operating characteristic (ROC) curve].
  - Where available, separately for each type of skin cancer (melanoma, BCC, SCC, rare skin cancers).
- Proportion of cancers missed and detected.
- Proportion of benign lesions missed and detected.
- Proportion of referrals confirmed to be skin cancer [positive predictive value (PPV)].

Implementation, resource use and practicality

Key outcomes relate to resource use and timing:

- Proportion of urgent cancer referrals:
  - needing a face-to-face hospital appointment with a specialist for review of lesion
  - converted to routine referral pathway
  - resulting in a diagnostic biopsy
  - booked for surgical procedure
  - discharged back to GP.
- Time to:
  - diagnosis

  - discharge
  - face-to-face consultant appointment
  - treatment (surgery).

- Cancer stage at detection.
- Ease of use/acceptability of AI software by healthcare professionals.
- Number of people consenting to use the technology.
- Test failure rates (with reasons, e.g. image capture issues).
- Proportion of suspicious skin lesions/patients excluded (with reasons, e.g. due to lesion location or scarring).

Clinical impact and patient benefit

- Clinical morbidities.
  - Including distant metastases and adverse outcomes of treatment.
- Mortality.
- Health-related quality of life (HRQoL).
- Non-clinical benefits to patients.
  - Reassurance that lesion is not cancerous.
  - Anxiety associated with waiting for a diagnosis.
  - Acceptability of AI technologies or processes

Costs

Costs were considered from an NHS and Personal Social Services (PSS) perspective. Costs for consideration include:

- Cost of annual subscription for AI software.
- Cost of training healthcare professionals to take images and to interpret AI software results.
- Cost of consultant dermatologist face-to-face appointments.
- Cost of staff time to upload images to AI software platforms and to interpret the results.
- Costs related to missed cancers.
- Costs of consultant dermatology triage team.
- Costs of teledermatology.
- Costs of new services required to support AI technologies (such as establishing new teledermatology services and setting up image capture).

Objectives

The aim of the project was to investigate the clinical and cost-effectiveness of AI technologies as decision aids to triage and diagnose suspicious skin lesions, specifically the two technologies (DERM and Moleanalyzer Pro) described in *Interventions*.

To achieve this, the following objectives were proposed:

Clinical effectiveness

- To perform a rapid systematic review, and if feasible a meta-analysis, of the diagnostic accuracy of the included AI technologies.
- To perform a rapid systematic review with a narrative synthesis of the clinical impact and practical implementation of the AI technologies.
- Based on the results of the rapid review, to identify evidence gaps and formulate recommendations for future research.

Cost-effectiveness

- To perform a rapid systematic review of published cost-effectiveness studies of alternative diagnostic strategies used to aid the diagnosis of skin cancer. This will focus on the included AI technologies but will also include alternative strategies if no evidence is identified for the included technologies.
- To develop a conceptual model that will identify likely drivers of health benefit, harms and cost associated with implementing the included AI technologies in the NHS.
- If evidence and time allows: to develop a budget impact model capturing the direct resource implications of implementing the included AI technologies in the NHS. This may additionally include threshold analysis to explore how health effects or indirect costs may impact cost-effectiveness.

Chapter 2 Methods

This report contains reference to confidential information provided as part of the NICE Diagnostic Assessment process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Systematic review methods

The systematic reviews were conducted following the general principles recommended in Centre for Reviews and Dissemination's guidance and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement[15,16] and its diagnostic test extension.[17]

The review was conducted as a rapid review, aimed at scoping the relevant literature and synthesise studies of key relevance to the UK health setting.

Search strategy

The aim of the literature search was to identify published and unpublished primary studies relating to the use of the proposed AI technologies (DERM and Moleanalyzer Pro) for identifying skin cancer.

An Information Specialist designed the search strategy in Ovid MEDLINE in consultation with the research team. The MEDLINE search strategy was checked by a second information specialist using aspects of the PRESS checklist.[18] This initial search strategy was then divided into two searches so that records highly likely to be about DERM or Moleanalyzer Pro could be identified and screened first. Search 1 contained terms for the two specific technologies and their company names. Search 2 consisted of search terms for skin cancer (in line with those types of skin cancer specified in the NICE scope document) combined with terms for AI and dermoscopy. Both searches were limited to records from 2015 onwards, reflecting the recent development of these technologies.

We note one minor spelling error in the MEDLINE strategy (line 24: 'acitinic' should read 'actinic'). However, this is very unlikely to have led to studies being missed.

Bibliographic databases were prioritised for searching, based upon relevance to the topic area of the assessment. The MEDLINE strategies were adapted to run on all the databases and resources specified in the protocol. The searches were run in October 2023 on the following databases and trial registries: MEDLINE ALL (Ovid), EMBASE (Ovid), Cochrane Central Register of Controlled Trials (CENTRAL, Wiley), the Association for Computing Machinery (ACM) Digital Library, ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform (WHO ICTRP). Records from the searches were imported into EndNote 21 [Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA] for deduplication.

Additionally, company websites were searched to identify relevant publications and other materials relating to the technology. The companies were contacted (via NICE) to provide details of all studies (completed or ongoing) that they have conducted. The search strategies are reported in *Appendix 1*.

Study selection

Titles and abstracts were screened by one reviewer (NU, AL or MS) and random spot checks were performed by a second reviewer to streamline the screening process. Records were prioritised for screening in EPPI Reviewer 6 to assist accurate screening. A cautious and inclusive approach was taken, with all abstracts of uncertain inclusion status checked by a second reviewer. Papers that examined AI technologies but where the technology used was unclear were identified, but did not proceed to full-text assessment.

Full papers of any records that were relevant were obtained and independently screened by two reviewers according to the inclusion criteria listed below. Any disagreements were resolved through discussion and, where necessary, consultation with a third reviewer.

A two-phase scoping process identified relevant studies. At the first phase, all relevant studies (according to the inclusion criteria in *Inclusion criteria*) were identified. A scoping process was then used to identify studies of highest quality and most relevance to the decision problem for full data extraction and synthesis (see *Data extraction* and *Methods of analysis and synthesis*).

Inclusion criteria

Population
People with skin lesions suspicious of cancer, presenting in primary care, local in-person diagnostic clinics, teledermatology, or secondary care settings. The applicability of populations and settings to the NICE scope was assessed and accounted for.

Interventions
DERM (Skin Analytics) and Moleanalyzer Pro (FotoFinder systems) used either alone or in combination with clinical judgement. All versions of the technologies were considered, and their applicability to current NHS practice was assessed and accounted for.

Comparators
Clinical judgement and triage of suspect skin lesions as part of the current diagnostic pathway, without AI use. This included, but was not restricted to, urgent teledermatology services and urgent face-to-face secondary care appointments. The applicability of comparators to the NICE scope was assessed and accounted for. Studies without a comparator were also eligible.

Reference standard
Histological confirmation or rejection of malignancy from a biopsy of the suspect lesion. For unbiopsied lesions, confirmation of non-malignancy by specialist dermatologists, or ground truth as established by panels of dermatologists, was accepted.

Outcomes
See *Key outcomes addressed as part of the assessment* for a full list of intended outcomes.

Study designs
All studies that included adult patients with skin lesions suspicious of cancer, of any design, were eligible for inclusion. Priority was given to studies with prospective recruitment of participants over retrospective reviews. Proof-of-concept, simulations and algorithm training studies were excluded.

Data extraction
A data extraction form was developed and piloted. For the initial scoping process data on intervention, study location and size, setting, type of outcomes reported, design and key quality indicators (randomisation, whether studies are comparative, prospective vs. retrospective design etc.) were extracted by one reviewer and independently checked by a second reviewer.

For studies selected for full data extraction and synthesis, full data on the intervention, patient characteristics and all reported outcomes were extracted by one reviewer and independently checked by a second reviewer. Discrepancies were resolved by discussion, with involvement of a third reviewer where necessary. Where feasible, data were electronically extracted from figures and tables presented in publications using WebPlotDigitizer software (https://automeris.io/).

Data from relevant studies with multiple publications were extracted and reported as a single study. The most recent or most complete publications were used in situations where we could not exclude the possibility of overlapping populations. Where there was evidence that an AI technology has developed or changed over time, only the most recent and complete studies were included. Studies reported as conference abstracts only were excluded.

Quality assessment

At the scoping phase, all studies were assessed for broad quality using the hierarchy presented in *Table 1*.

Prioritised diagnostic accuracy studies that reported sensitivity and specificity were assessed using the quality assessment of diagnostic accuracy studies-2 (QUADAS-2) tool[19] and comparative diagnostic accuracy studies (i.e. with more than one index test) were assessed using the quality assessment of diagnostic accuracy studies-comparative (QUADAS-C) tool, which include items on the quality and applicability of studies.[20] The review team ensured signalling questions for QUADAS-2 and QUADAS-C were relevant to the review question, and input from an experienced clinical dermatologist was sought as appropriate to ensure relevant signalling questions were interpreted appropriately and consistently across assessments. Included studies were assessed by at least one reviewer and checked by a second.

Methods of analysis and synthesis

Scoping review

Initially a scoping process was used to classify identified studies for relevance to the decision problem, based on study quality, setting, outcomes reported and relevance to the NHS and population in the NICE scope (people referred on the urgent suspected skin cancer pathway). The priority hierarchy for the quality of diagnostic accuracy and clinical evidence studies that was used is presented in *Table 1*.

For each included AI technology only, the studies at highest priority level for that technology were taken forward for full data extraction and narrative synthesis. For example, if there are randomised controlled trials (RCTs) of a technology, non-randomised studies were not considered further. Studies at lower priority levels were taken forward if they were of particular relevance to the NHS and the population in scope or report outcomes were not presented in higher-quality studies.

Studies conducted in teledermatology settings, or equivalent early diagnostic clinics, were preferred for full data extraction and synthesis. However, given variation in diagnostic processes in different countries, other settings, including primary care and specialist dermatology clinics, in studies outside of the UK were considered where no evidence in the preferred settings is available.

Narrative synthesis

For studies taken forward from the scoping phase for full synthesis, a narrative synthesis approach was used following appropriate guidance.[21] The results of data extraction for each outcome were presented in structured tables and figures as appropriate, with a text summary. Studies were grouped by population and intervention characteristics. Tabulated results were then compared across studies, interventions and outcomes to identify the broader evidence of effectiveness. Evidence was summarised for specified subgroups (skin colour, skin type and socioeconomic status) where available.

Meta-analysis

Where sufficient data on diagnostic accuracy were available, the EAG had planned to pool data relating to sensitivity and specificity by AI technology using bivariate meta-analytic techniques. As data were insufficient for this, separate meta-analyses of sensitivity and specificity were performed instead, using standard random-effects methods. Subgroup

TABLE 1 Study priority hierarchy for scoping review

Priority level	Diagnostic accuracy	Clinical and implementation evidence
1 (highest)	Prospective cohort comparative studies	Randomised controlled trials
2	Prospective cohorts of AI technology only	Non-randomised cohort studies
3 (lowest)	Retrospective and case-control studies	Retrospective and case-control studies Patient or clinician surveys

analyses were intended for relevant subgroups (skin colour, skin type and socioeconomic status), but no suitable data were available. Heterogeneity was investigated by examining data plots and ROC curve plots.

Methods for synthesising evidence of cost-effectiveness

Relevant cost-effectiveness evidence on the use of AI technologies with class IIa designation (DERM, Moleanalyzer Pro) for early detection of benign skin lesions were identified and narratively summarised. The aim of the review was to examine existing decision-analytic models used to assess the cost-effectiveness of the named AI software options against any comparator, in order to inform parameterisation of a conceptual model to identify key issues, evidence gaps and areas of uncertainty to help direct future data collection and research.

Identifying and reviewing published cost-effectiveness studies

The searches described in *Search strategy* were used to identify relevant economic evaluations of named AI technologies in people with suspicious skin lesions in any setting. Study designs included in the review were budget impact models, return on investment analysis, and other cost-only analyses, as well as full economic evaluations considering both costs and consequences (including cost-effectiveness, cost–utility and cost–benefit analyses).

It was anticipated that no relevant studies would be identified for the named technologies; therefore, additional searches were conducted to identify studies looking at any technology used to aid diagnosis of skin cancer in an NHS setting. The search strategy combined terms for skin cancer with terms for economic evaluations. A search filter was applied to limit retrieval to UK studies,[22,23] along with date limit of 2013 onwards and a limit to studies published in English. MEDLINE (Ovid) and EMBASE (Ovid) were searched on 6 November 2023.

Identified economic models were reviewed and discussed in detail, with the aim of informing the design and parameterisation of conceptual model. Material provided by submitting stakeholders pertaining to the value case for their product was also reviewed.

We aimed to answer the following decision questions on the basis of the identified published evidence, and material submitted by the developers of the included technologies:

1. What are the cost and resource use implications of the use of AI technologies following an urgent suspected skin cancer referral to identify benign skin lesions?
2. What would a health economic model to estimate the cost-effectiveness of AI technologies to identify benign skin lesions in this setting look like, and what are the key evidence requirements necessary to populate such a model?

Development of a conceptual cost-effectiveness model

The structure of a conceptual model for AI tools will be necessarily pragmatic and flexible in terms of the number of different diagnostic and care pathways included, and the two potential use cases for AI technologies in a post-referral setting. The EAG is also clear on the structural limitations of a model of this design, which, while based on precedent, may not be able to provide a granular representation of the diagnostic accuracy and outcomes for the many indications included under the skin cancer umbrella, and may not fully represent the impact of these technologies upon consultant capacity and waiting times, among other important motivating factors for the present assessment.

The conceptual model described comprises an overview of the structure of a cost–utility model for the assessment of AI technologies for the triage of suspected cancer cases referred on the 2-week wait (2WW) pathway. The structure of the conceptual model was designed considering the strengths and limitations of previously published diagnostic models for skin cancer in an NHS setting, and evidence submitted by stakeholders. The exercise sought to identify key inputs necessary for the linkage of short-term diagnostic accuracy metrics with long-term outcome.

The conceptual model was developed in alignment with the NICE reference case and is described in full in *Chapter 6*.

Handling information from the companies

All information submitted by the companies received by the EAG in October 2023 was fully assessed. Information supplied during November 2023 was subject to a more limited assessment. All material supplied was assessed to determine whether it met the inclusion criteria for the reviews. Included studies were data extracted in accordance with the procedures outlined in this protocol.

Chapter 3 Results: diagnostic accuracy and clinical impact

Search results

Figure 3 presents an overview of the study selection process in a PRISMA flow diagram. A first bibliographic search with named AI technology terms was complemented by a second search with no named technologies, references from company submission and hand-searching. A total of 1946 unique records were retrieved and screened. After title and abstract screening, 86 references were retrieved for full-text selection. Six studies, including four evaluations of DERM,[24–27] and two studies of Moleanalyzer Pro,[28,29] were included in the review. In addition, 13 unique studies, including 11 studies of Moleanalyzer Pro,[30–40] and 2 studies of DERM[41,42] that were considered lower priority were

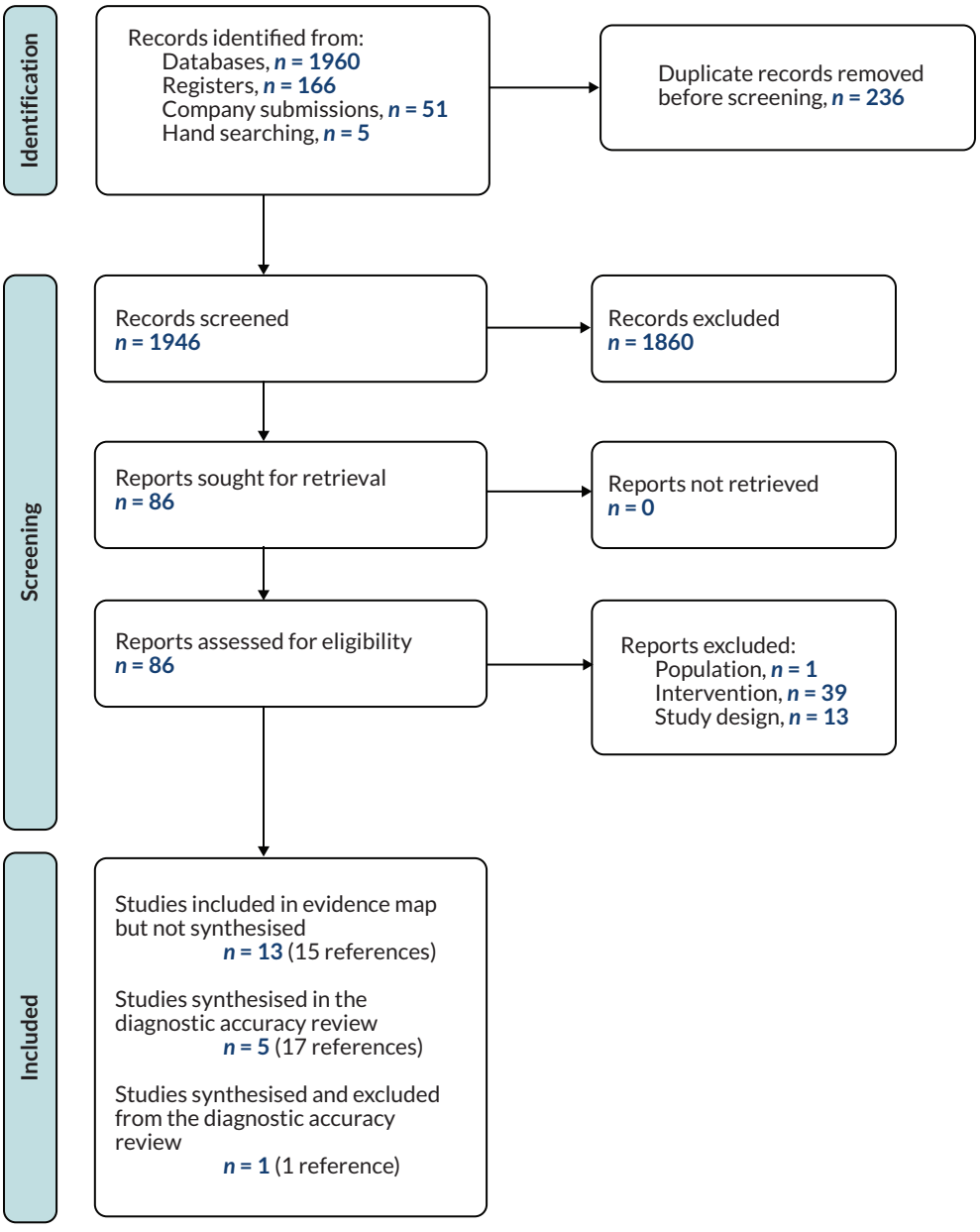


FIGURE 3 Study selection process (PRISMA diagram).

included in an evidence map only, and were not fully synthesised. These studies were classed as lower priority because they either were conducted outside of clinical practice (e.g. retrospective design on selected sample of images) or evaluated an older or outdated version of an AI technology. A list of studies excluded at full-text screening stage is reported in *Appendix 2*. The publications identified from database searches corresponded with those listed on the company website. The submissions from Skin Analytics and FotoFinder did not include any additional eligible studies, although they provided further details from relevant studies not contained in published material.

DERM

Summary of DERM studies

Table 2 summarises the six studies of DERM included in the evidence map. Two studies evaluated an early algorithm version to test the accuracy of DERM for melanoma detection. Phillips (2020)[42] used a retrospective design to train an AI algorithm to detect melanoma from a selected sample of lesions including histologically confirmed melanoma and benign pigmented lesions, along with a meta-analysis of naked-eye examination with or without dermoscopy. Phillips (2019)[41] was a diagnostic cohort where images of suspicious and control skin lesions were collected prospectively in UK hospitals on

TABLE 2 Summary of DERM studies identified

Study	Linked material	Design *N* patients (lesions)	Setting	Period	Diagnostic (index) tests	Outcomes reported
Included in the review						
DERM-003 (NCT04116983)[24]	Marsden[43] Austin[44]	Prosp. DA cohort *N* = 544 (585)	Hospital	June 2020–February 2022	DERM v3.0 (confidential information has been removed) Dermatol.	DA
DERM-005 Chelsea and Westminster (NCT04123678)[25]	Kawsar 2023[45] DERM 2023_Q3[46] Skin Analytics 2023[47]	Prosp. DA cohort *N* = 617 (782)	Hospital	February 2020–August 2021	DERM (confidential information has been removed) Dermatol (TD)	DA Referrals Patient views Economic
UHBFT and WSFT[26]	Andrew[48] DERM 2023_Q3[46] Skin Analytics 2023[49] Jenkins (undated)[50]	Prosp. DA cohort *N* = NR (8571)	Hospital/'TD hub'	July 2021–October 2022	DERM version A (confidential information has been removed) DERM version 'B' (confidential information has been removed)	DA
UHL[27]	Baker 2023[51] Skin Analytics 2023[52] Baker (undated)[53]	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed[a]	DERM (version NR)	confidential information has been removed[b]
Included in evidence map only						
Phillips 2019[41]	NA	Algorithm training and prosp./retros. DA cohort *N* = 501 (551)	Hospital	September 2018–February 2019	DERM[c] (pre-August 2019)	DA
Phillips 2020[42]	NA	Algorithm training + MA *N* = NR (7102)	NA	NR	NR	DA

DA, diagnostic accuracy; Dermatol., dermatologist assessment; MA, meta-analysis; NR, not reported; prosp., prospective; retros., retrospective; TD, teledermatology; UHBFT, University Hospital Birmingham Foundation Trust; UHL, University Hospital Leicestershire; WSFT, West Suffolk Foundation Trust.
a AI TD introduced in March 2022.
b Referrals, procedure duration, waiting time.
c 'Earlier version' than DERM v3, only for melanoma, pre-August 2019 update.

three different cameras, and retrospectively analysed. Both studies were excluded from the main review as they evaluated an earlier version of DERM for the detection of melanoma only, in a selected sample of histologically diagnosed lesions.

A further four studies were identified, including three prospective diagnostic cohorts,[24–26] and a before-and-after study.[51] All evaluated a more recent version of DERM in a post-referral setting in England. Two studies reported being conducted in a 'teledermatology hub' for triage within the 2WW referral pathway.[26,27] These studies included patients with a suspicious skin lesions with a GP referral to attend a teledermatology hub where a CP/HCA captured standardised photographic images of their lesion(s). Following DERM assessment, lesions classed as high risk were triaged to urgent virtual review by a hospital dermatologist. Lesions classed as low risk were sent for remote review by a second reader (consultant dermatologist), who would either discharge the patient if in agreement with AI or overturn the AI risk assessment and proceed with an urgent referral to a hospital dermatologist.

Three studies evaluated sensitivity and specificity of DERM alone against a reference standard that combined histopathology and/or clinical assessment (for non-excised lesions);[24–26] of those, two compared the accuracy of DERM against dermatologist assessment alone concurrently.[24,25] One unpublished study only reported sensitivity estimates for lesions with histopathological diagnosis;[27] however, the study was included in the review as it also reported clinical output outcomes, and clinician and patient views.

Based on clinical trial registration, two completed or ongoing studies with no published results were identified.[54,55] Both are outside the UK, so may be of less relevance to this assessment. These are summarised in *Appendix 3*, *Table 22*.

A number of evaluations are being carried out across the UK in the post-urgent suspected cancer referral setting, as well as in the pre-referral community setting. The company reported in their November 2023 submission to NICE that outcome data for a number of these evaluations were expected in the fourth quarter of 2024. Further details are presented in *Appendix 3*, *Tables 23* and *24*.

Characteristics of studies

Table 3 summarises the characteristics of participants of DERM studies included in the review. Further participant selection criteria are summarised in *Appendix 4*, *Table 25*. Where reported, the large majority of participants were white and very few patients had darker skin (Fitzpatrick types IV–VI). Lesions were most often located on the face and scalp, followed by the chest/back. The proportion of lesions with melanoma was lower in DERM 005, and SCC and BCC rates were higher in DERM 003. No participant characteristic details were reported for the Leicestershire study.[27]

Risk of bias

Results of the quality and applicability assessment are reported in *Table 4*. All studies were at high risk of selection bias due to the exclusion of a significant proportion of participants (15.6–27.4% where reported) that would have otherwise been eligible for assessment in clinical practice.[26] The performance of AI is likely to be significantly improved by the exclusion of some of these lesions (e.g. images with body hair, tattoos, subungual lesions).

Two studies (DERM-005 and Thomas 2023)[25,26] reported separate results for pre-specified thresholds and post hoc thresholds; therefore, the risk of bias was low and high for these respectively. As is standard practice, a significant proportion of lesions did not undergo histology. However, the risk of bias regarding the reference standard was considered to be low in studies that confirmed the absence of cancer using expert consensus and sufficient follow-up. One study (DERM-003)[24] did not report sufficient details on the conduct of the reference standard and was at unclear risk of bias for this domain. There were no significant concerns regarding flow and timing.

All studies raised concerns with regard to the applicability of their populations; the high rate of exclusion of participants with suspected lesions that would normally be seen in practice is a significant limitation. In response to a clarification request, the company noted that the versions of DERM used in all three studies [DERM v3.0 and (confidential information has been removed)] were older than the current version used in the UK (confidential information has been removed) which, among other elements, includes a different set of thresholds for sensitivity and specificity. Therefore, the applicability of the diagnostic accuracy results for DERM to current practice is uncertain. The teledermoscopy devices used in two studies [Dermlite DL1 Basic (DermLite LLC) system][24,25] were considered out of date following

TABLE 3 Participant characteristics of DERM studies included in the review

	N patients (lesions)	Age (range)	% female	Fitzpatrick skin type (%)	Ethnicity (%)	Lesion location (%)	Cancerous lesions (%)[a]
DERM-003[24]	544 (585)[b]	Median 73 (18–97)	50	I: 21 II: 57 III: 20 IV: 1 V–VI: 1	White: 94 Black: 0 Asian: 1 Other: 0 Missing/ NR: 4	Face/scalp: 46 Posterior chest and back: 15 Arms: 14 Legs: 12 NR/missing: 13	Melanoma: 2.7 SCC: 7.5 BCC: 33.7 Other: 0.3
DERM-005 Chelsea and Westminster[25]	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
UHBFT and WSFT[26]	7625 (8571)[c]	NR (18–93)	NR	I: 8 II: 25 III: 18 IV: 3 V–VI: 1 NR: 44	NR	NR	Melanoma: 2.9 SCC: 3 BCC: 7.2 Other: 0.2

NR, not reported; UHBFT, University Hospital Birmingham Foundation Trust; WSFT, West Suffolk Foundation Trust.
a Expressed as % of all lesions with confirmed diagnosis.
b Patient/lesions with DERM assessment and confirmed diagnosis.
c Participants/lesions received DERM assessment with confirmed diagnosis following referral to trust (and second read for lesions classed by DERM as low risk).

clinical advice and therefore raised concern about their applicability to current practice. There were no concerns regarding the applicability of reference standards.

Diagnostic accuracy data from publications

The three fully included studies all reported diagnostic accuracy data for DERM. Studies reported diagnostic accuracy for all melanomas combined and by melanoma type. In all studies, the diagnostic accuracy reported was for autonomous use of DERM, without additional assessment by dermatologists.

TABLE 4 Quality assessment of DERM diagnostic accuracy studies

		Risk of bias				Applicability concerns		
Study	Test	P	I	R	FT	P	I	R
DERM 003	DERM	✗	✓	?	✓	✗	?	✓
DERM 005	DERM	✗	✓/✗[a]	✓	✓	✗	?	✓
Thomas (2023)	DERM	✗	✓/✗[b]	✓	✓	✗	?	✓
DERM 003	Dermato.	✓	✓	?	✓	✗	✗	✓
DERM 005	TD	✓	✓	✓	✓	✗	✗	✓
DERM 003	DERM vs. dermato.	✓	✓	?	✓	N/A	N/A	N/A
DERM 005	DERM vs. TD	✓	✓/✗[a]	✓	✓	N/A	N/A	N/A

Dermato., dermatologist assessment; FT, flow and timing; I, index test; N/A, not applicable; P, patient selection; R, reference standard; TD, teledermatologist assessment.
✓, indicates low risk; ✗, indicates high risk; ?, indicates unclear risk.
a Low risk for the main analyses (pre-specified thresholds for sensitivity and specificity), and high risk for the results of post hoc analyses where the target sensitivity estimates for melanoma, SCC and BCC were amended to match the DERM algorithm to other settings in 'live development'.
b Low for version A (pre-algorithm change), high for version B (post-algorithm change). The algorithm was changed during the study to improve specificity.

DERM-003[24] reported diagnostic accuracy for three different smartphone cameras when used to take images of lesions (iPhone 11, iPhone 6s, Samsung 10). The EAG have chosen to only report results for the iPhone11, as this was the most recently released phone considered. It should be noted that there were variations in diagnostic accuracy according to phone used. It also reported diagnostic accuracy for dermatologists without AI use.

Thomas (2023) [University Hospital Birmingham Foundation Trust (UHBFT) and West Suffolk Foundation Trust (WSFT)][26] reported results separately for Birmingham and West Suffolk centres. It also reported results for two versions of DERM: DERM-vA (used July 2021–April 2022) and DERM-vB (used April–October 2022). As DERM-vB appears to have superseded DERM-vA we only report results for the more recent DERM-vB for this study.

Results for DERM-005 were extracted from a preprint manuscript by Marsden *et al.*[25] This compared DERM to standard of care (dermatologists without AI). (confidential information has been removed). Data were extracted from *Figure 2* of this preprint, which reported the full categorisation of lesions by true diagnosis and test result, from which sensitivity and specificity estimates were calculated.

Figure 4 summarises the diagnostic accuracy from the three included studies.

Meta-analyses of sensitivity and specificity were performed where two or three studies reported data. These were separate, univariate analyses as data were too limited for bivariate meta-analysis. The meta-analysis results are presented in *Table 5*. These results suggest a high sensitivity when using DERM autonomously without assessment by a dermatologist is achievable, and may be higher than achievable using a standard diagnostic pathway without DERM. However, some malignant lesions will still be missed. The specificity of DERM is lower than for dermatologists. In particular, specificity was much lower for detecting SCC and BCC, suggesting that DERM has some difficulty in distinguishing these types of cancer from benign lesions.

In DERM-003, for detecting benign lesions, the sensitivity of DERM was significantly lower compared with face-to-face dermatologist assessment {DERM: 43.9% [95% confidence interval (CI) 37.4 to 50.6]}; dermatologist: 73.9% (95% CI 67.6 to 79.4)], although it had comparable specificity [DERM: 93.3% (95% CI 90.0 to 95.6); dermatologist: 93.7% (95% CI 90.5 to 95.9)]. Hence, around 56% of benign lesions were classified as not benign by DERM, compared with 26% for dermatologists, and approximately 7% and 6% of non-benign (but mostly pre-malignant) lesions were misclassified as benign by DERM and dermatologists, respectively.

Further diagnostic accuracy results from studies of DERM are reported in *Appendix 4*, *Table 26*.

Subgroup data by skin type

Two studies reported separate diagnostic data for Fitzpatrick skin types V and VI. In Thomas (2023), of 159 lesions assessed, 94 lesions had a final diagnosis, including BCC (n = 1) and IEC (n = 1), and AK (n = 1), all correctly referred by DERM (vA or vB).[26] Three atypical nevi were pending face-to-face assessment, and the remainder were benign with a benign specificity of 44.3% (39/88). DERM 003 found no Fitzpatrick skin types V and VI.[24]

TABLE 5 Meta-analysis of diagnostic accuracy from DERM publications

Test	Cancer	Studies	Sensitivity	95% CI	Specificity	95% CI
DERM	Any malignancy	All	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
DERM	Melanoma	All	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Dermatologists	Any malignancy	DERM-003 and DERM-005	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed

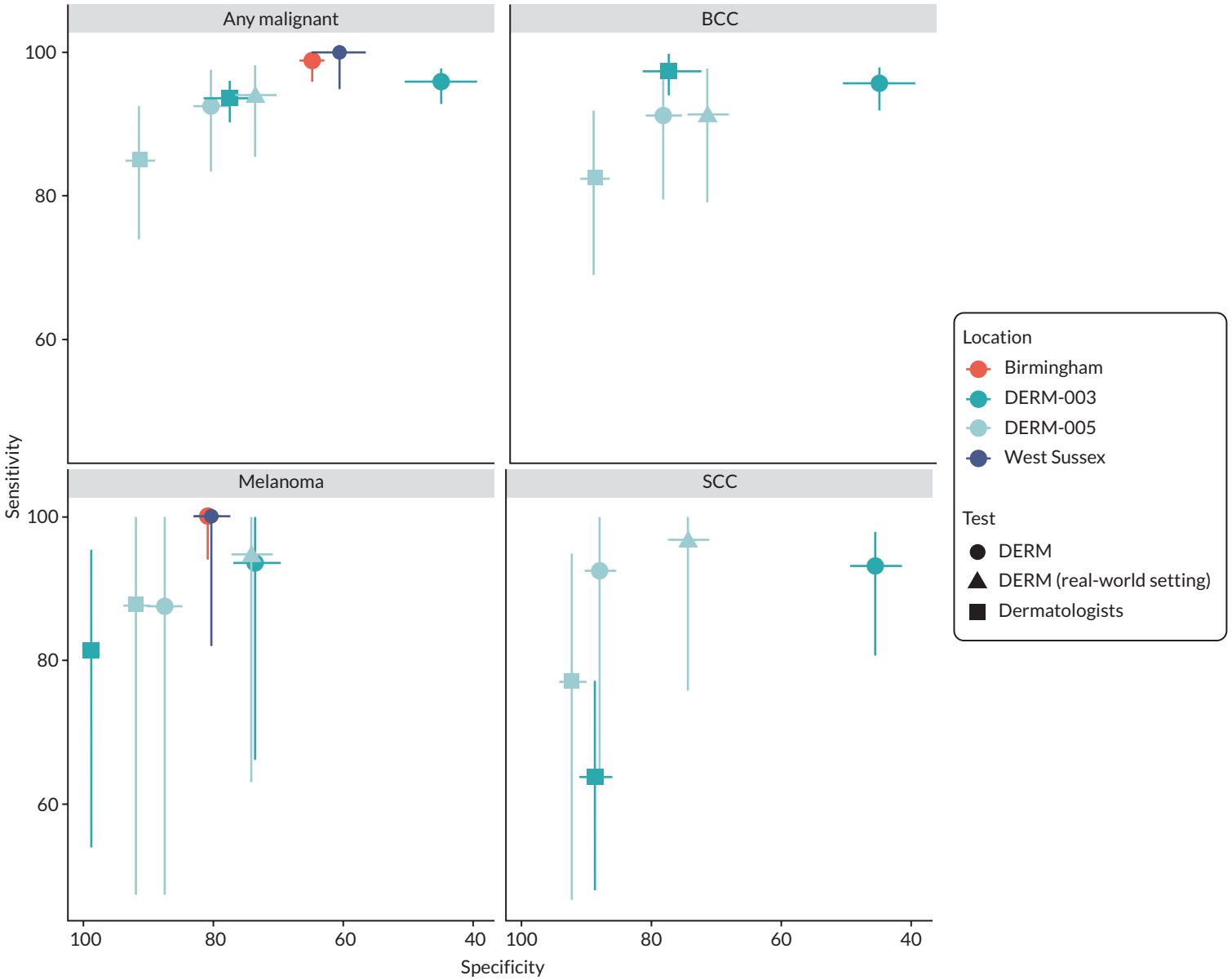


FIGURE 4 Diagnostic accuracy results from DERM publications.

Diagnostic accuracy from unpublished data

In addition to data in published and unpublished papers, Skin Analytics also provided some original data from the Birmingham and Chelsea and Westminster study centres. These data reported all lesions assessed in those centres from April 2022 up to end of September 2023. These data therefore overlap with the data from publications but appear more up to date. The EAG assumes that all patients were assessed using DERM-vB, given the initiation date. We assume that these data include all DERM-vB data from the UHBFT and WSFT study up to October 2022, as reported in Thomas (2023). We assume this includes some patients from DERM-005, although the overlap is unclear.

The supplied data also reported detailed numbers of patients by both DERM results and 'ground truth' diagnosis, enabling a more thorough analysis of the diagnostic accuracy of autonomous use of DERM than was possible using published data. Diagnostic accuracy was calculated for these data in two ways. The 'Exact' analysis considered a DERM result to be a true positive (TP) only if it matched exactly the ground truth diagnosis (i.e. a melanoma diagnosed by DERM was a melanoma; a SCC diagnosed by DERM was a SCC). An 'All malignant' analysis considered a DERM result to be a TP if a malignant lesion was diagnosed as malignant even if not in the correct category (i.e. if a SCC was diagnosed by DERM as a melanoma, or vice versa). This 'All malignant' analysis approximately matched that performed by the company.

Diagnostic accuracy results from combining the Birmingham and Chelsea and Westminster centres are summarised in *Figure 5*, and results for the two centres separately are given in *Figure 6*.

These results show a high sensitivity of DERM for detecting malignant lesions when using the 'all malignant' classification. For example, detecting any malignant lesion had a sensitivity of 96.1% (95% CI 95.4 to 96.8), and sensitivities were 95% or higher for all types of cancer. Sensitivities were similar in Birmingham and London. The specificity for detecting any malignancy was 65.4% (95% CI 64.7 to 66.1). Specificities varied by type of cancer and were slightly lower in Birmingham than in London, but were generally between 60% and 70%. These results are broadly similar to those extracted from publications.

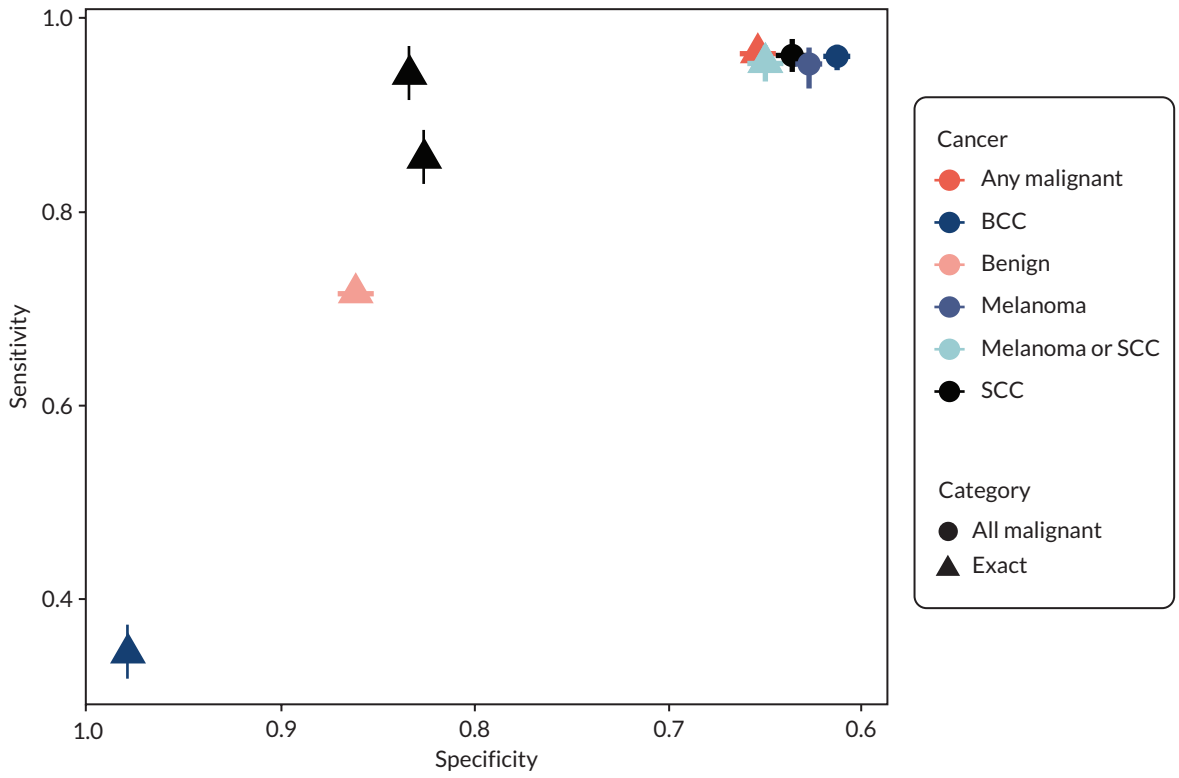


FIGURE 5 Diagnostic accuracy of DERM from pooled Birmingham and London data. Note: The 'All malignant' category considered a DERM result to be a TP if a malignant lesion was diagnosed as malignant even if not in the correct cancer category. The 'Exact' category considered a DERM result to be a TP only if it matched exactly the ground truth diagnosis.

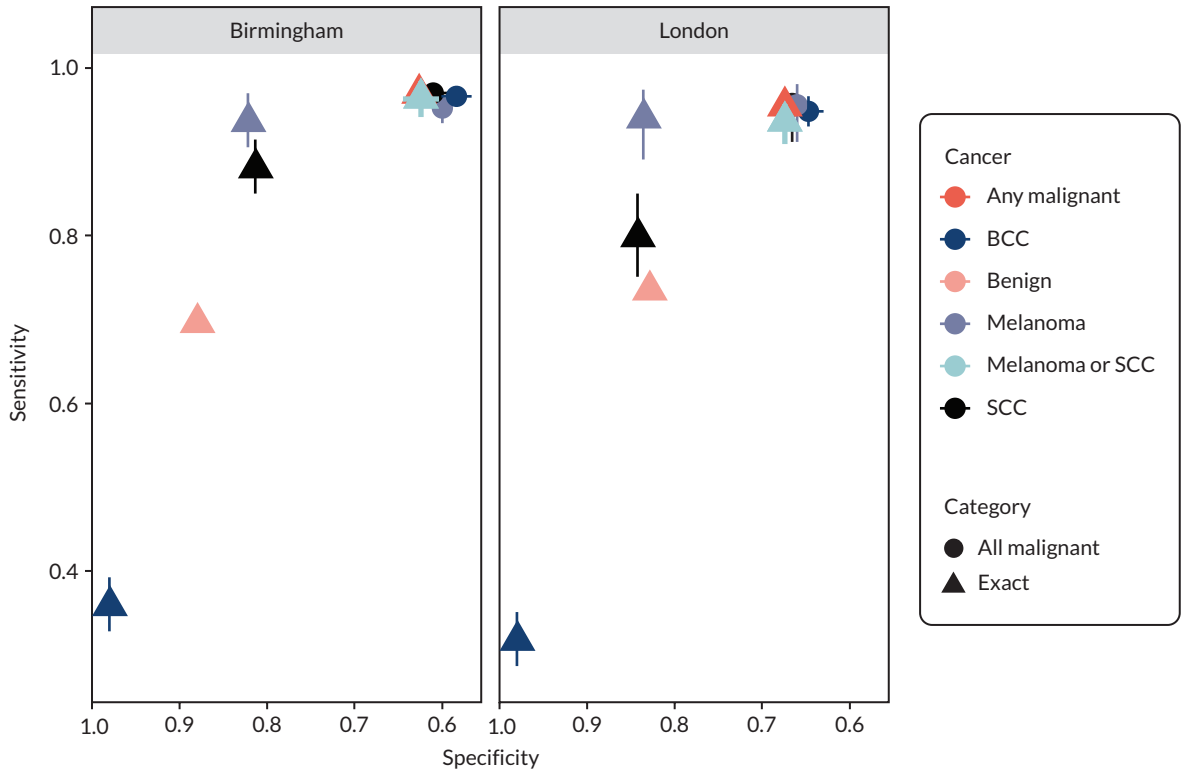


FIGURE 6 Diagnostic accuracy of DERM from separated Birmingham and London data. Note: The 'All malignant' category considered a DERM result to be a TP if a malignant lesion was diagnosed as malignant even if not in the correct cancer category. The 'Exact' category considered a DERM result to be a TP only if it matched exactly the ground truth diagnosis.

When using the 'Exact' classification, there is a decrease in accuracy. For melanoma, the sensitivity remains at near 95%, but for SCC and BCC, the sensitivity declines substantially. This suggests that both SCC and BCC lesions may be misclassified as more serious malignancies by DERM (i.e. SCC as melanoma and BCC as SCC or melanoma).

For the detection of explicitly benign lesions, the sensitivity was 71.5% (95% CI 70.7 to 72.3) for a specificity of 86.2% (95% CI 85.4 to 87.0). Hence, around 28% of benign lesions were classified as not benign by DERM, and 14% of non-benign (but mostly non-malignant) lesions were misclassified as benign.

It should be noted that the reference standard in this analysis was usually a 'ground truth' diagnosis made by dermatologists where the lesion was judged to be non-malignant. Therefore, the diagnostic accuracy of DERM may be slightly incorrect as some genuinely malignant lesions may have been incorrectly classified as benign by dermatologists. This also means that estimates of the diagnostic accuracy of dermatologists without DERM may not be reliable.

Diagnostic accuracy of full teledermatology pathway

Diagnostic accuracy reported in publications and in the original trial data provided relates only to autonomous use of DERM, and not to the full teledermatology pathway, with or without DERM.

Diagnostic accuracy of the full pathway is largely unknown. Data on assessments by dermatologists after DERM assessment were not reported in publications. In all studies, patients who were discharged by a dermatologist were not tested further, so there was no diagnostic reference standard applied.

The unpublished Edge Health report[27] on the Leicestershire study included some data on dermatologist assessment of lesions after the DERM assessment. A summary of these data is given in *Table 6*. This suggests that a 'second read' of lesions classed as benign by DERM (confidential information has been removed) when compared to using autonomous DERM.

TABLE 6 Results of 'second read' assessment in the Leicestershire study

DERM result	After 'second read'	After final assessment by Trust dermatologist	Number of lesions	Number of malignant lesions
Benign	Benign	(*Not used*)	Confidential information has been removed	Confidential information has been removed[a]
Benign	Possibly malignant	Benign	Confidential information has been removed	Confidential information has been removed[a]
Benign	Possibly malignant	Possibly malignant	Confidential information has been removed	Confidential information has been removed
Malignant	(*Not used*)	Benign	Confidential information has been removed	Confidential information has been removed[a]
Malignant	(*Not used*)	Possibly malignant	Confidential information has been removed	Confidential information has been removed

a No reference standard applied; dermatologist assessment assumed correct.

This suggests that using a 'second read' for lesions classed as benign by DERM could (confidential information has been removed). After a final teledermatology assessment, this would (confidential information has been removed). The sensitivity is uncertain because of the lack of a perfect reference standard. However, if the sensitivity of autonomous DERM is 95%, then use of a 'second read' could (confidential information has been removed) based on the Leicestershire data.

Referral status

As the supplied data included full data on number of malignancies, it was possible to estimate how autonomous use of DERM (without a 'second read' by a dermatologist) would impact on onward referrals and discharge rates. For this analysis, it is assumed that all melanoma, SCC or other-non-BCC malignancy cases should receive an urgent referral; BCC and Bowen's disease should receive a routine referral, and all other case should be discharged or treated locally without referral. We note that this may not be exactly what might happen in practice. The results are summarised in *Table 7*.

The results of this analysis suggest that autonomous use of DERM could approximately halve the number of referrals to a dermatologist (among lesions that can be assessed by DERM). However, a small number of lesions, slightly under 1%, would be both malignant and incorrectly discharged (FN). Most of those incorrect discharges would be BCC cases and only 0.2% of lesions would be melanomas or SCC and also discharged.

Most referrals would be false positives (FPs), with around 64% of all referrals being benign lesions. Among urgent referrals, the substantial majority (around 85%) would be FPs. Routine referral would be uncommon (around 9%). This is partly due to a substantial overdiagnosis of BCC cases as being SCC or melanoma.

Implementation, resource use and related outcomes

One study reported data on referral and exclusion rates.[26]

Two studies of DERM reported data that related to implementation outcomes (as listed in *Implementation, resource use, and practicality*).[26,27] Data on these outcomes were mostly taken from the unpublished Edge Health report of patients in Leicestershire.[27] Two studies of DERM reported data on cancer stages as diagnosed by a reference standard. Most melanoma had superficial spreading and had Breslow thickness < 1.0 mm. Most SCC identified were stage 1. Further details are reported in *Appendix 4*, *Tables 27* and *28*.

No evidence, published or unpublished, was identified for numbers of patients transferred to surgery, or test failure rates.

TABLE 7 Percentages of patients by referral status with autonomous DERM use

Group	Percentage of total DERM population	95% CI
Urgent referrals	**39.0**	**38.3 to 39.6**
Correct urgent referrals (melanoma or SCC, TP)	5.8	5.5 to 6.2
Needless urgent referral (FP)	33.1	32.5 to 33.8
Missed urgent referral (FN)	0.3	0.2 to 0.4
Underdiagnoses (urgent referral classified as routine)	0.1	0.1 to 0.2
Routine referrals	**8.7**	**8.3 to 9.1**
Correct routine referrals (TP)	3.8	3.5 to 4.0
Needless routine referral (FP)	4.9	4.6 to 5.2
Missed routine referral (FN)	0.6	0.5 to 0.7
Overdiagnoses (routine referral classified as urgent)	7.4	7.1 to 7.8
All referrals (urgent and routine)	**47.7**	**47.0 to 48.4**
Correct referrals (urgent and routine) (TP)	17.1	16.6 to 17.6
Needless referral (FP)	30.6	29.9 to 31.2
Missed referral (FN)	0.8	0.7 to 0.9
Discharged or treated locally	**52.3**	**51.6 to 53.0**
Correct discharge (TN)	51.5	50.8 to 52.2
Incorrect discharge (FN)	0.8	0.7 to 0.9

TN, true negative.

Referral and exclusion rates

Thomas (2023)[26] reported data on the diagnostic pathway for patients assessed with DERM-vB. This is summarised in *Table 8* and *Appendix 4*, *Table 29*.

There were some differences between the two locations in terms of rate of use of DERM and referral rates, suggesting that use of DERM may vary by location. A notable issue was the substantial number of lesions that could not be assessed using DERM.

TABLE 8 Diagnostic pathway for patients in Thomas (2023) when using DERM-vB

		Birmingham	West Suffolk
Not assessed using DERM		25%	17%
Referred to dermatologist by DERM	Total	44%	62%
	Malignant lesions	7.5%	9.7%
Judged non-malignant by DERM	Total	31%	21.6%
	Discharged at second read	18.7%	10.7%
	Discharged after referral	4.8%	2.7%
	Malignant lesions	0	0

Note
All % are out of total *n* of cases/patients, including those not assessed by DERM.

With DERM vB, between 64% and 76% of lesions eligible for AI assessment and judged non-malignant by DERM were subsequently discharged after second read or referral: none of these lesions that were subsequently biopsied were malignant.

Impact on resource use

In the Leicestershire study resource use data was reported but has not been published (confidential information has been removed).

Timings

In the Leicestershire study timing data was reported but has not been published (confidential information has been removed).

No other data on waiting times, including time to discharge and time to treatment, were reported for any of the DERM studies.

Cancer stage

No evidence on cancer stages at times of diagnosis was identified.

Acceptability to healthcare professionals

One study of DERM (versions not reported) collected feedback from healthcare professionals on benefits and limitations of the tool.[27]

In the study conducted across Leicestershire community hubs, clinicians shared their views on their confidence with DERM, its impact on the trust and on patients. Response rates were not reported. Confidence in DERM was limited among consultants: 33% reported they felt confident when reviewing images of skin lesions taken at the Community Hub, 17% felt confident that AI could reliably distinguish benign and malignant reasons, and 17% agreed that there was no need for a dermatologist (from Skin Analytics) to review lesions classed as benign. A minority of consultants agreed that AI brings benefits for the trust (33%) and for patients (17%).

(confidential information has been removed)

Clinical impact and patient benefit

Clinical morbidities and mortality

The included studies did not report medium- or long-term data on clinical morbidities such as metastases or adverse outcomes of cancer treatment, nor were data on mortality reported.

Health-related quality of life

No data identified in the included studies.

Non-clinical benefits to patients

Two studies of DERM (versions not reported) collected feedback from patients on benefits and limitations of the tool.[27,45] A total of 266 respondents (38.2% response rate) completed questions on their experience with DERM as part of the DERM-005 study.

Reassurance that lesion is not cancerous

In the DERM-005 study, patients expressed confidence in DERM being used on a visual analogue scale from 0 to 100, with higher scores indicating a higher level of agreement. Participants generally responded positively when considering AI as a tool to help doctors, but more cautiously when considering the use of AI to replace a dermatologist. Median levels of agreement with interquartile range (IQR) are illustrated in *Figure 7*.

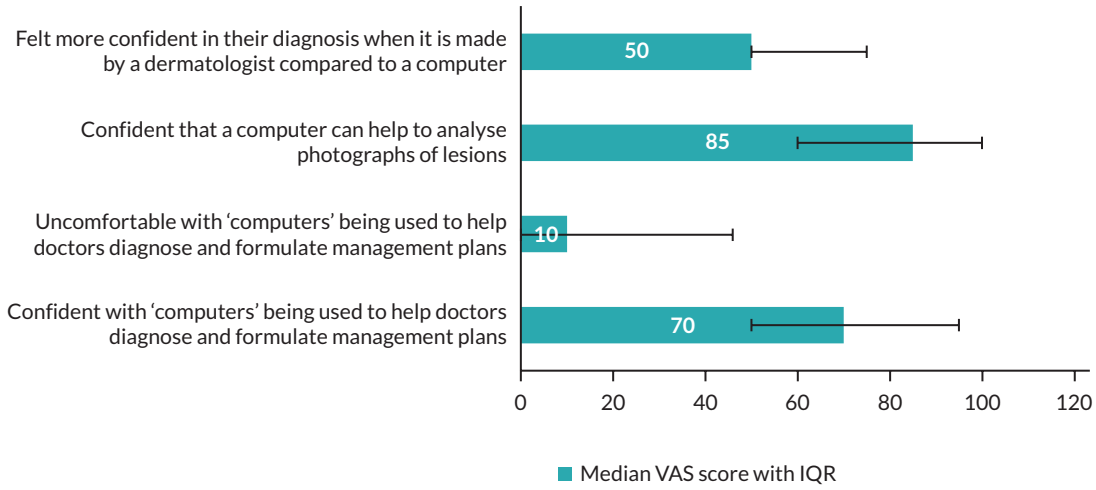


FIGURE 7 Summary of responses DERM-005 relating to confidence in AI diagnosis. Figure created by EAG based on Kawsar (2023),[45] Table 3. VAS, visual analogue scale.

Waiting for a diagnosis and associated anxiety

In the DERM-005 study, the photography service was generally considered an efficient use of patients' time, and respondents agreed that a computer assessing the photographs saves time compared to face-to-face consultation (*Figure 8*). No respondents felt the time needed to take photographs was too long (median score 0), and most would rather have their lesion assessed by a computer than waiting weeks to see an in-person dermatologist.[45]

Of the patient responses in the Leicestershire study, (confidential information has been removed).

Acceptability of AI technologies or processes

In the DERM-005 study, patients generally indicated they felt comfortable with the use of AI and the dermoscopic images required, but there was a mixed response to a statement on preference for a face-to-face dermatologist appointment. No participants found it embarrassing to have photos taken (median score 0, IQR 0–5) (*Figure 9*).

In the Leicestershire study, (confidential information has been removed).

Moleanalyzer Pro

Summary of Moleanalyzer Pro studies

A total of 13 distinct studies of Moleanalyzer Pro were identified, in 2 prospective, cross-sectional diagnostic accuracy cohort studies,[28,29] and 11 retrospective reviews of image data sets.[30–40]

The two prospective cohorts were multicentre studies conducted in a post-referral, secondary care setting; Winkler (2023)[29] was conducted in Germany, and MacLellan (2021)[28] in Canada. Winkler (2023) evaluated the accuracy of Moleanalyzer Pro for detecting melanoma against dermatologist assessment with and without Moleanalyzer Pro in patients with suspected melanocytic lesions; final diagnosis was confirmed by biopsy (in 55% of patients) or clinical follow-up and/or expert consensus. In addition to diagnostic accuracy, the study reported the number of unnecessary excisions, and acceptability of the AI-tool from dermatologists and patients. MacLellan (2021) evaluated the diagnostic accuracy of Moleanalyzer Pro for detecting malignancies against: a dermatologist face-to-face assessment, remote dermatologist assessment, and other non-invasive technologies beyond the scope of this assessment. Clinical management decisions were recorded, and all suspected lesions were excised regardless of the clinical decision or AI output.

Eleven studies performed a retrospective review of existing image data sets to test the diagnostic accuracy of an AI-algorithm against a reference standard test. They are summarised in *Table 9*. Where reported, the Moleanalyzer

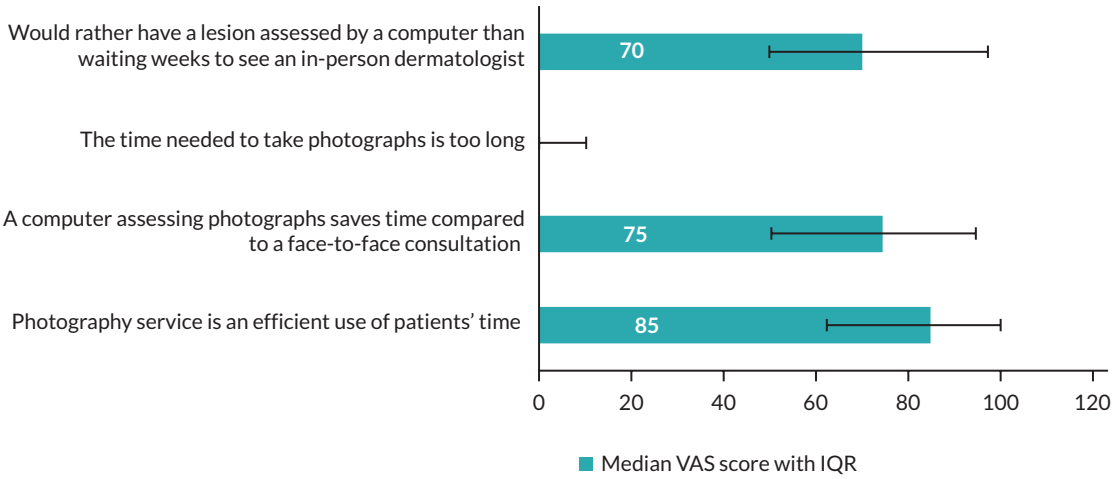


FIGURE 8 Summary of responses DERM-005 relating to waiting for a diagnosis. Figure created by EAG based on Kawsar (2023), Table 3. VAS, visual analogue scale.

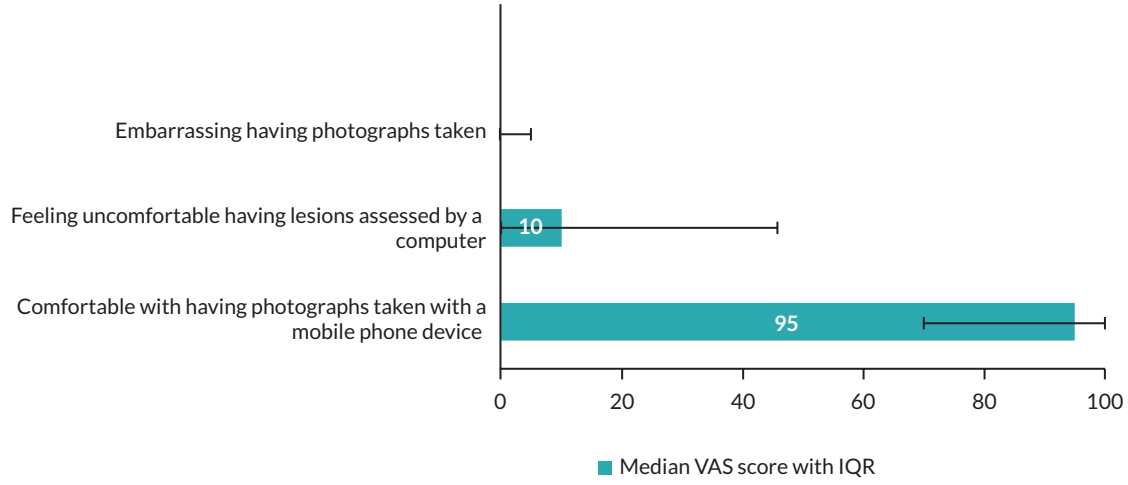


FIGURE 9 Summary of responses DERM-005 relating to acceptability. Figure created by EAG based on Kawsar (2023), Table 3. VAS, visual analogue scale.

algorithm in these studies was based on a modified version of Google's Inception v4 convolutional neural network (CNN) architecture. The reference standard in these studies included histopathology, dermatologist consensus and/or clinical follow-up; four studies analysed only or nearly only excised lesions.[30,32,33,38] Five studies compared the accuracy of AI-algorithm against dermatologists' assessment.[30–33,40] and three were compared against other AI tools.[28,34,40] Due to the lack of prospective evaluation in a clinical setting, 11 studies were excluded from the main review,[30–40] and 2 studies were retained for full data extraction, quality assessment and synthesis.[28,29]

Characteristics of Moleanalyzer Pro studies

Table 10 summarises the characteristics of participants included in MacLellan[28] and Winkler (2023).[29] Further participant selection criteria are summarised in *Appendix 4*, *Table 25*. The large majority of patients had lighter skin colours (Fitzpatrick types II–III). Where reported, lesions were most often located on the trunk, followed by extremities. The prevalence of melanoma (respectively 28.2% and 16.7%) was high in both studies compared with an urgent referral population in the UK.

Risk of bias

Results of the quality and applicability assessment are reported in *Table 11*.

TABLE 9 Summary of Moleanalyzer Pro (FotoFinder) studies included in the evidence map

Study	Design	*N* participants (lesions)	Diagnostic (index) tests	Outcomes
Fink 2020[30]	Retrospective review	72 (72)	Moleanalyzer Pro Dermatologists	Diagnostic accuracy
Haenssle 2018[31]	Retrospective review	NR (300)	Moleanalyzer Pro Dermatologists	Diagnostic accuracy
Haenssle 2020[32]	Retrospective review	100 (100)	Moleanalyzer Pro Dermatologists	Diagnostic accuracy
Kommoss 2023[33,56]	Retrospective review	100 (100)	Moleanalyzer Pro Dermatologists	Diagnostic accuracy Clinical management decisions
MacLellan 2021[28]	Prospective cohort	184 (209)	Moleanalyzer Pro Dermatologist (with/without dermatoscope) Teledermatologists Various TD-AI tools[a]	Diagnostic accuracy Clinical management decisions
Sies 2020[34]	Retrospective review	435 (1981)	Moleanalyzer Pro Moleanalyzer Dynamole	Diagnostic accuracy
Sies 2021[35]	Retrospective review	108 (233)	Moleanalyzer Pro	Diagnostic accuracy
Sies 2022[36]	Retrospective review	465 (1549)	Moleanalyzer Pro	Diagnostic accuracy
Winkler 2020[37]	Retrospective review	180 (780)	Moleanalyzer Pro	Diagnostic accuracy
Winkler 2021[38]	Retrospective review	30 (30)	Moleanalyzer Pro Dermatologists	Diagnostic accuracy
Winkler 2021[39]	Retrospective review	NR (130)	Moleanalyzer Pro	Diagnostic accuracy
Winkler 2022[40,57]	Retrospective review	59 (236)	Moleanalyzer Pro Dermatologists Other AI-tool[b]	Diagnostic accuracy
Winkler 2023[29]	Prospective cohort	188 (228)	Moleanalyzer Pro Dermatologists Both combined	Diagnostic accuracy Unnecessary excisions Dermatologist and patient acceptability

NR, not reported; TD, teledermatologist assessment.
a Teledermoscopy DermEngine, MetaOptima, MelaFind, Verisante Aura.
b Based on resnet34 architecture trained with images from the HAM10000 database.

Both studies were at high risk of selection bias due to the exclusion of participants who would have otherwise been eligible for assessment in clinical practice, including non-melanocytic lesions, and for MacLellan (2023), Fitzpatrick skin types higher than III. The threshold for a positive diagnosis with AI was not reported in MacLellan (2023); therefore, the index test domain was at unclear risk of bias. The reference standard tests were at low risk of bias. As with DERM studies, excision and histology were not performed in all participants in Winkler (2023). However, the risk of bias regarding the reference standard was low, due to the use of clinical follow-up data and expert consensus for non-excised lesions. There was insufficient information from the study to assess the flow of study participants and exclusions from analysis; therefore, this domain was at unclear risk of bias.

Study exclusions in both studies (notably non-melanocytic lesions), the high prevalence of melanoma, and the inclusion of lesions deemed 'challenging' by a dermatologist in MacLellan (2023) limit the applicability of both studies to an urgent referral population. The model of dermatoscope used was not reported in Winkler (2023), and it was out of date

TABLE 10 Participant characteristics of the Moleanalyzer Pro studies included in the review

	Mean age (range)	% female	Fitzpatrick skin type (%)	Ethnicity (%)	Lesion location (%)	Melanoma lesions (%)
MacLellan 2021[28]	52 (31–86)	46	I: 3 II: 60 III: 36 IV–VI: < 1	NR	Head/neck: 24[a] Trunk: 42[a] Extremities: 31[a] Acral: 3[a]	28.2
Winkler 2023[29]	53 (19–91)	48	I: 3 II: 34 III: 56 IV: 6 V–VI: 1	NR	Head/neck: 8 Trunk: 65 Upper extremities: 10 Lower extremities: 15 Acral: 1 Nail: 1	16.7

NR, not reported.
a Only reported for lesions confirmed as melanoma.

TABLE 11 Quality assessment of Moleanalyzer Pro diagnostic accuracy studies

		Risk of bias				Applicability concerns		
Study	Test	P	I	R	FT	P	I	R
MacLellan 2021	Moleanalyzer	✗	?	✓	?	✗	✓	✓
Winkler 2023	Moleanalyzer	✗	✓	✓	?	✗	✓	✓
MacLellan 2021	Dermato.	✗	✓	✓	?	✗	✗	✓
Winkler 2023	Dermato.	✗	✓	✓	?	✗	?	✓
Maclellan	Teledermato.	✗	✓	✓	?	✗	✗	✓
Maclellan	Moleanalyzer vs. dermato. and teledermato.	✗	?	✓	?	N/A	N/A	N/A
Winkler	Moleanalyzer vs. dermato.	✗	✓	✓	?	N/A	N/A	N/A

Dermato., face-to-face dermatologist assessment; FT, flow and timing; I, index test; N/A, not application; P, patient selection; R, reference standard; teledermato., remote dermatologist assessment on images. ✓, indicates low risk; ✗, indicates high risk; ?, indicates unclear risk.

in MacLellan (2023), which limited the applicability of dermatologist assessments. There were no concerns regarding the applicability of reference standards.

Diagnostic accuracy

Winkler (2023) reported the diagnostic accuracy of using Moleanalyzer Pro both with and without clinical input; MacLellan reported results for Moleanalyzer Pro, face-to-face dermatology and remote dermatologist diagnosis alone. The results presented were for the diagnosis of melanoma only; no data were reported for other types of skin cancer or for premalignant and benign lesions. Results for Winkler (2023) and MacLellan (2021) are presented in *Tables 12* and *13* respectively. Winkler (2023) also reported ROC curves of diagnostic performance. These are reproduced, in a simplified form, in *Figure 10*. PPV and negative predictive value (NPV) were not reported.

A meta-analysis of the two studies found that Moleanalyzer Pro alone had a sensitivity of 84.4% (95% CI 73.9 to 91.0) and a specificity of 84.5% (95% CI 72.0 to 92.1) to detect melanoma. In both studies, Moleanalyzer Pro had

TABLE 12 Diagnostic accuracy in Winkler (2023)

	Sensitivity[a]	Specificity[a]	Accuracy[a]
Moleanalyzer Pro alone	81.6 (66.6 to 90.8)	88.9 (83.7 to 92.7)	87.7 (82.8 to 91.4)
Dermatologist alone	84.2 (69.9 to 92.6)	72.1 (65.3 to 78.0)	74.1 (68.1 to 79.4)
Dermatologist with Moleanalyzer Pro	100.0 (90.8 to 100.0)	83.7 (77.8 to 88.3)	86.4 (81.3 to 90.3)

a Results expressed as % and 95% CI.

TABLE 13 Diagnostic accuracy in MacLellan (2021)[a]

	Sensitivity[a]	Specificity[a]
Moleanalyzer Pro alone	88.1 (79.4 to 96.9)	78.8 (71.5 to 86.2)
Dermatologist alone	96.6 (91.9 to 100)	32.2 (18.4 to 46.0)
Teledermatologist alone	89.8 (79.6 to 96.2)	66.0 (57.8 to 73.5)

a Results expressed as % and 95% CI.

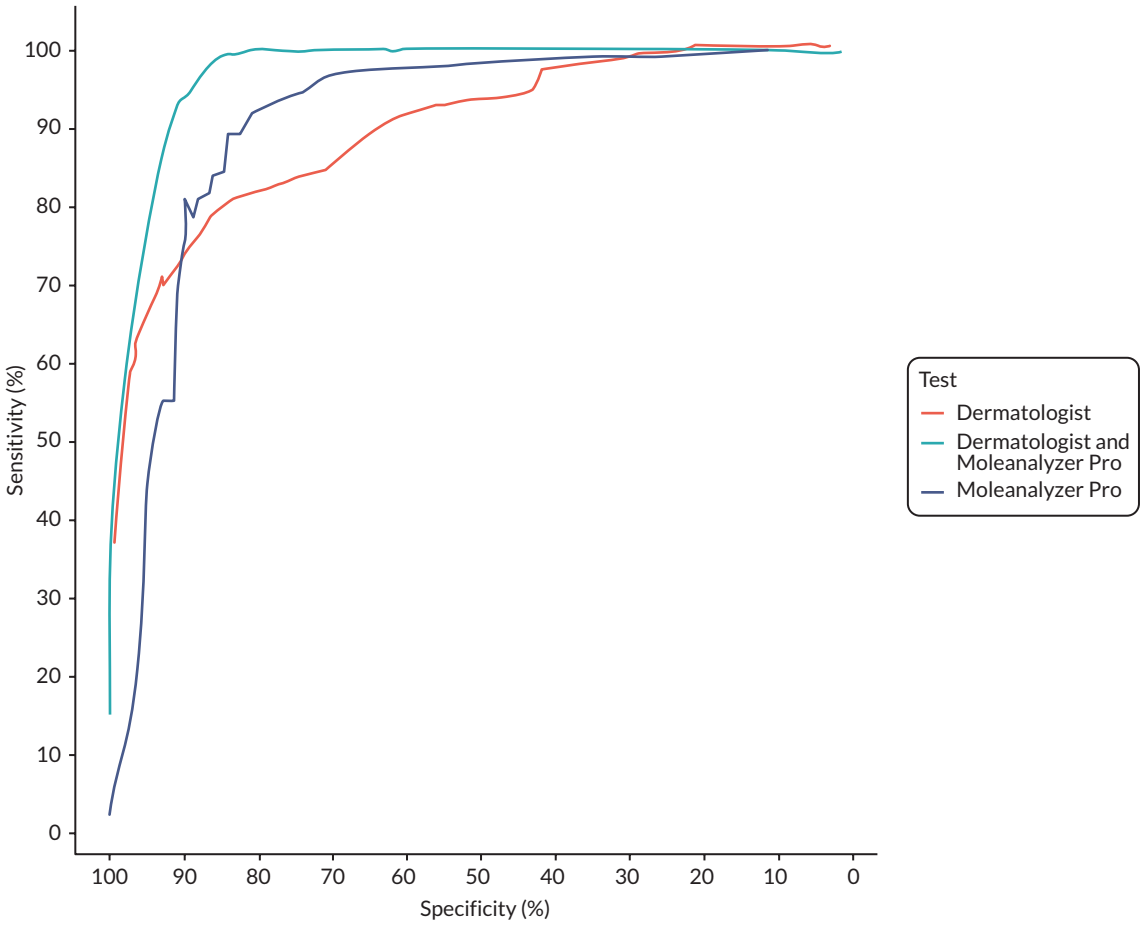


FIGURE 10 Receiver operating characteristic curves for Moleanalyzer Pro for melanoma diagnosis (adapted from Winkler 2023).

somewhat poorer sensitivity, but higher specificity for detecting melanoma than face-to-face dermatologists. Compared with teledermatology, Moleanalyzer Pro had slightly lower sensitivity and higher specificity in MacLellan (2021). Combining Moleanalyzer Pro with dermatologist assessment had higher sensitivity and specificity than assessment by dermatologists alone. The estimated sensitivity is lower than observed for DERM, but the ROC curve in *Figure 10* suggests that Moleanalyzer Pro could achieve a specificity of around 60–75% at a sensitivity of over 95%, which is similar to that observed for DERM (see e.g. *Figure 5*).

The EAG did not identify any evidence on the diagnostic accuracy of Moleanalyzer Pro for the detection of SCC, BCC or malignant lesions in general.

Implementation, resource use and related outcomes

Referral rates

One study of Moleanalyzer Pro evaluated referral decisions with face-to-face dermatology alone and following the integration of AI into decision-making. In Winkler (2023),[29] dermatologists originally recommended the excision of 104 of 190 (54.7%) benign nevi. After reviewing and integrating Moleanalyzer Pro results into decision-making, the estimated rate of unnecessary excisions was reduced by 19.2% from 104 to 84 nevi ($p < 0.001$), while the rate of excision of malignant lesions was not significantly changed ($p > 0.99$). The percentage of nevi managed by follow-up examinations was increased with the integration of Moleanalyzer Pro results into decision-making (from 37.9% to 44.7%, $p = 0.053$).

The EAG did not identify any other evidence from Moleanalyzer Pro studies included in the synthesis on implementation, resource use and related outcomes.

Acceptability to healthcare professionals

Winkler (2023) reported on feedback from dermatologists.[29] Dermatologists were asked after every assessment of a lesion whether or not they judged the CNN scores to be helpful and/or reassuring. For 205 out of 228 lesions, dermatologists completed the evaluation. Out of 205 replies, 159 indicated CNN scores were reassuring (77.6%) and 173 CNN scores were perceived to be helpful (84.4%).

Clinical impact and patient benefit

The EAG did not identify any evidence on Moleanalyzer Pro, published or unpublished, on any clinical outcomes.

Non-clinical benefits to patients

Patients recruited to Winkler (2023)[29] were provided with a questionnaire including 10 statements, based on the 'trust in medical technology' instrument. For each item, response categories indicated the level of agreement with a statement, from very high to none, and undecided. Results are summarised below, pragmatically grouped by categories referring to the outcomes of interest for this report (reassurance, waiting for diagnosis, acceptability), although several items could be considered to contribute to multiple outcomes.

Reassurance that lesion is not cancerous

Responses indicated that patients generally trusted the CNN results (76% very high/high agreement) (*Figure 11*). CNN results were considered trustworthy by 81.5% of respondents (very high/high agreement) and the CNN exam provided a feeling of increased safety for 88.5% of respondents (very high/high agreement). The same level of reassurance was not found when considering autonomous use of Moleanalyzer Pro. When asked whether the AI tool may offer a higher diagnostic quality than a physician, 41.1% of respondents indicated low or no agreement. The overwhelming majority of respondents indicated that they would like the opinion of an expert physician besides an AI-assisted diagnosis (97.8% very high/high agreement).

Waiting for a diagnosis and associated anxiety

Patients were asked whether they would accept longer examination times for an additional CNN-assisted diagnosis, and 33% expressed very high or high agreement with this statement.

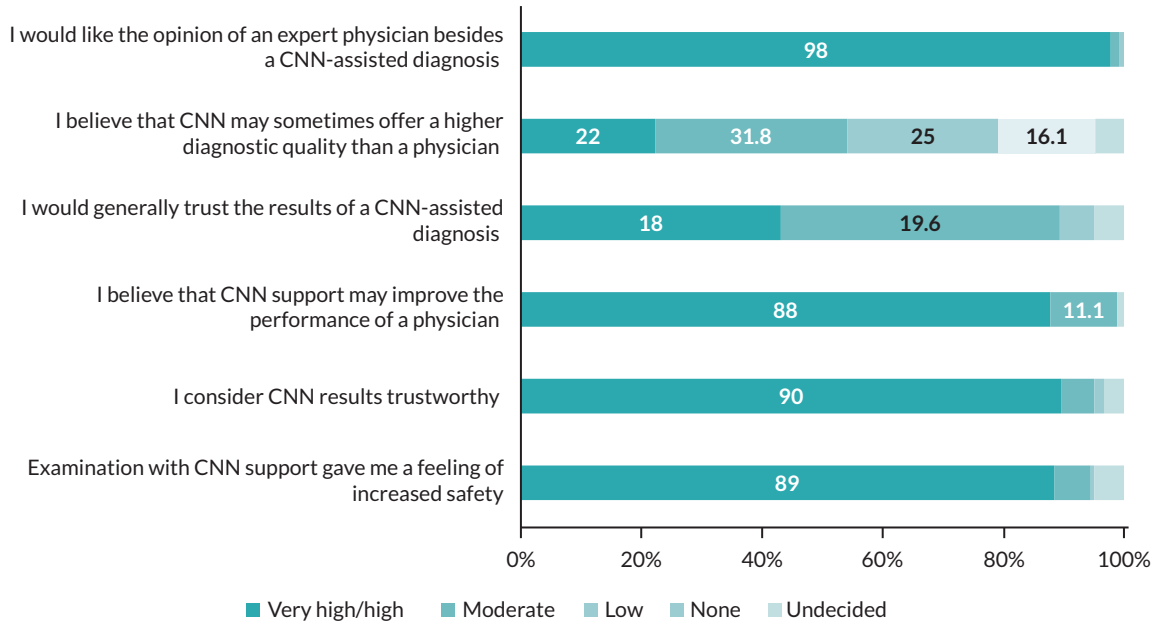


FIGURE 11 Reassurance offered by Moleanalyzer Pro results, percentage agreement. Figure was created by the EAG based on data from Figure S2.[40]

Acceptability of AI technologies or processes

Three questionnaire items related to acceptability of using Moleanalyzer Pro in the diagnostic process (*Figure 12*). Respondents generally did not believe that a CNN may completely replace the examination by a physician (26% moderate agreement, 23% low agreement, 28% no agreement). However, responses relating to the use of AI to assist the diagnosis made by a clinician were more favourable, with patients generally indicating they accepted the use of the tool by clinicians (85% no agreement with statement that CNN should not be used).

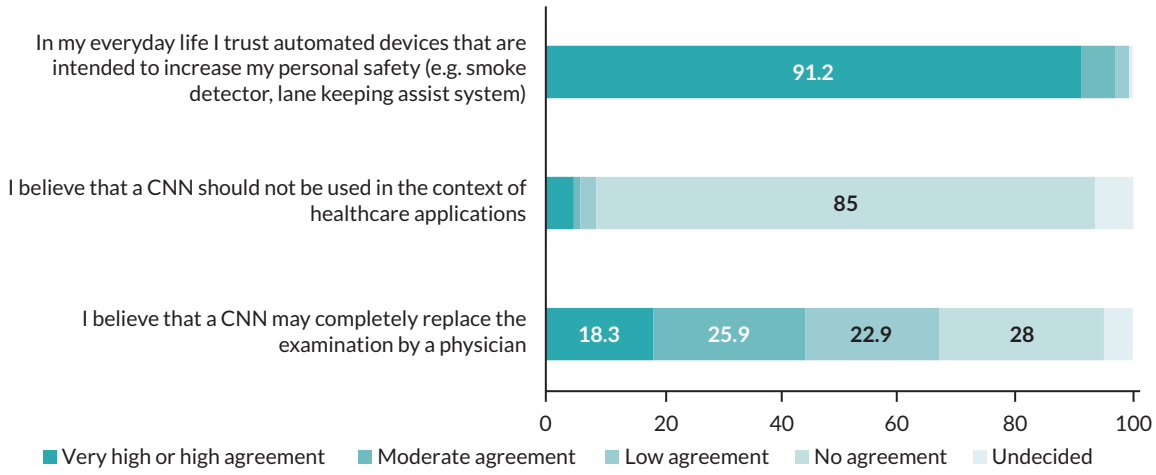


FIGURE 12 Acceptability of Moleanalyzer Pro in diagnostic process, percentage agreement. Figure was created by the EAG based on data from Winkler (2022), Figure S2.[40]

Conclusions

DERM

The review identified three recent studies of DERM that were suitable for assessment, and one currently unpublished study that was also considered. All were performed in the UK and embedded DERM within a post-referral setting.

Diagnostic accuracy

Both published and unpublished data sources for DERM suggested it has a high diagnostic accuracy for detection of malignant lesion when used autonomously: with a summary sensitivity of around 96.1% (95% CI 95.4 to 96.8) for a specificity of around 65.4% (95% CI 64.7 to 66.1). Diagnostic accuracies for detecting specific types of cancer (melanoma or SCC) were similar to this. There was some evidence that DERM might tend to misdiagnose BCC, with many BCC cases being classified as SCC or melanoma. The summary sensitivity when detecting benign lesions was 71.5% (95% CI 70.7 to 72.3) for a specificity of 86.2% (95% CI 85.4 to 87.0).

The diagnostic accuracy for autonomous use of DERM was broadly similar to the diagnostic accuracy of dermatologists. Results are similar to a previous systematic review of dermatology which found a summary sensitivity of 94.9% and specificity of 84.3%.[58]

The diagnostic accuracy of the whole teledermatology pathway including DERM could not be assessed because of a lack of any independent reference standard of diagnosis. This is a key area of uncertainty in assessing the actual clinical value of using DERM.

Clinical outcomes

The EAG identified very limited published evidence on any clinical outcomes. Unpublished data suggested that autonomous use of DERM could approximately halve the number of referrals to a dermatologist (among lesions that can be assessed by DERM). However, a small number of lesions, slightly under 1%, would be both malignant and incorrectly discharged (FN).

(confidential information has been removed)

Patient and clinician perspectives

Some evidence was found for patient and clinician opinions of the use of DERM. Consultants overwhelmingly thought that AI should not be used autonomously, and there was a concern that AI used as a decision-aid was increasing patient time on the diagnostic pathway. However, the evidence is limited to very small samples of responders.

Patients were perhaps more positive than clinicians about the use of DERM alongside a face-to-face diagnostic appointment with a clinician. Patients with experience of having a lesion assessed with DERM were generally accepting of the use of DERM as a tool aiding clinical diagnosis, but up to 50% of patients indicated they preferred a face-to-face dermatology appointment.

Moleanalyzer Pro

Fourteen studies of Moleanalyzer Pro were identified, but only two prospectively evaluated patients in practice and so only they were considered for full synthesis. Neither was performed in the UK. No relevant unpublished material was identified.

Diagnostic accuracy

A meta-analysis of two studies found that Moleanalyzer Pro alone had a sensitivity of 84.4% (95% CI 73.9 to 91.0) and a specificity of 84.5% (95% CI 72.0 to 92.1) to detect melanoma. In both studies, Moleanalyzer Pro had somewhat poorer sensitivity, but higher specificity for detecting melanoma than face-to-face dermatologists. The diagnostic accuracy of Moleanalyzer Pro for the detection of SCC, BCC or other malignant lesions is unknown.

Clinical outcomes

The EAG did not identify any evidence for Moleanalyzer Pro for any clinical outcome.

Patient and clinician perspectives

The use of Moleanalyzer Pro was generally supported by both clinicians and patients, and its results were trusted. However, the overwhelming majority of patients indicated that they would like the opinion of an expert physician besides an AI-assisted diagnosis.

Chapter 4 Results: cost-effectiveness review

Results of literature searches

Two sets of database searches were conducted to identify any cost-effectiveness evidence on the named studies and to inform the development of a conceptual decision-analytic model. The first of these searches was strictly confined to economic studies relating to the use of the named technologies. The second comprised a targeted literature search to identify economic evaluations of any approach to skin cancer diagnosis in an NHS setting. Conference abstracts were excluded from this search. Search strategies can be found in full in *Appendix 1*. Identified studies were summarised narratively. No formal data extraction or quality appraisal was undertaken.

Economic studies relating to the named technologies

Four hundred and seventy-nine records were identified through database searches related to the named technologies. Only one of these records related to health economics review – a clinical trial registration for an economic evaluation of DERM (Skin Analytics), for which there were no corresponding publications or abstracts. As a result, no studies from this search were considered in the literature review.

Economic studies related to diagnostics in skin cancer

The broader search for economic studies relating to the diagnosis of skin cancer in a UK setting returned 999 unique records (date limit of 2013 onwards). Three cost-effectiveness studies were identified following full-text screening, namely Wilson *et al.* (2013), Edwards *et al.* (2016) and Wilson *et al.* (2018).[59–61] These studies were considered relevant to the development and parameterisation of the conceptual model, although none related specifically to adjunctive or autonomous use of AI technologies for diagnosis of lesions suspicious of skin cancer.

Other identified studies

A submission from Skin Analytics provided two unpublished reports relevant to the cost-effectiveness of DERM, with some relevance for the development of the conceptual model. The first of these comprised an evaluation of a pilot of DERM implemented across the University Hospitals of Leicester NHS Trust in the 2WW pathway.[27] The second study comprised a preliminary report describing a de novo cost–utility model produced by the University of Exeter and Skin Analytics. No executable model was made available to the EAG. As these two studies are directly relevant to the decision problem, they are discussed separately in *Chapter 5*.

A report commissioned by the NHSE AI Award group also included economic analyses. This report is only subject to a brief overview in *National Health Service England artificial intelligence in health and care economic evaluations* as it was made available to the EAG only shortly before the end of the project. The documentation provided was also incomplete and did not include an executable model.

Summary of identified evidence

The characteristics of the identified studies are summarised in *Table 14*. All three identified studies were decision-analytic models. In line with inclusion criteria for the broad review of any approach to skin cancer diagnosis, all were from a UK perspective.

Wilson et al. (2013)

Wilson and colleagues developed a decision-analytic model to assess the cost-effectiveness of the MoleMate handheld SIAscopy scanner and proprietary algorithm as a diagnostic aid for primary care clinicians to direct more appropriate referral of pigmented lesions to specialists, compared to current practice. The economic model drew on data generated in the MoleMate UK trial, which enrolled 1293 participants across 15 English general practices.

TABLE 14 Characteristics of identified studies

Study details	Intervention and comparator	Study population, study design, data sources	Costs (perspective, description and values) and outcomes (description and values)	Results: cost-effectiveness
Wilson *et al.* 2013[59] English primary care setting Cost-effectiveness analysis	MoleMate diagnostic aid plus best practice vs. best practice alone	Patients aged 18 or over who have at least one suspicious pigmented lesion, that could not immediately be diagnosed as benign in a primary care setting Study design: modelling study, decision tree with Markov extension design, lifetime time horizon, 3.5% discount rate applied to costs and benefits Source of clinical data: MoleMate RCT *n* = 1293 patients in 15 general practices in the East of England Source of resource-use data: MoleMate trial, published literature Unit costs: NHS reference costs, published literature Utility data: published literature	NHS perspective Costs included: intervention costs including MoleMate device and annual maintenance costs, GP staff time; referral costs and follow-up tests and procedures. Treatment costs associated with TPs were based on 2010 UK guidelines for the management of cutaneous melanoma comprising biopsy excision, staging, and definitive surgery Outcome measure: QALY	MoleMate strategy is estimated to cost an extra £18 compared to best practice alone, and yield 0.01 QALYs per patient. Corresponding ICER is £1896/QALY
Wilson *et al.* 2018[60] United Kingdom Cost-effectiveness analysis	Alternative risk-stratified surveillance policies (based on Williams score) vs. current practice (ad hoc presentation)	UK population Study design: modelling study, patient-level simulation design, 30-year time horizon, 3.5% discount rate applied to costs and benefits Source of clinical data: published literature, expert opinion Source of resource-use data: published literature, guidelines Unit costs: NHS reference costs Utility data: published literature	NHS perspective Costs included: primary care costs, referral, diagnosis, treatment, follow-up and end-of-life costs Outcome measure: QALY	The most cost-effective surveillance strategy (highest net benefit) was for those with a Williams score of 15–21 to be offered a one-off full-body skin examination, and for those with a score of 22 or more to be enrolled into a quinquennial monitoring programme, rising to annual recall for those with a risk score > 43. For implementation of the overall surveillance programme, the ICER was £10,199. Per patient QALYs are improved by 0.016 and costs are increased by £165.
Edwards *et al.* 2016[61] UK Systematic review and cost-effectiveness analysis	VivaScope 1500 and 3000 imaging systems vs. routine management and monitoring	Three study populations were considered: 1. People with suspected melanoma who have equivocal lesions following dermoscopy 2. People with suspected BCC whose lesions have an equivocal or positive result on dermoscopy, to make or confirm diagnosis, as an alternative to diagnostic biopsy 3. Patients with LM prior to surgical management Study design: modelling study, decision tree with Markov extension design, lifetime time horizon, 3.5% discount rate applied to costs and benefits Source of clinical data: systematic literature review of available evidence of VivaScope Unit costs: company data, NHS tariff and reference costs Utility data: published literature	NHS and PSS perspective Costs included: intervention costs (including equipment, maintenance, consumables, staff training, staff time), comparator costs (biopsies, histological examination, monitoring, clinician time), costs associated with management of positive and negative results, and future health events (e.g. recurrence, progression) Outcome measure: QALY	Where VivaScope is used exclusively in the melanoma population, the ICER was between £8877 and £19,095. Incremental health was improved by 0.009–0.016 QALYs, incremental costs were between £138 and £178 (ranges indicate use of different diagnostic accuracy data). When also used for other indications, VivaScope becomes the dominant strategy in this population For use exclusively in the BCC population, results show a dominant strategy – average per patient costs are reduced by £52 and QALYs are increased by 0.011. In the LM population, the model indicates a cost-effective strategy (ICER: £10,241) where VivaScope is used only for LM mapping where average per-patient costs are increased by £70.75 and QALYs are increased by 0.007. Where VivaScope is used across indications, per-patient average costs are reduced by £74.12 and QALYs increased by 0.007 (indicating a dominant strategy).

ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year.

Model structure

The authors adopted a decision tree model structure to capture the initial decision to *refer* or *not refer* patients to specialist care. Three Markov models were used to estimate the long-term costs and health outcomes of patients based on their diagnosis at the terminal nodes of the decision tree [i.e. TP, FN, and true negative (TN)/FP]. The reference standard (i.e. the definition of an appropriate referral) was whether secondary care clinicians decided to biopsy or monitor a lesion – matching the reference standard used in the trial. The model did not structurally distinguish between melanoma and other types of skin cancer, and it only accounted for disease stage at diagnosis.

Patients correctly identified (TPs) were assumed to be appropriately treated at the point of diagnosis, and thus remained within the same Markov state according to stage at diagnosis until death, that is treated patients cannot experience progression. Patients with a FN diagnosis similarly entered the Markov model according to stage at diagnosis, but could experience disease progression, could be diagnosed and treated (entering a corresponding Markov state according to their post-treatment prognosis by stage at diagnosis), or could die. Patients without cancer (correctly identified or not) simply followed a normal life expectancy with zero cost or health consequences.

Mechanism of cost consequences

For patients in both arms of the model, two potential outcomes at primary care were possible: referral or non-referral to secondary care. Improved specificity reduced the number of (inappropriate) referrals and therefore reduced costs associated with follow-up investigations in secondary care. In the model, the specificity of MoleMate was lower than that of best practice (82.1% vs. 89.2%), suggesting increased costs versus current care. Improved sensitivity had the effect of increasing immediate costs associated with follow-up investigations and treatment but lowered the cost associated with treatment of later-stage disease from initially unidentified melanomas. As treatment costs differed by disease stage at diagnosis (i.e. higher treatment costs for more severe disease), the net cost impact of improved sensitivity depends on the scale of the cost difference between treating early and late-stage disease and the effect of discounting. The MoleMate system was more sensitive than current practice (98.4% vs. 95.6%). Use of the MoleMate system itself was associated with a small additional cost (about £14) and the MoleMate strategy increased average patient costs by £18, suggesting little cost impact beyond the cost of the device itself.

Mechanism of health consequences

Staging of disease at diagnosis was based on the American Joint Committee on Cancer (AJCC) Melanoma Staging Database report from 2009.[62] The distribution of disease stage at diagnosis was implicitly assumed not to be affected by underlying diagnosis. Incidence of malignancy in patients referred to specialist care was 5%, the majority of whom had stage 1a/b disease.[62]

Undiagnosed disease was associated with a 70–80% annual probability of remaining at the current stage, a 10% probability of being detected in a given year, and a 10–20% probability of progressing one or more stages. These transition probabilities were based on an earlier cost-effectiveness model for screening of melanoma.[63] Health state utilities were derived from a 2004 conference abstract – Bendeck *et al.*[64] stage 4 disease was associated with the most significant quality-of-life impact (no cancer: 1.00 vs. stage 4: 0.52) where other stages were associated with more modest impacts on quality of life. Disease prognosis worsened commensurately with disease stage, with risk of death calculated using a log-odds ratio versus stage 1a melanoma where patients were at greater risk of mortality in later stages. Patients with stage 4 disease had a log-odds ratio of death of 5.743, based on the AJCC report. Given the opportunities in the model for disease progression, and the worse outcomes associated with later disease stages in terms of mortality and quality of life, missing a case of cancer at the point of diagnosis has a negative health consequence in the model.

As MoleMate was associated with increased sensitivity compared to current practice (98.4% vs. 95.6%), more patients with skin cancer were correctly referred to specialist care and were subsequently treated, generating a small quality-adjusted life-year (QALY) benefit of 0.093 versus current practice. This improvement in patient health offsets cost increases associated with lower specificity of the MoleMate system (as described in *Mechanism of cost consequences*). MoleMate was associated with an incremental cost-effectiveness ratio (ICER) of £1896 per QALY gained in the base-case analysis compared to current practice.

Edwards et al. (2016)

Edwards and colleagues performed a systematic review and economic evaluation to evaluate the clinical and cost-effectiveness of VivaScope 1500 and 3000 for the diagnosis of equivocal skin lesions and in lesion margin delineation prior to surgical excision. VivaScope is a technology designed to be used in conjunction with dermoscopic examination to aid diagnosis of suspicious lesions.

For the evaluation of cost-effectiveness, three 'part' models were built covering three populations: (1) people with suspected melanoma who have equivocal lesions following dermoscopy, (2) people with suspected BCC whose lesions have an equivocal or positive result on dermoscopy, to make or confirm diagnosis, as an alternative to diagnostic biopsy, and (3) patients with LM prior to surgical management. Only the first two of these models related to the diagnosis of suspicious lesions and therefore are the models relevant to this review; both are discussed in detail below.

Model structure: diagnosis of melanoma

The authors employed a decision tree model structure to calculate the short-term outcomes of patients with suspected melanoma with equivocal lesions following dermoscopic assessment and used Markov models to represent the long-term outcomes of patients.

The current practice arm of the model details the current patient pathway – some patients directly undergo biopsy and excision (where melanoma status is confirmed) and some undergo monitoring. Monitoring can result in referral for biopsy and excision if melanoma is suspected or discharged if not. Given that biopsy is considered the gold standard test, the melanoma status of all patients who have undergone biopsy and excision is ultimately known. Some patients without melanoma are biopsied unnecessarily (and suffer associated health losses and procedure costs). In the VivaScope arm of the model, all patients undergo an examination with VivaScope, where positive cases are excised and biopsied, while negative cases are discharged without further investigation or treatment. Those patients with a positive VivaScope result (or patients undergoing biopsy in standard care) without melanoma will have unnecessarily undergone biopsy (and its associated harms) and those with melanoma who tested negative at VivaScope (or discharged at monitoring) will have been discharged inappropriately. Patients then enter one of three Markov models based on their diagnostic outcome and true disease status.

The first Markov model represents the outcomes of patients who are correctly identified as having melanoma (VivaScope TP or identified using biopsy). Identified melanomas were assumed to be identified as either in situ (60% of lesions) or stage 1 (1a or 1b) (40% of lesions). A number of key assumptions were made:

- following identification and treatment, melanomas were assumed not to progress
- patients with identified melanomas had a reduction in their HRQoL applied as a one-off disutility at treatment which then returned to that of the general population
- patients with lesions on their head and neck experienced an additional permanent reduction in HRQoL due to scarring following excision and biopsy
- patients with an identified melanoma stage 1b were at increased risk of mortality for 10 years, returning to the general population thereafter.

The second Markov model represents FNs with melanoma. Key model assumptions were as follows:

- all melanomas were assumed to be in situ or stage 1 (1a or 1b) at the time of assessment
- melanoma could progress by only a single stage
- all unidentified melanomas are identified when they reach stage 2 (2a, 2b, 2c), or within 5 years of the first assessment
- people with an unidentified melanoma had a HRQoL equal to that of the age-adjusted general population until their melanoma was identified (when a decrement is applied)
- people with a lesion on their head and neck experienced an additional permanent reduction in their HRQoL due to scarring following excision and biopsy
- unidentified melanomas did not incur any costs
- melanoma was assumed to be successfully treated upon diagnosis (no further progression)

- people with an unidentified or identified melanoma at stage 1b or 2 were at increased risk of mortality, from the outset of the model until 10 years after diagnosis, after which point mortality risk was equal to that of the general population.

The third Markov model represented people without melanoma (VivaScope FP or TN, or negative following biopsy/monitoring). Key assumptions were as follows:

- people with a lesion on their head and neck experienced an additional permanent reduction in their HRQoL due to scarring following excision and biopsy
- otherwise, HRQoL was equal to that of the general population of the same age.

Transition probabilities in the model were informed by assumptions regarding the progression of patients. A progression probability of 15.3% was used, calculated based on the assumption that the mean duration of stage 1 melanoma is 50 months, and 50% melanomas progress. This transition probability was applied to progression from both in situ to stage 1, and stage 1–2. An annual probability of opportunistic diagnosis (given initial non-identification) of 35% was applied based on the assumption that all unidentified melanomas would be diagnosed by the time they reach stage 2 at the latest, and it was structurally imposed that yet unidentified melanomas at 5 years were diagnosed.

Mechanism of cost consequences: diagnosis of melanoma

The costs associated with the VivaScope pathway in the model depend on (a) whether the VivaScope technology is used for other potential indications (and thus the fixed costs of VivaScope are spread over a larger population) and (b) which diagnostic accuracy figures are used in the model (Alarcon *et al.*[65] vs. Pellacani *et al.*[66]) where the former has the greatest impact on pathway costs. If VivaScope has higher sensitivity, more cases will be identified correctly, and treatment costs will be higher.

In the model, there is little difference in treatment costs across disease stages [except for the increased use of sentinel lymph node biopsy (SLNB) in later stages] and therefore identification of melanoma at an earlier stage is unlikely to drive value in terms of reducing modelled treatment costs. This contrasts with Wilson *et al.* (2013),[59] where there was a steeper gradient in costs. Higher specificity for VivaScope results in fewer non-melanoma patients undergoing unnecessary excision and biopsy (thus costs saved – excision and biopsy is £151). Given that monitoring is not available on the VivaScope pathway, monitoring costs are also saved (£93). The incremental cost results are shown in *Table 15* and are shown according to different diagnostic accuracy inputs and whether VivaScope is used for melanoma only, for the two diagnostic indications or for all three indications. The cost of the device itself appears to be a large driver of pathway costs.

Mechanism of health consequences: diagnosis of melanoma

A more specific VivaScope test will reduce the number of patients undergoing excision and biopsy – this reduces the number of patients experiencing anxiety while waiting for biopsy results, and the number of patients experiencing permanent disutility from scarring on their head and neck.

Unlike other studies in this review, the key driver of value in the model appears to be the reduction in health harms associated with biopsy and excision used for the detection of melanoma. In the model, under routine management, 67% of lesions were excised despite a prevalence of melanoma of only 15%. Given the large health decrement applied in the model following biopsy and excision, the main value case of VivaScope appears to be the reduction of the use of biopsy

TABLE 15 Incremental costs associated with VivaScope pathway

	Alarcon diagnostic accuracy data	Pellacani diagnostic accuracy data
VivaScope for melanoma diagnosis only	£137.99	£177.03
VivaScope for melanoma and BCC	−£52.71	−£13.67
VivaScope for all indications	−£56.95	−£17.91

and excision and its associated harms. It is unlikely that earlier diagnosis of melanoma would be a key driver of patient health in the model, as it assumes that patients with unidentified melanomas have HRQoL equal to that of the general population until diagnosis, when a one-off decrement applies at the point of treatment (in addition to a permanent decrement for some patients with scarring). This contrasts with other studies such as Wilson *et al.* (2013)[59] which assumed a persistent impact on HRQoL for treated patients (thus making the health consequence for a missed case higher). The fact that structural limitations are placed on progression in Edwards *et al.* (2016)[61] also means that patients with unidentified cancers are unable to progress to later stages and incur greater health decrements, again meaning that the health consequence is lower for a missed case of melanoma.

The incremental QALY results associated with each pathway are shown in *Table 16*.

Model structure: diagnosis of basal cell carcinoma

A decision tree model structure was used to estimate the short-term outcomes of patients with suspected BCC lesions and positive or equivocal findings in dermoscopy. According to the model structure, patients with lesions suspicious of BCC are examined either according to current practice (who all receive diagnostic biopsy) or with VivaScope. Given that diagnostic biopsy is the gold standard for diagnosis, all patients in the current practice arm have their treatment status correctly determined and are treated or discharged accordingly. Diagnostic biopsy incurs a one-off disutility related to the procedure (−0.02), a 6-week disutility related to anxiety while waiting for biopsy results (−0.008), and a permanent disutility for 5% of patients with scarring on their head or neck (−0.016). Although all patients are appropriately treated or discharged in the current practice arm, some patients undergo unnecessary diagnostic biopsy and experience a utility decrement. In the VivaScope arm, patients testing positive at VivaScope progressed to treatment (without the need for diagnostic biopsy). Patients for whom VivaScope indicated a negative result received diagnostic biopsy (because the original dermoscopic outcome suggested malignancy) and are discharged or treated as appropriate. All patients with the BCC in the VivaScope arm were correctly treated. A proportion of patients who tested positive at VivaScope will be FPs and therefore will have been inappropriately treated.

Treatment in the model comprised both surgical and non-surgical therapies. Patients undergoing surgical treatment (75% of patients) experienced a utility decrement (−0.004) from the procedure itself, and a proportion experienced a permanent disutility associated with scarring (−0.019 for surgical excision, −0.021 for Mohs surgery). Given that all patients in both arms of the model with BCC are correctly identified as having the condition, only one Markov model is required for those patients who have experienced scarring from unnecessary biopsy as the net difference in long-term treatment outcomes between arms is because of scarring.

Mechanism of cost consequences: diagnosis of basal cell carcinoma

The immediate mechanism by which VivaScope impacts costs in the model is by reducing diagnosis costs – the cost of biopsy in the model is £134, whereas the cost of VivaScope is £70 (exclusive use on BCCs). This cost benefit in favour of VivaScope will be somewhat reduced by the unnecessary treatment costs incurred through treatment of FP patients at VivaScope.

Mechanism of health consequences: diagnosis of basal cell carcinoma

It appears that the mechanism by which health is impacted in the model is through the avoidance of health harms associated with diagnostic biopsy, which carries a large health decrement in the model associated with anxiety while waiting for results, the procedure itself, and scarring. The driver of the value of VivaScope appears to be as a result of the fact that it is non-invasive (unlike diagnostic biopsy) and so not associated with scarring and not associated with a long wait for results, thus no anxiety-related decrement. VivaScope allows some patients (who test positive) to proceed

TABLE 16 Incremental QALYs associated with VivaScope pathway

	Alarcon diagnostic accuracy data	Pellacani diagnostic accuracy data
VivaScope arm	0.016	0.009

directly to treatment without biopsy, appearing to generate value in the model through a reduction in the proportion of patients receiving diagnostic biopsy.

Given that all patients with BCC in both arms are treated, there is no mechanism for health gains associated with improving identification of disease in the model.

The model results showed that where VivaScope is used exclusively for suspected melanomas with equivocal dermoscopy, the ICER is £8877–19,095 depending on clinical data used. When also used for other indications, VivaScope becomes the dominant strategy in these patients. For use exclusively in the BCC population, results show a dominant strategy.

Wilson et al. (2018)

The authors adapted a previously developed decision-analytic model (Wilson *et al.* 2013[59]) to evaluate the potential cost-effectiveness of a risk-stratified population surveillance programme. The authors estimated the costs and outcomes associated with surveillance strategies of different risk groups. The population was segmented by Williams score, a clinical tool for identifying the risk of melanoma. The main purpose of this study was to identify the risk score cut-off at which it is most cost-effective to enrol patients into a surveillance programme consisting of (1) a one-off visit to the patient's primary care practitioner, (2) an ongoing primary care-based monitoring programme (and the optimal frequency of visits). The authors estimated outcomes over 30-year time horizon.

Model structure

The authors employed a patient-level simulation model based on the structure of Wilson. The model is comprised of two 'modules' – patients enter the model in the natural history module according to the distribution of prevalent melanomas and their disease stages. When contact is made with the health system, the patient enters the clinical module which has a decision tree-like structure where referral, treatment and discharge decisions occur. The clinical module allows patients to present in primary care: both of their own initiative and if they are told to do so following a risk assessment. Following presentation, any suspicious moles are inspected at primary care and a decision is made to either refer to secondary care or discharge the patient. The model categorises melanoma into four main types: superficial spreading, LM, acral lentiginous and nodular, each with nine stages of invasion (1a – 4) plus an in situ stage (except for nodular melanoma). The authors assumed that invasive disease would progress at the same rate irrespective of the primary melanoma subtype, but the model allowed different progression probabilities from in situ disease. Patients with melanoma correctly identified as such (TPs) receive appropriate treatment according to their disease stage – they are then flagged by the model as having a history of melanoma and are at risk of stage-specific mortality. FN patients are discharged and returned to the natural history module in which they are at risk of disease progression and mortality. FPs incur the cost of referral and are discharged into the community. The authors assume that patients who are unaware they have melanoma suffer no impairment in quality of life. At the point of diagnosis, a disutility is assigned.

Mechanism of cost consequences

All optimal surveillance strategies were associated with incremental costs, which included the cost of the surveillance strategy itself and increased costs associated with the treatment of identified cases. The benefits of surveillance were primarily driven by health consequences not cost savings.

Mechanism of health consequences

In the model, early disease detection of disease prevents progression to later stages which are associated with greater health decrements and higher rates of mortality. Early detection via surveillance therefore generates health benefits by avoiding cases of late-stage diagnosis compared to when the disease is identified opportunistically.

The most cost-effective surveillance strategy (highest net benefit) was for those with a Williams score of 15–21 to be offered a one-off full-body skin examination, and for those with a score of 22 or more to be enrolled into a quinquennial monitoring programme, rising to annual recall for those with a risk score > 43. The overall ICER associated with the implementation of the surveillance strategies was £10,199.

Discussion

All three studies identified in the cost-effectiveness review employed similar model structures – a decision tree structure to represent the short-term outcomes associated with different diagnostic pathways, and Markov models to estimate long-term outcomes.

All three studies incorporated multiple indications, but none were so broad as the scope of the present assessment, and the extent to which different diagnoses were distinguished between prognostically and diagnostically varied. In Wilson (2013), the model tracked outcomes of malignant skin disease which comprised BCC, SCC and malignant melanoma. The model did not distinguish between melanoma or SCC and estimated outcomes based only on disease stage at diagnosis. Wilson (2018) included melanoma only but distinguished between the following subtypes: superficial spreading, LM, acral lentiginous and nodular. The model assumed that invasive disease would progress at the same rate irrespective of type but allowed the rate of progression from in situ disease to vary by subtype. The Edwards (2016) model accounted for melanoma, BCC and LM which were considered individually within three separate 'part' models.

The approach of the authors to the progression of undiagnosed cancers incorporated a range of data sources and fixed assumptions. Wilson (2013, 2018) used data from Losina (2007)[63] and Wilson (2018),[67] respectively, as sources for expert-elicited progression probabilities. The authors assumed a 10% annual probability of opportunistic detection of previous FNs. Wilson (2013, 2018) did not place limits on the progression of patients with FN test results, whereas the Edwards (2016) model assumed all undetected cancers would be opportunistically detected by the time they progress to stage 2. For the BCC model presented by Edwards *et al.*, no cancers remained undetected and so there was no progression possible in the Markov model component. In the Edwards melanoma model, an annual progression probability of 15.3% was applied, regardless of the current disease stage. A 35% annual probability of identification (if initially undetected) was applied based on the assumption that all unidentified melanomas should be identified by the time they reach stage 2, or 5 years after initial assessment. All studies assumed differential mortality rates according to disease stage for both identified and unidentified melanomas. BCC was assumed not to be associated with elevated mortality rates.

The mechanism by which costs and health outcomes are impacted in the three publications differs substantively. Value in the Edwards model was driven through the reduction of inappropriate procedures (most notably biopsy and excision and diagnostic biopsy) on health outcomes and costs. Biopsy and excision for melanoma were associated with a permanent disutility from scarring (for those with a lesion on their head or neck), temporary disutilities from anxiety while waiting for test results, and a disutility from the procedure itself. This model placed less value on diagnostic sensitivity, assuming that unidentified cancers have a utility equal to that of the general population until later diagnosis. In the BCC model, implicit in the structure is that 100% of BCC cases are always correctly identified as such and so there is no cost or health consequence from improved diagnosis.

This approach contrasts with that of the other two publications, whereby increased sensitivity drove value. In Wilson (2013), MoleMate increased average per patient testing costs, but improved patient health because of increased detection of cancers which were associated with improvements in long-term health outcomes. This is likely due to two structural differences: firstly, the model assumed a differential utility decrement by cancer stage independent of diagnosis, and secondly, any health decrement associated with procedures undertaken in secondary care (e.g. scarring from biopsy and excision) was not captured. However, Wilson (2018) differs from Wilson (2013), in that they assume undiagnosed melanomas only impact HRQoL after diagnosis (in line with Edwards) but did allow disease progression beyond that assumed in Edwards. Neither Wilson (2013) nor Wilson (2018) captured the health impact of scarring on patient health or other harms of procedures such as diagnostic biopsy, for example anxiety while waiting for results.

Chapter 5 Economic models submitted by skin analytics

Cost–utility model (Exeter Test Group/Skin Analytics)

Skin Analytics provided a preliminary report on a cost–utility model developed with Exeter Test Group during the latter part of the early value assessment (EVA) process. The executable model itself was not made available to the EAG for review. Due to the late provision of the company cost-effectiveness report, and the incomplete description of the analysis in the submitted documents, the EAG is unable to provide the usual level of scrutiny of a company cost-effectiveness model and does not accord with the template used in the assessment of company model used within the single technology process.

The decision problem considered in this analysis aligned with the scope of the EVA, that is triage of patients referred from primary care via the dermatology urgent skin cancer referral pathway. The model assesses two models of implementation of DERM in this setting: DERM with a second read, in which the images from DERM-negative patients are assessed by a consultant prior to discharge; and DERM without a second read, where DERM-negative patients are discharged without a further assessment. The model considered two comparators: face-to-face assessment and teledermatology.

Modelled population

The characteristics of the modelled population were based on NHS sources. It was assumed that 87.2% of patients screened had precancerous or benign lesions. Of the patients, 5.9% were assumed to have melanoma, SCC and rare skin cancers, and 6.9% had BCC. The model assumed that disease stage of melanoma at the point of diagnosis would also apply to SCC and other rare cancers. Evidence supporting this assumption was not presented in the provided report.

Model structure

The model structure was adapted from Wilson *et al.* (2013)[59] [described in *Wilson et al. (2013)*], comprising a decision tree with Markov models at each terminal node to link specific diagnostic outcomes with long-term costs and outcomes. The model described differs from that presented in Wilson *et al.* (2013), in that it explicitly models BCC as a separate diagnostic category to the high-risk cancers (i.e. melanoma, SCC, and rare cancers), reflecting the different prognosis and treatment of these indications. The model applies three diagnostic categories, each with a distinct diagnostic accuracy profile for each strategy, and associated treatment costs. These are:

- 'High-risk cancers', including melanoma, SCC and 'other high-risk cancers'
- BCC
- low-risk lesions (benign and precancerous).

The model adopted a lifetime time horizon (up to 100 years of age), with a 1-year cycle length in the Markov phase of the model. A half-cycle correction was applied. The model adopted an NHS and PSS perspective. Costs and benefits are discounted at 3.5% per annum.

There are possible four diagnostic pathways represented by decision trees. While the report does not contain a complete model schematic, it can be inferred from the provided description.

DERM without a second read

- DERM-positive patients are referred to a face-to-face dermatologist, and can then be diagnosed malignant (TP and FP) or benign (TN and FN).
- DERM-negative patients are discharged and enter the FN or TN Markov model.
- Patients who are ineligible for DERM assessment or whose DERM assessment is unsuccessful are referred directly to a face-to-face assessment.

DERM with a second read

- DERM-positive patients are referred to a face-to-face dermatologist, and can then be diagnosed malignant (TP and FP) or benign (TN and FN).
- DERM-negative patients undergo a virtual triage by a consultant dermatologist from which they can be discharged or referred to a face-to-face dermatologist who can diagnose malignant (TP and FP) or benign (FN and TN).
- Patients who are ineligible for DERM assessment or whose images are unsuccessful are referred directly to face-to-face assessment.

Face-to-face assessment

- Patients are assessed by a dermatologist and are either discharged (FN and TN) or referred for histological assessment (TP and FP).

Teledermatology

- Patients who are ineligible for teledermatology assessment are referred directly to face-to-face assessment.
- images are assessed remotely by a consultant, and patients are either discharged (FN and TN) or referred for a face-to-face assessment.

There are five Markov models used to represent the differing prognoses of patients by diagnostic outcome and indication beyond the terminal nodes of the decision tree:

High-risk cancer (melanoma, squamous cell carcinoma, rare cancers)

- TPs: Patients enter a health state corresponding to the stage of their disease at the point of diagnosis and treatment. The prognosis of patients with in situ or stage 1a cancer is equal to that of the general population for the remainder of the modelled time horizon. Later cancer stages are essentially modelled using a series of three tunnel states, wherein a patient is subject to an elevated mortality risk for the first 5 years which declines for the following 5 years, and returns to that of the general population thereafter.
- FNs: Patients enter a health state corresponding to the stage of their disease. Every year the patient can remain undetected and remain at the same stage, progress to a more advanced stage, or be opportunistically diagnosed and treated. The outcomes of these patients upon diagnosis are modelled in the same way as TPs.

Basal cell carcinoma

- TPs: Patients correctly diagnosed with BCC are treated and experience general population mortality risk. A small disutility is applied to some patients reflecting the impact of scarring on the head or neck upon HRQoL. This Markov model comprises two health states – alive and dead.
- FNs: Patients with undetected BCC have a 20% annual probability of being opportunistically diagnosed and treated. There is no risk of progression associated with having undetected BCC, nor is there any impact on HRQoL. A proportion of patients whose BCC is detected and treated experience a small utility decrement as above. A four-state model is described – undiagnosed BCC, opportunistic detection and treatment, treated, and dead. It is unclear what purpose the separate health state representing detection and treatment serves.

Mechanism of cost consequences

Costs relating to diagnosis include face-to-face assessment, biopsy/excision and multidisciplinary team meetings (MDT). Costs associated with each of the diagnostic processes are replicated in *Table 17* for comparison.

The costs of further follow-up and treatment following a referral to a face-to-face appointment differ by the modelled indication, with further costs associated with more advanced stage at diagnosis. On the melanoma pathway, initial

TABLE 17 Exeter model diagnostic costs

Parameter	Cost	Source
Photo clinic appointment (medical photography for DERM, teledermatology)	£14.30	Skin Analytics – 45 minutes of Band 3 time.
Teledermatology review	£25.00	Skin Analytics – 10-minute slot (2020 PSSRU cost – hospital-based consultant, medical) plus 15 minutes Band 3 administration time.
Teledermatology system price per image	£7.00	Skin Analytics – list price of Cinapsis, Dermicus.
DERM second read	£17.00	Skin Analytics consultant time[a]
DERM assessment price per image	£38.20	Skin Analytics list price[a]
Face-to-face dermatologist appointment	£142.00	WF01B, 2023–5 NHS Payment Scheme. NHS England

a Unit prices provided to the EAG differed from those applied in the Exeter model.

biopsy/excision and SLNB had a unit cost of £507, and was applied in addition to a MDT (£123) to all patients who were not discharged following their face-to-face assessment, as was a vitamin D test at a cost of £178. The source of the £178 cost of a vitamin D test was unclear, and appeared to be substantially higher than other literature sources,[68] which tend to inflate from a figure of £16.50 based on previous NICE guidance.[69] The costs associated with biopsy and treatment appear high relative to the studies discussed in *Chapter 4*, and were not consistently based on NHS Reference Costs/Personal Social Services Research Unit (PSSRU) costs. This punishes diagnostic strategies with lower specificity and may inflate the potential cost savings associated with higher-specificity strategies.

Frequency of clinical follow-up was determined by disease stage at diagnosis, with a unit cost of £77 for each visit. Patients with stage 1b or higher disease were assumed to require frequent ongoing follow-up imaging (e.g. MRI, CT, ultrasound). Terminal care costs of £15,531 were applied to patients who died with stage 1b or higher disease. Costs of further investigations were applied to melanoma patients with stage 2 or higher disease at diagnosis, including histology testing and further medical imaging. Further surgical and systemic treatment was included for patients with stage 3 or 4 disease.

Treatment costs associated with BCC were calculated using a weighted cost of £556.82 per patient, comprising various alternative treatment strategies from McFerran *et al.*,[70] with costs inflated to 2024 values using the EPPI-Centre cost converter. It was noted that phototherapy, which contributed £38.84 to the weighted cost, is not used for treatment of BCC on the NHS. As there are no health consequences of a missed BCC diagnosis, the only meaningful outcome of a correct BCC diagnosis is incurring this cost. This counterintuitively means technologies with poorer sensitivity generate value by having a lower sensitivity.

The primary mechanism of cost savings in the model was the avoidance of face-to-face assessments and biopsy. The specificity of a face-to-face assessment with a consultant was 79.7%, resulting in a proportion of patients receiving costly biopsy unnecessarily if they are not discharged using teledermatology or DERM. A diagnostic pathway with higher sensitivity also avoids of missed cases which have the potential to develop into advanced disease, with substantially increased treatment costs. An important assumption with the model is that the sensitivity of face-to-face assessments is increased to 99% following triage with either DERM or teledermatology. This means that fewer cancers are missed in model pathways including an additional triage step. The plausibility of this assumption is not clear and may not reflect real-world practice given the low assumed specificity of DERM and teledermatological assessment.

The assumption that a relatively high number of unnecessary biopsies resulting from face-to-face assessments and the improvement in sensitivity of face-to-face assessment following triage are likely key drivers of benefit in the model and as such the associated diagnostic parameters are central to the value proposition.

DERM (with or without second read) versus teledermatology

Results of the company's model suggest DERM either with or without second read generates costs savings relative to teledermatology (*Table 18*). A simple comparison of first-line assessment costs inclusive of DERM however suggests that both DERM strategies are more costly than teledermatology (£72 vs. £57 average cost per patient). These higher costs associated with both DERM strategies are driven fewer by patients being eligible for assessment by DERM than teledermatology (81% vs. 90%). This results in more patients receiving more expensive face-to-face assessments.

The first-line incremental costs associated with both DERM strategies are, however, offset by improved specificity relative to teledermatology which results in higher effective discharge rates. Effective discharge rates are 36.9% for DERM without second read, 15.7% for DERM with a second read and 30.9% for teledermatology. The higher discharge rates associated with DERM without a second read generates cost savings as fewer face-to-face appointments are required and fewer biopsies conducted, while DERM with a second read generates cost savings through the avoidance of missed diagnoses. The specificity of teledermatology was assumed to be 35% based on an average observed across UK DERM pilot pathways and other real-world data sources; this compares with a specificity of 42% assumed for DERM without second read based on performance across secondary care pilot sites. The specificity of DERM with a second read can be estimated using the DERM specificity of 42% and the specificity of the second read of 60%. The assumed specificity of teledermatology however appears low compared with published sources. Teledermatology specificity was reported as 84.3% in the Cochrane review referenced in the preliminary Exeter report which may indicate the assumed specificity is lower than in practice.[58]

The total average costs of the peri-referral pathway (i.e. between referral and initial secondary care consultation) are approximately £146 for teledermatology, £118 for DERM without a second read and £172 for DERM with a second read (assuming there is no additional step which can overrule Skin Analytics dermatologists). That is, the reduction in face-to-face dermatology referrals achieved by DERM used autonomously generates cost savings per patient referred from primary care. DERM with a second read may be the costliest approach, but may be associated with non-cash-releasing benefits related to outsourcing of teledermatology review to Skin Analytics consultants. Note that using the modelled assumptions, the inclusion of teledermatology in this pathway is more costly than simply referring all patients to a face-to-face assessment. However, if the Cochrane diagnostic accuracy values are applied for teledermatology, DERM strategies become more costly than teledermatology. Teledermatology also becomes cost saving versus the traditional pathway.

DERM without second read versus face-to-face assessment

Compared to face-to-face assessment, results of the company's model suggest both DERM strategies incur lower costs (see *Table 18*). As above, this is driven by lower costs associated with unnecessary referrals and inappropriate biopsies. In the BCC population, additional cost savings are also generated due to the lower sensitivity for DERM compared to face-to-face assessment (90% vs. 95%). This occurs because of the assumption that missed cases of BCC have no consequences in terms of costs.

DERM with a second read versus DERM without a second read

DERM with a second read is associated with incremental costs compared with DERM without a second read (see *Table 18*). The cost difference between the two strategies is in part driven by the addition of the second read which increases costs in the DERM with a second read strategy. However, the incremental costs associated with DERM with a second read are partially offset by the lower rate of missed diagnoses.

Mechanism of health consequences

The annual risk of progression with melanoma, SCC and other rare cancers is derived from Wilson (2013) as described in *Wilson et al. (2013)*. As in Wilson (2013), the distribution of disease stage at diagnosis was implicitly assumed not to be affected by the underlying diagnosis. Health outcomes in the model were a consequence of both treatment and underlying disease which were applied as either a utility decrement or mortality modifier. Specific assumptions were applied for the BCC population, and melanoma, SCC and other rare cancer populations.

TABLE 18 Results of cost-effectiveness of DERM

			Inc. (vs. usual care)		Inc. (vs. teledermatology)		
Strategy	Cost (£)	QALY	Cost (£)	QALY	Cost (£)	QALY	ICER
DERM + second read	465.84	11.1925	−31.14	+ 0.0077	−6.27	+ 0.0039	24,655.23
DERM	445.09	11.1917	−51.89	+ 0.0069	−27.02	+ 0.0031	–
Teledermatology	472.11	11.1886	−24.87	+ 0.0038	–	–	Strictly dominated
Usual care – baseline	496.98	11.1848	N/A				Strictly dominated

Inc, incremental; discrepancies due to rounding.

The model assumes BCC does not progress if not diagnosed, and undiagnosed BCC has no further health consequences. There is a 20% annual probability of opportunistic detection of undiagnosed BCC in the Markov phase of the model, with all patients assumed to achieve general population health outcomes following treatment, with no risk of recurrence. The treatment of BCC is associated with costs and causes a permanent disutility in 15% of the 58.9% of patients with scarring on their head or neck. There are therefore negative outcomes in terms of both costs and QALYs associated with correctly diagnosing a case of BCC, meaning that in the model it appears that more benefit is yielded by missing a given case of melanoma than by detecting it. The assumptions underpinning the modelling of BCC may not be clinically plausible. This means that a diagnostic strategy with a higher sensitivity for BCC is likely to be less cost-effective than one that misses BCC more often and postpones diagnosis. The sensitivity of DERM for BCC is 90%, lower than the 95% assumed for face-to-face assessment and teledermatology. This is likely to lead to increased costs and reduced QALYs for the latter two strategies, despite achieving a better diagnostic outcome. The clinical plausibility of this is unclear and runs counter to expectations that improving diagnostic outcomes improves health outcomes.

Melanoma, SCC and other rare cancers were assumed be associated with lower quality of life dependent on disease stage. Utilities were based on a 2014 study by Tromme and colleagues,[71] which used the EuroQol-5 Dimensions, five-level version (EQ-5D-5L) questionnaire in a population with melanoma to derive utility weights according to disease stage and whether patients were actively undergoing treatment or were in remission. These utilities were adjusted to the mean age of the modelled population using Sullivan *et al.*[72] Utilities reflecting the impact of treatment were applied as a one-off disutility in the first year of treatment. Patients with stage 1b or 2 disease were assumed to return to an age-adjusted general population-equivalent utility 2 months after treatment. Those with stage 3 or 4 cancer at diagnosis have a reduced quality of life for the remainder of their lifetime. These utilities were based on small samples and are not necessarily logically consistent – for example a patient who has recovered from stage 3 cancer has a utility of 0.701, but 0.797 for a patient who has recovered from stage 4 disease. A single utility representing recovered patients with stage 2 or above cancer may have been more appropriate.

Melanoma, SCC and rare cancers were assumed to occur on the head or neck in 40.2% of patients, 15% of whom experienced scarring following treatment, and a permanent disutility, the magnitude of which was not reported. It was also assumed that patients would experience a disutility of −0.505 for the period over which they are waiting for a result after GP referral to capture the impact of anxiety and psychological distress.

Mortality rates for high-risk cancers were taken from Edwards *et al.* (2016), ultimately based on Balch *et al.* (2009).[62] As described above, patients with high-risk cancers at stage 1b or higher had an increased risk of cancer-related mortality for the first 10 years following diagnosis and treatment, after which time they have the same mortality risk as a healthy member of the general population (using Office for National Statistics data). An annual probability of death was calculated from 5-year survival data. The mortality risk and thus potential QALY loss associated with undiagnosed high-risk cancers is a potentially significant driver of benefit generated by more sensitive treatment strategies.

The most effective strategy was DERM with a second read, generating 11.1925 QALYs at a cost of £465.84. The ICER for DERM with a second read compared to DERM alone was £24,655 per additional QALY gained. Both DERM

strategies were predicted to be less costly and more effective than teledermatology and usual care. Teledermatology was less costly and more effective than usual care. The observed shortfall in QALYs accrued on usual care is likely to be driven by the assumption that the sensitivity of a face-to-face assessment is significantly improved in patients who have undergone previous DERM triage. This structurally confers health benefits onto strategies employing an intermediary step between primary and secondary care and may not be reflective of real patient outcomes.

Summary of critique

The submitted model represents the most recent and complete attempt to represent the NHS urgent skin cancer referral pathway but is subject to a number of weaknesses which may mean it does not appropriately characterise the main drivers of value in this pathway.

As the most common form of skin cancer, the consequences of diagnosis and treatment of BCC is an influential driver of cost accrual in the model. The model essentially punishes correct BCC diagnoses, as excision is associated with accrual of costs and a QALY decrement. This introduces a disincentive to improve diagnostic sensitivity, and indeed DERM is less sensitive for BCC than teledermatology or face-to-face assessment. This may reduce QALYs and increase costs on the two comparator pathways, and this is somewhat concealed in the cohort structure.

The model also structurally imposes a 99% sensitivity for face-to-face assessment following triage, without evidential support. This means that the simple introduction of a triage step (i.e. DERM, teledermatology) prior to consultation with a dermatologist reduces missed diagnoses and avoids the associated cost and health implications. This assumption may not reflect the respective real-world holistic sensitivity of these pathways but would invariably result in better cost-effectiveness estimates for DERM and teledermatology.

The costs associated with biopsy and treatment of high-risk cancers drives cost accrual in triage strategies with lower specificity, as more patients will undergo unnecessary and expensive diagnostic biopsy, in addition to the costs of a face-to-face consultation. Significant value is therefore generated by triage strategies with higher specificity. The magnitude of uncertainty surrounding the specificity of teledermatology is vital to understanding the potential for cost-effective use of DERM in this model structure. Given higher rate of ineligibility for assessment with DERM versus AI, the true discharge effective rates are closer than implied by simple comparison of the respective diagnostic accuracy statistics of each technology, as a higher proportion of patients on the DERM pathway proceed immediately to face-to-face assessment. The specificity of teledermatology reported in published sources is substantially higher than that observed in the pilot sites (which were largely not set up for teledermatology services). It is therefore highly plausible that in the presented model structure, teledermatology would be more cost-effective than a pathway incorporating DERM.

East Midlands academic health science network (2023)

The authors report an evaluation of a pilot of a Skin Analytics AI-powered teledermatology (i.e. DERM with a second read) for the skin cancer 2WW pathway at UHL sites in March 2022. The evaluation uses a mixed-methods framework, combining patient and staff feedback surveys with quantitative data collected as part of the pilot. The existing pre-intervention pathway prior to the implementation of the pilot involved patients referred from primary care on the urgent skin cancer referral pathway (NG12).

(confidential information has been removed)

National Health Service England artificial intelligence in health and care economic evaluations

The contents of this report remain confidential at the time of submission.

(confidential information has been removed)

Chapter 6 Model conceptualisation and identification of evidence gaps

Model conceptualisation

The following sections describe a conceptual model based primarily on a synthesis of the economic analyses identified in the economic review, and evidence submitted by skin analytics. While cost–utility models have recently been built to address the present decision problem [see *Cost–utility model (Exeter Test Group/Skin Analytics)* and *National Health Service England artificial intelligence in health and care economic evaluations*], the EAG considers currently available evidence insufficient to answer the issue of the potential cost-effectiveness of AI technologies for detecting benign lesions following referral from primary care. This section expands upon the EAG reasoning for this conclusion and details key data necessary to fully address the decision problem.

Decision problem

The outlined conceptual model considers both use cases for AI technologies proposed for this evaluation, that is autonomous AI triage following referral from primary care, and AI triage with a second read 'safety net' prior to discharge following referral from primary care. The use case of AI technologies to be assessed is the identification of benign lesions and the direction of discharge prior to contact with secondary care. These decisions could be made autonomously by AI or following dermatologist review (second read). A holistic modelling approach to the diagnostic accuracy of these technologies is necessary in order to assess the potential value to the NHS.

The modelled population should include all patients referred on the urgent referral skin cancer pathway from primary care. The prevalence of cancer subtypes should be sourced from appropriate and recent UK national sources. Staging of disease at the point of entry into the model should be based on UK data if available. If there are differences by stage at presentation according to indication, this should also be reflected. See *Prevalence of disease and distribution by disease stage* for further discussion.

To reflect current service provision, two alternative comparator diagnostic pathways are considered: the teledermatology model and the conventional model of referral to face-to-face assessment model. Current provision varies across the English NHS, with no nationally standardised alternative model to the usual referral pathway. Fully reflecting regional variations may therefore require additional comparator pathways to be modelled. Modelled outcomes should include diagnostic outcomes, that is TP, FP, FN, TNs; costs; and QALYs. Disaggregation of outcomes by indication should be possible.

The proposed model should be built in full alignment with the NICE Reference Case and should adopt an NHS and PSS perspective. Costs and benefits should be discounted at 3.5% per annum. A lifetime time horizon should be applied on the basis of the age of the modelled population.

Proposed model structure

In line with previous economic analyses, the EAG proposes a cohort model in which all patients enter a common decision tree structure, regardless of underlying indication. Different Markov models would then be used to reflect differences in long-term costs and outcomes as a result of diagnostic outcome. Differences in model inputs relating to costs and health outcomes would allow the model to be parameterised to address specific indications, for example, melanoma, SCC and BCC.

The level of granularity possible in the model will be data dependent. However, as in previous models, it will be important to differentiate melanoma and other high-risk cancers from BCC, as the costs and consequences of diagnosis and misdiagnosis can be radically different.

The proposed model applies three broad diagnostic categories, with each having distinct long-term consequences, which are represented by different Markov models. The capacity of the conceptual model to account for specific diagnoses within these categories is dependent upon the availability of data to inform specific diagnostic accuracy and natural history parameters. The diagnostic categories are as follows, based on the groupings proposed in the Exeter Test Group/Skin Analytics model described in *Cost–utility model (Exeter Test Group/Skin Analytics)*:

- 'high-risk cancers', including melanoma, SCC and other rare high-risk cancers
- BCC
- low-risk lesions (benign and precancerous)

A key concern regarding the use of AI technologies for the diagnosis of skin cancers is the identification of rarer indications. Given that these technologies may have limited experience of rare cancers, there remains uncertainty as to whether their high sensitivity to melanoma and SCC is maintained across these rarer indications. Treating them as a single diagnostic category in terms of diagnostic accuracy, stage at diagnosis, rate of progression and impact upon mortality may therefore be subject to uncertainty. Where possible, sensitivity analysis should be undertaken in which rare cancers are categorised separately, and alternative sources of diagnostic and prognostic data are used to parameterise this sub-population in the model.

Decision trees

Patients enter the decision tree following an urgent referral from primary care, according to the chosen approach to AI implementation (i.e. with or without a second read), and to each comparator (face-to-face and teledermatology). The decision tree directs patients through a series of tests and clinical decision points, determining their accumulation of any costs associated with testing and appointments, and their ultimate diagnostic classification, that is TP, FP, FN and TN.

The comparator combinations of AI with and without second read with teledermatology and the direct referral pathways generate four diagnostic pathways, illustrated in the simplified decision tree schematics in *Figure 13*. While the schematic depicts individual head-to-head comparisons, the proposed model would generate results in a fully incremental format.

Only a proportion of patients are eligible for AI and teledermatology assessment. This is represented in the decision trees by a third initial branch. Different proportions of patients are eligible for each of these technologies, with current eligibility criteria more restrictive for the use of AI triage technologies than for teledermatology. This may have a significant impact on the costs and outcomes achieved on each pathway. Patients ineligible for AI/teledermatology are routed straight to face-to-face assessment, with a proportion whose ineligibility was not assessed prospectively, and were thus subject to additional costs associated with unsuccessful photography/an indeterminate AI result. The diagnostic accuracy of a consultant dermatologist may also differ for patients whose lesions are ineligible for each technology. Where possible, this should be accounted for in the economic analysis or otherwise explored in relevant sensitivity analysis.

All decision trees determine the proportion of patients with TP, FP, TN and FN under each diagnostic strategy, with long-term outcomes for each determined by each of the respective Markov models depicted in *Figure 14*. At the terminal nodes representing TP and FP, patients are assumed to undergo biopsy and/or treatment appropriate to their stage at diagnosis.

Markov models

Patients correctly identified at the terminal nodes of the decision tree enter Markov model A (see *Figure 14*); these patients have ongoing mortality and HRQoL implications following treatment depending on the disease stage at the point of diagnosis. This Markov model comprises a Markov state (not depicted) for every possible disease stage at the point of diagnosis and treatment, and a series of tunnel states reflecting mortality risks post treatment. The use of tunnel states permits declining risks of post-treatment mortality to be modelled (per Edwards *et al.*) and may be applied as long as clinically appropriate, at which point they return to a general population risk of mortality (see *Clinical input parameters*).

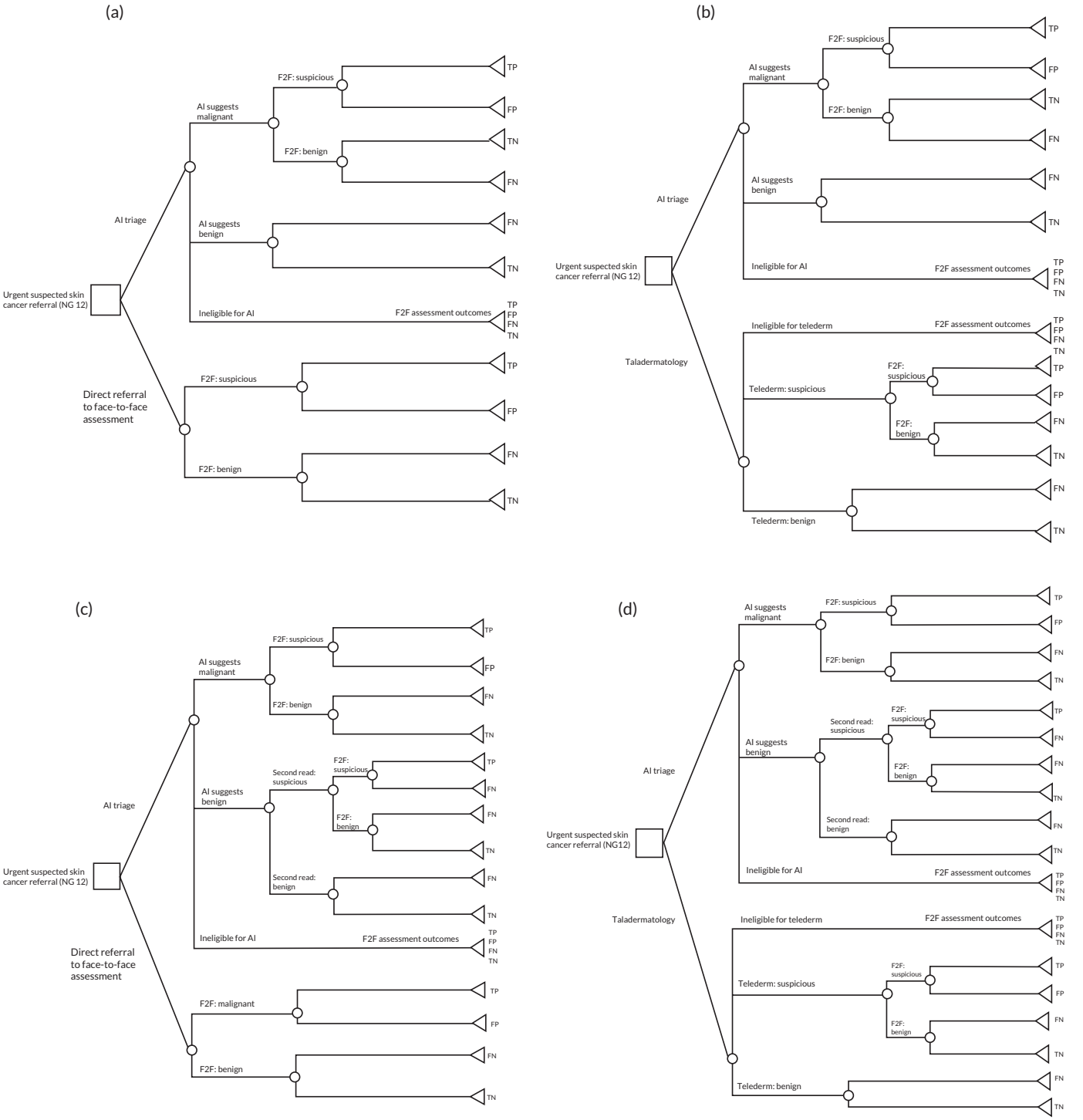


FIGURE 13 Proposed model structure: decision tree schematic (a) AI without second read vs. referral to face-to-face assessment; (b) AI without second read vs. teledermatology; (c) AI with second read vs. referral to face-to-face assessment; (d) AI with second read. F2F, face-to-face.

Patients who reach a FN terminal node enter Markov model B (see *Figure 14*), and have a stage-specific risk of progression, mortality and opportunistic detection. Patients with a TN or FP diagnosis enter Markov model C (see *Figure 14*). These patients have general population mortality and HRQoL outcomes. Utility decrements may be applied to account for the long-term impact of scarring due to inappropriate biopsy on the head and neck in FP patients.

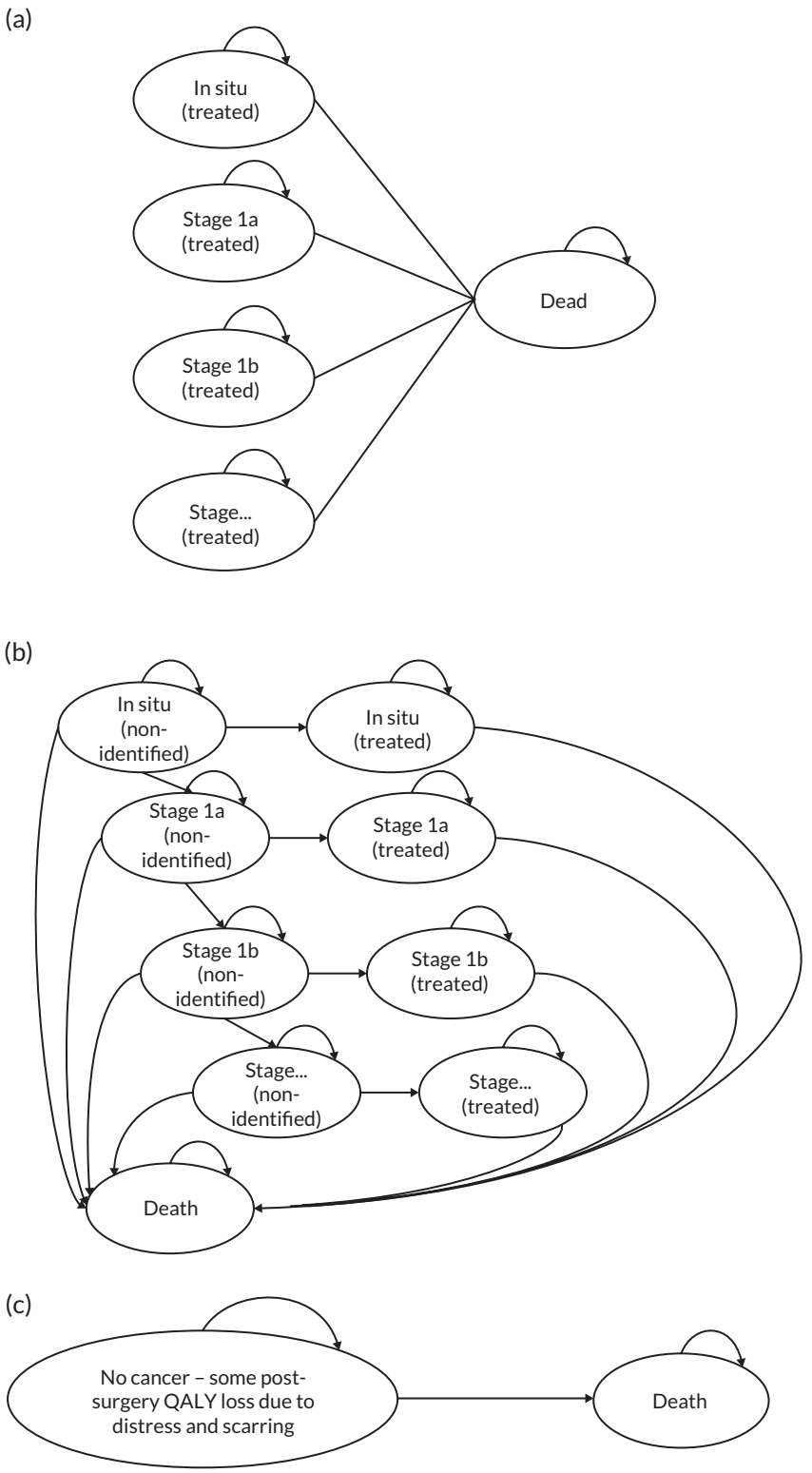


FIGURE 14 Markov model components: (a) TPs; (b) FNs; (c) TNs and FPs.

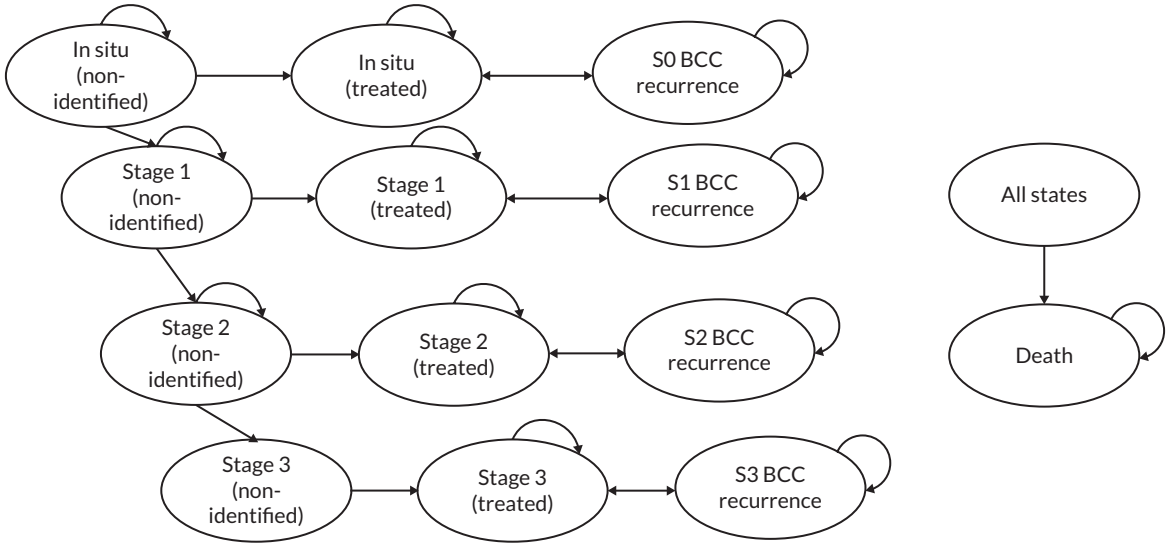


FIGURE 15 Markov model component capturing outcomes of BCC FNs.

The EAG note that existing modelling approaches assume no adverse implications of a missed BCC in terms of cost or health outcomes. The clinical plausibility of this approach is unclear and, in the context of a cost–utility model, this essentially rewards strategies with lower sensitivity, as the costs associated with BCC are avoided or postponed (and are subject to more discounting). The EAG therefore proposes an alternative approach to capturing the long-term impact of missed diagnoses of BCC. Under this approach, TPs, TNs and FPs follow the same Markov model structures as the high-risk cancers, but in the proposed model, FNs for BCC follow the structure presented in *Figure 15*. While BCC is associated with a low risk of spread and progression to metastatic forms of the disease, if left undiagnosed, some subtypes can be invasive and can cause local destruction of deeper tissues such as muscle and bone,[73] which can be particularly impactful for lesions located on the head and neck. Untreated BCC may become more advanced over time and can be prone to higher rates of recurrence.[74–76] While recurrence of BCC remains manageable, it is associated with additional treatment costs.[12] The Markov structure in *Figure 15* therefore intends to capture the slow development of non-identified BCC and its opportunistic detection. Following detection and treatment, patients are then subject to a stage-specific recurrence rate. It is assumed that recurrent BCC is immediately detected and treated (with an associated cost), and patients return to the stage-specific 'treated' health state, with an ongoing risk of further recurrence. This assumes that recurrence of BCC does not have a modifying effect on the probability of future recurrences. Mortality is possible from any health state. Ideally, stage-specific treatment costs and disutilities would be modelled to allow differences in treatment costs arising from differences in the complexity of surgical intervention and reconstruction, and the potential impact upon HRQoL to be accounted for in the model.[77] The modelling of BCC in this way will be dependent on the availability of data to inform progression and stage-specific recurrence rates.

This model structure is in broad alignment with the analyses described in *Chapter 4*, including that built by the Exeter Test Group, which in itself was adapted from Wilson *et al.* (2013)[59] and Edwards *et al.* (2016),[61] with the addition of a Markov component to capture the long-term outcomes of a missed case of BCC. A model of this design captures the differential in core costs and consequences of alternative diagnostic tests which impact the routing of patients through a diagnostic pathway. This includes the financial consequences of appropriately discharging patients with benign lesions, and avoiding unnecessary resource-intensive face-to-face consultations, but also the impact of missed diagnoses on cost and health outcomes.

An important omission from the proposed model structure is the ability to capture non-cash resource benefits. Doing so would require a more complex approach which estimates the effect of the technologies upon downstream dermatologist capacity, and the impact of its deployment in this and other populations upon health outcomes.

Clinical input parameters

Prevalence of disease and distribution by disease stage

Estimates of prevalence are required for the proposed model and should be based on the population described in the decision problem, that is patients referred on the urgent referral pathway from primary care. A systematic review should be conducted to identify the prevalence of each disease type considered in the model for the UK urgent referral population or identify sources of NHS data to inform this parameter. All three studies identified in the cost-effectiveness review in *Summary of identified evidence* relate to different patient populations and are, therefore, not relevant to the current decision problem. The Exeter model used post-market surveillance data from Skin Analytics to obtain estimates of prevalence but noted this was a placeholder with a preference for acquiring national data in the future. Sensitivity analysis should be conducted on plausible estimates of prevalence to represent uncertainty or regional variation in prevalence estimates.

Also required for the proposed model is the distribution of each disease by stage at identification. Data should be obtained based on stage at presentation for lesions examined on the urgent referral pathway across each disease type considered in this model. Distribution of disease by stage will impact cost and health outcomes estimated by the model. If presentation is typically at later stages of disease, there will be reduced scope to generate benefits via early detection and vice versa.

Diagnostic accuracy data

To inform a future cost-effectiveness model, data on diagnostic accuracy are required for all relevant diagnostic strategies These data should ideally be obtained for each indication considered in the model, as diagnostic accuracy may differ by condition.

The use case for AI technologies in the proposed model involves the identification of benign lesions to allow patients to be discharged following referral, but prior to face-to-face assessment. The key statistic to estimate the capacity of a test to correctly identify TN cases is the specificity associated with this pathway. Discharge of patients with benign lesions reduces the cost and health implications associated with unnecessary investigations. While diagnostic accuracy may be framed with regard to its sensitivity to benign lesions at this point in the diagnostic pathway, it is helpful to refer to sensitivity and specificity for detection of malignancy, for consistency with the intent of subsequent/comparative face-to-face assessment.

The value implications of differing diagnostic performance across the comparators under consideration will depend on the following assumptions: follow-up costs for patients after a positive test (e.g. cost of biopsy and excision), the health consequence of treatment itself (e.g. scarring due to excision, anxiety while waiting for biopsy results), Markov state stage-dependent treatment costs/health decrements, and assumptions regarding the progression of patients. The net effect of this (along with the impact of discounting) will determine how diagnostic accuracy drives costs and outcomes.

In the case of the present pathway, this relationship is somewhat complicated by the application of sequential tests, for example, the use of a second read following AI assessment. For two sequential tests with imperfect accuracy, independence between tests would imply that overall specificity of a pathway would decrease. This may not be reflective of actual practice and would punish pathways with more steps. It is unlikely that test accuracy is fully independent between steps, a lesion deemed malignant by AI may be more likely to be deemed malignant by a read by a dermatologist and so it may be inappropriate for subsequent accuracy values to be applied to one other. Equally, assuming interdependence of two sequential tests may also not be completely appropriate, especially where testing is subjective.

Given these complexities and the number of diagnostic decision points in the decision trees described in *Proposed model structure*, care must be taken that diagnostic accuracy values are not simply pieced together from different sources to estimate whole-pathway sensitivity and specificity. In order to understand the resource use implications of post-referral use of AI or teledermatology, data on the sensitivity and specificity of both the whole pathway and its constituent components must be collected. These data should ideally be generated comparatively on the same clinical population

(i.e. having undergone the same pre-screening) in the same conditions. In the case of teledermatology, it is important to ensure the intention is the same as AI, that is with the express intention of identifying and ruling out benign lesions (as opposed to triage/prioritisation of all lesions); otherwise estimates of specificity are not comparable with the use case of AI technologies in this space.

The clinical evidence supplied in support of the DERM and MoleAnalyzer technologies is described and synthesised in *Chapter 3*. This evidence is largely derived from pilot studies. The EAG consider that further development of the evidence base is required to inform a future cost-effectiveness model. To inform a future cost-effectiveness model, data should be based on studies with the following characteristics:

- **Setting:** UK post referral (before secondary care investigations)
- **Intervention:** AI technologies (with and without human confirmatory read)
- **Comparators**: Face-to-face and teledermatology with intent to exclude benign lesions
- **Outcome**: Diagnostic accuracy (sensitivity and specificity) of individual component tests and the overall pathway

Progression and opportunistic detection parameters

Parameters describing the ongoing probability of undiagnosed progressing or being opportunistically detected are necessary to inform the transition between states in the long-term Markov components of the model, representing the natural history of skin cancer in patients with a FN diagnostic outcome. These parameters are likely to be influential in determining the mechanism of benefit in a future cost-effectiveness model. A model which applies more rapid progression or a lower chance of subsequent detection will impose greater value on improved sensitivity of a diagnostic pathway.

Data on the progression of unidentified skin cancers to inform progression probabilities appear limited. The approach taken by cost–utility models reported in the cost-effectiveness review and the Exeter Test Group model relied on expert-elicited progression probabilities, including Losina *et al.* (2007)[63] and Wilson *et al.* (2017).[67] In the absence of more recently published alternative data sources, the proposed model may need to adopt transition probabilities based upon these studies.

Two contrasting approaches were taken by identified studies to structural assumptions regarding the opportunistic detection of FNs of patients – Edwards *et al.* assumed that FNs must be identified 5 years following initial assessment or upon progression to stage 2. However, this latter restriction may be reflective of the typically earlier staging at presentation of the population considered in the Edwards study. The other identified studies placed no structural limitations on patient progression. However, it may be appropriate to impose a time-based limit on the period over which a malignant lesion remains undetected, to avoid implausibly long durations of patients living with progressed disease.

Consideration should also be paid to whether it is appropriate to use common progression and identification parameters across multiple diseases in a future cost-effectiveness model or whether separate values should be used if disease processes are sufficiently different.

In the conceptual model, Markov state-specific mortality rates are likely appropriate, that is the mortality rate for a modelled patient is dependent upon their disease stage at presentation (if correctly identified and treated), or a patient's current disease stage (where undetected). Within previous models, mortality risks have increased with disease stage, with mortality risks converging with that of the general population following successful treatment. A permanent increase to mortality rates may also be appropriate in patients who experienced more aggressive treatment at later stages of disease. Mortality in patients with benign lesions can be reflected by general population rates.

A consensus across the models considered in this report is that people with stage 1a melanoma have a risk of mortality close to that of the general population and so no additional risk was assumed. Regarding patients with disease initially identified (or subsequently identified following initial non-identification), an assumption should be made regarding the duration of elevated mortality following treatment, reflecting the residual risk associated with the disease. Edwards

assumed that following identification, patients would experience elevated mortality for 10 years after which their risk of mortality would return to that of the general population – 5 years at a higher rate and 5 years at a lower rate. An alternative approach applied by Wilson (2013)[59] (parameters obtained from Balch[62]) and Wilson (2018)[60] (parameters obtained from a previous NICE appraisal[78]) calculated log-odds ratios for each stage and applied them to general population mortality rates.

All models identified in this report applied differential rates of mortality according to disease stage, reflecting differences in prognoses. Given the large sample size of the Balch *et al.* reference (n = 30,946),[62] the EAG consider this a suitable source for populating a future cost-effectiveness analysis but may require reanalysis reflecting more recent techniques. Given the age of this study, further searches should be undertaken to identify more recent estimates of mortality in this population (or other secondary analyses of Balch *et al.*), although this is unlikely to be an important driver of model outcomes.

Health-related quality of life

Health-related quality of life is represented in the model through the application of health utilities. Previous models have applied utilities to represent dimensions including:

- utility decrements representing the disutility associated with diagnostic and treatment procedures (e.g. anxiety associated with the wait for biopsy results, scarring as a result of biopsy/excision)
- health-state utilities representing diagnostic status and disease stage (or presence of disease).

The utilities reported in Tromme *et al.* (2014) (adapted for use in Edwards *et al.* 2016) to represent health-state utilities specific to disease stage and treatment status may be adequate to represent the impact of skin cancer and its treatment upon HRQoL. However, these data should be reanalysed – perhaps by pooling EuroQol-5 Dimensions scores for patients with stage 3 and 4 melanoma to avoid logical inconsistencies arising from the small sample size of patients with stage 4 disease in remission. EQ-5D-5L summary scores should be cross-walked to EuroQol-5 Dimensions, three-level version (EQ-5D-3L) using the Hernández Alava mapping algorithm,[79] and should be adjusted for age and sex balance using the EEPRU value set established by the NICE Decision Support Unit.[80] Given the age of the Tromme *et al.* data set and its aforementioned limitations, a systematic review of HRQoL studies should be undertaken to identify any more recently published data sources. Where alternative values are identified, these should be mapped to EQ-5D-3L for consistency with the NICE Reference Case. Utilities should also be adjusted for age and sex balance using the EEPRU value set.[80]

An anxiety-related disutility in line with that used in Edwards *et al.* (2016) could also be applied for the period over which patients await a final diagnostic result following GP referral. The impact of AI technologies on this interval should be identified from existing evidence sources (such as the DERM pilot studies), and its effects explored in sensitivity analysis, reflecting the potential for lengthened waiting times as seen in the UHL pilot. Previous models have also applied utility decrements with scarring on the head or neck following treatment. A disutility of an appropriate magnitude should be identified from literature sources.

A systematic review of HRQoL values should also seek to identify disutilities associated with BCC treatment. In the absence of disease-specific HRQoL data, it may also be appropriate to apply a one-off disutility equivalent to that applied for the treatment of melanoma in situ, which is typically managed using excision in a similar manner to BCC.

Cost and resource use parameters

Relevant costs in the proposed cost-effectiveness model include those related to diagnosis (e.g. the cost of the technologies, comparators and clinical appointments), treatment and investigation-related costs (e.g. biopsy, excision, imaging), and long-term state-dependent management costs based on treatment and disease stage. Those related to the technologies themselves should be based on information provided by the companies and any implementation

costs likely to be incurred should be considered in the model (e.g. staff training, establishing new medical photography infrastructure).

The costs and resource use assumptions applied are likely to be a key driver of the value of technologies in this space. There is a degree of control over the valuation of each diagnostic accuracy parameter in models of diagnostic technologies, that is, greater value can be ascribed to improving sensitivity by emphasising the costs of a missed diagnosis on the cost of delayed treatment. Equally, a technology which prioritises specificity may be made to generate more apparent value through increasing specificity and thus avoid unnecessary further investigations. The proposed model should aim for consistency in sources of cost data with precedent in NICE appraisals to ensure costs to the NHS and PSS are represented as accurately as possible.

Any costs associated with NHS procedures should be based on the latest national sources in alignment with the NICE methods guide for consistency with previous (and future) NICE decisions. These sources include the Unit Costs of Health and Social Care,[81] NHS Reference Costs to and the NHS Drug Tariff.[82] Any costs without appropriate NHS reference costs (e.g. long-term state-dependent costs) should be based on a synthesis of the available evidence with costs inflated to the current cost year. The application of unit costs in the model should be made based on treatment guidelines provided by NICE and authoritative clinical guidelines.

Technology costs

Costs of the relevant technologies were provided by Skin Analytics and Moleanalyzer as part of the assessment process. Available information for each company is described below.

Skin Analytics DERM

Skin Analytics provided information regarding pricing for DERM. Pricing information is provided according to two options on a per-year basis: (a) per 10,000 catchment population covered; and (b) per 2WW referral. It is unclear whether both pricing models are available to trusts, or if the cost per 2WW referral is for indicative purposes only, as annual payments are stated to be made upfront.

The pricing options are presented in *Table 19*. The total cost per 2WW image processed is £30.00, with an additional optional unit cost of £8.20 per referral to store images in order to allow remote review by trust clinical staff.

The company state that pricing is inclusive of training and data storage costs. The proposed model should identify relevant costs of establishing the infrastructure necessary to take and process photographs, administer patients through the DERM process, and any further steps further to the implementation of the technology in settings with and without existing teledermatology services.

TABLE 19 Skin Analytics DERM pricing

Component	£ per 10k	£ per 2WW	Description
Base platform with DERM review	3300	30.00	Image and medical history capture platform, DERM assessment, PDF report with suspected diagnosis and recommended next steps.
Teledermatology functionality add on (optional)	900	8.20	Specialist teledermatology functionality within Skin Analytics' system to allow clinical staff to virtually review patient's cases and decide on the most appropriate outcome.
Discount if contributing outcome data (optional)	(250)	(2.30)	Discount provided if > 50% of biopsy results for patients through the pathway are shared with Skin Analytics.
Total cost per year (ex VAT) – *with outcomes discount*	3950	35.90	
Total cost per year (ex VAT) – *without outcomes discount*	4200	38.20	
Second read (Skin Analytics dermatologist)	£17 per case		

TABLE 20 Moleanalyzer pricing

Pricing option	Cost
FotoFinder Moleanalyzer AIMEE scoring (flat per year)	£1210
FotoFinder Moleanalyzer Pro includes AIMEE offline package (per year)	£1750

AIMEE, artificial intelligence mole examination and evaluation

FotoFinder Moleanalyzer Pro

FotoFinder provided details of the costs associated with Moleanalyzer Pro. The company provided the costs in *Table 20* for the technology. It was unclear from the company's submission how these pricings applied, for example, whether on a per-user basis or otherwise. The company stated that there was no cost for training and indicated that there was a discount for multi-user access. Full pricing details should be incorporated into a future cost-effectiveness model.

TABLE 21 Cost items required for the proposed model

Parameter	Exeter model value	EAG identified costs	EAG comment
Dermatological appointment (outpatient)	£142 NHS reference costs (WF01B 330 – first attendance)	**WF01A** – non-admitted, follow-up: Non-consultant led: £129.26 Consultant-led: £163.41 **WF01B** – non-admitted, first visit: Non-consultant led: £143.81 Consultant-led: £163.39	Clarification should be sought as to the appropriate reference cost.
Teledermatology	£25 10 minutes of 'hospital-based consultant' time, with additional 15 minutes band 3 administration time – unit costs from PSSRU	**WF01C** – non-admitted, non-face to face, follow-up: Non-consultant-led: £121.20 Consultant-led: £115.44 **WF01D** – non-admitted, non-face to face, first visit: Non-consultant-led: £284.09 Consultant-led: £114.52	As NHS reference costs appear considerably higher than the values applied in the Exeter model, clarification should be sought as to appropriate unit costs.
Biopsy + excision	£507 – inclusive of biopsy, SLNB, and surgical treatment in a single sitting, NIHR costing	**JC42C** – outpatient, intermediate skin procedures, 19 years and over: £257.43	
SLNB	See above	**WH54A** – admitted patient care, day case, CC Score 1 +: £1584.52 **WH54B** – admitted patient care, day case, CC Score 0: £1510.75	
BRAF testing and reporting	Testing: £374 Reporting: £113	£37 – Olaparib STA[83]	
Ultrasound	£248	**RD43Z**: ultrasound scan duration 20 minutes+: By department code: IMAGDA: £155.34 IMAGOP: £293.54 SI: £160.26	
CT scan	£108	**RD26Z**: computerised tomography scan, three areas, with contrast: By department code: IMAGDA: £139.49 IMAGOP: £146.34 IMAGOTH: £88.74 SI: £164.08	

Diagnosis, treatment and follow-up costs

As discussed, any future cost–utility model should be parameterised using NHS Reference Costs and costs provided by the PSSRU for consistency with other models considered by NICE. Unit costs should be applied to resource use assumptions informed by NICE guidelines.

The EAG have outlined a non-exhaustive list of unit costs in *Table 21* that could be adapted to implement into a future cost-effectiveness model, alongside a comparison with the values used by the Exeter/Skin Analytics model. For implementation into a future cost-effectiveness model, unit costs should be updated based on the most recent published reference costs.

Strengths and limitations of the proposed modelling approach

The conceptual model described by the EAG is based primarily on a synthesis of the economic evidence identified in the economic review, as well as evidence submitted by Skin Analytics. The presented model considers the currently available evidence and identifies areas where further research is required.

Strengths of the EAG's approach to the conceptual model include that it draws on precedent within the indication and other analyses considered by NICE to inform the structure, key assumptions and parameterisation. The conceptual model better aligns with the NICE reference case, through the use of more consistent cost and utility data sources and methods of analysis. The alternative structure proposed by the EAG for patients with BCC better represents the long-term consequence of BCC in terms of recurrence and therefore better captures the consequence of a FN case.

Limitations of the model proposed by the EAG include that the model cannot capture one of the primary benefits of the system, namely non-cash-releasing benefits (in common with other identified models). The hybrid structure proposed (a decision tree and Markov extension) cannot meaningfully quantify the impact of reducing demand on services in terms of reducing waiting times (and potential improvements in quality of care) for a specialist consultation across all dermatological indications. A more complex modelling approach would be required to capture demand, capacity and temporal dynamics.

Summary of evidence requirements

To inform a future cost-effectiveness model, future research should focus on addressing the limitations of the clinical evidence that would allow greater certainty in comparative diagnostic accuracy of AI technologies against comparators. As discussed above, the clinical evidence identified in *Chapter 3* was based on heterogeneous pathways and settings and may not provide appropriate diagnostic accuracy inputs for the pathway described in this model. The EAG consider that studies reporting the diagnostic accuracy of should have the following characteristics:

- **Setting:** UK peri-referral (following referral from primary care, before secondary care investigations)
- **Intervention:** AI technologies (with and without human confirmatory read)
- **Comparators**: Face-to-face and teledermatology
- **Outcome**: Diagnostic accuracy (sensitivity and specificity) of individual tests and the overall pathway

Chapter 7 Discussion

Statement of principal findings

Diagnostic accuracy and clinical impact

DERM

Three studies of DERM were examined to assess diagnostic accuracy. Autonomous use of DERM appears to have a high diagnostic accuracy for detection of malignant lesions: with a summary sensitivity of around 96.1% (95% CI 95.4 to 96.8) for a specificity of around 65.4% (95% CI 64.7 to 66.1). Similar diagnostic accuracies were found for detecting specific types of cancer (melanoma or SCC). There was some evidence that DERM might misdiagnose BCC cases as SCC or melanoma. Results for malignancy were similar across published and unpublished data. The summary sensitivity when detecting benign lesions was 71.5% (95% CI 70.7 to 72.3) for a specificity of 86.2% (95% CI 85.4 to 87.0).

The diagnostic accuracy of autonomous use of DERM appears to be similar to the diagnostic accuracy of dermatologists without DERM. (confidential information has been removed). The diagnostic accuracy of the whole teledermatology pathway including DERM could not be reliably assessed because of a lack of any independent reference standard of diagnosis.

The EAG found very limited evidence on the broader clinical impact of DERM, most of it unpublished. The evidence suggested that if DERM were used on its own around half of all patients would be discharged, and half referred for further assessment (either in person or through teledermatology). About 0.8% of patients would be discharged with a malignant lesion, mostly with BCC. (confidential information has been removed)

Patient opinion was broadly supportive of using DERM in some form as part of their diagnosis, but patients were divided on whether they preferred teledermatology to face-to-face appointments. Clinicians were generally very resistant to using DERM in isolation without human assessment of lesions.

Moleanalyzer Pro

Two prospective studies of Moleanalyzer Pro were identified; neither were performed in the UK. Moleanalyzer Pro alone had a sensitivity of 84.4% (95% CI 73.9 to 91.0) and a specificity of 84.5% (95% CI 72.0 to 92.1) to detect melanoma from a meta-analysis of the studies. This appeared similar to the accuracy of dermatologists alone. No eligible evidence was found for the diagnosis of SCC, BCC or other cancers.

The EAG did not identify any relevant evidence on the clinical impact of using Moleanalyzer Pro.

Patient and clinician opinion was generally supportive of using Moleanalyzer Pro in some way to aid diagnosis. However, the overwhelming majority of patients indicated that they would like the opinion of an expert physician besides an AI-assisted diagnosis.

Cost-effectiveness review and stakeholder submissions

No published assessments of the cost-effectiveness of the named AI technologies in an NHS setting were identified. Three published cost-effectiveness studies were identified evaluating any diagnostic technology for skin cancer in an NHS setting. All three studies focused on melanoma but also consider other skin cancers (e.g. BCC). While all identified studies adopted similar model structures, the mechanisms by which diagnostic accuracy generated value (in terms of either cost savings or QALY gain) differed substantively across studies. In particular, diagnostic sensitivity had less value in some models with value instead generated by the avoidance of unnecessary referral and diagnostic procedures. This is exemplified in one identified model of BCC in which it was assumed that all cases were correctly identified, and, as such, there were no cost or health consequences from improving diagnostic accuracy. Conversely, in other models, improved sensitivity and reduced frequency of missed diagnoses were the main drivers of benefits. In these

models, greater emphasis was placed on the consequences of missed diagnoses, with more granular modelling of the consequences of disease progression and mortality.

The EAG received several submissions that included relevant economic analysis. This included a preliminary report describing a cost–utility model developed by Exeter Test Group and Skin Analytics, a pilot evaluation of DERM for the skin cancer 2WW pathway at UHL, and several economic analyses commissioned by NHSE (Unity Insights and University of Surrey). All three submissions assessed the value of using DERM in an NHS setting. No economic evidence in support of Moleanalyzer Pro was submitted.

The most comprehensively reported and relevant of these was the cost–utility model developed by Exeter Test Group. This model built upon the three previous skin cancer models identified in the EAG's review. Aligning with the proposed use case, this model represents an assessment of DERM in a post-referral setting, with and without a second read, compared with teledermatology and the conventional urgent referral model (face-to-face). It considered three diagnostic categories: high-risk cancer, BCC and non-/pre-cancer.

The EAG considered the model structure largely appropriate to assess core aspects of the potential value of AI technologies for identifying benign lesions in a post-referral setting, but noted several issues which may mean that the main value drivers may not be appropriately characterised. Namely, the model imposed disincentives for the correct diagnosis and treatment of BCC, which rewarded the comparatively lower sensitivity of DERM; assumptions around post-triage diagnostic accuracy of face-to-face assessment which structurally assumed benefits for any strategy incorporating a triage step; costs associated with diagnostic investigations and treatment may be inconsistent with sources generally used in NICE appraisals, and may overvalue specificity in terms of generating cost savings; and the derivation of the HRQoL value set is not aligned with the NICE Reference Case. It remains highly uncertain whether currently available diagnostic accuracy evidence is sufficient to reliably populate a cost–utility model, particularly with regard to the comparative specificity of AI technologies to an effectively implemented teledermatology service. Therefore, while this analysis predicted that DERM with or without a second read would dominate all other options, this was highly dependent on the relative specificity of teledermatology.

Conceptual model

The EAG outlines a conceptual model which aims to provide an alternative to that described in the Skin Analytics submission. The proposed conceptual model seeks to address methodological issues identified in the reviewed literature and to explore the necessary structure and evidence required for future model development. For patients with high-risk cancers, the model structure described in the Skin Analytics model would be preserved. An alternative structure is, however, proposed to capture the natural history of BCC in FNs, to better reflect the long-term health- and cost-consequences of BCC.

While cost–utility models have recently been built in support of the present decision problem, the EAG consider the available evidence inadequate to characterise the potential value of these technologies in an NHS setting. In particular, the EAG highlights limitations in comparative diagnostic accuracy evidence for the named technologies. Current evidence for both DERM and Moleanalyzer Pro is lacking with regard to the diagnostic accuracy of the whole diagnostic pathway (i.e. inclusive of subsequent steps). Availability of these data is essential to understanding the likelihood of missed cases which cannot be inferred from the partial data currently available. Similarly, comparable diagnostic accuracy data describing current service provision is lacking, particularly for the teledermatology pathway. Without comparative evidence on the diagnostic accuracy of AI technologies and teledermatology, their relative value for safe and cost-effective identification of benign lesions will remain unclear.

The EAG also note a lack of robust data available to inform progression probabilities in undiagnosed disease, and a focus on expert-elicited parameters in previous cost–utility models. Establishing rates of progression and ultimately the consequences of missed diagnosis is important to characterising trade-offs in sensitivity and other potential cost savings. While adjunctive AI technologies have principally been positioned as means of more efficiently identifying benign lesions, the introduction of further triage steps may also impact pathway sensitivity and are likely to represent part of the value case for AI technologies.

The EAG propose that a future cost-effectiveness model should use unit costs obtained based on the NICE Reference Case from national sources, namely the latest NHS Reference Costs and Unit Costs of Health and Social Care (PSSRU) where available, with costs supplemented with those identified by a systematic review of the literature. The EAG also note that costs of establishing the necessary services to implement the technology in trusts without existing teledermatology infrastructure have not been characterised. It may be appropriate to also include these start-up costs within any economic analysis.

Strengths and limitations of the assessment

Strengths

This report presents an extensive systematic review of all published and unpublished evidence on DERM and Moleanalyzer Pro. The consistency between evidence identified through database searches, and that supplied by the companies, suggests that this report covers all the relevant evidence on the two technologies.

Skin Analytics supplied a large quantity of evidence on DERM, including raw study data and unpublished study reports and economic analyses. This enabled a more thorough investigation of the clinical value and cost-effectiveness of DERM than would have been possible if using only published studies.

The outlined conceptual model addresses limitations with currently proposed models to more comprehensively evaluate both the short-term costs and consequences associated with alternative diagnostic strategies.

Limitations

Given the short time frame for this project, a rapid review approach was used. Database searches were more limited than for a full review and were focused on publications explicitly naming DERM or Moleanalyzer Pro. We acknowledge that some relevant material may have been missed, although the consistency of our findings with material supplied by the companies reduces this risk.

The use of a rapid review approach also meant that we restricted full data extraction and synthesis to studies with prospective inclusion of patients, and to the most recent versions of the two technologies. This may mean that some useful evidence has not been considered. However, we consider that our approach has focused on the highest quality evidence of most relevance to practice.

The rapid review approach and limitations in the evidence base meant that the capacity to synthesise evidence was limited. Meta-analysis was not feasible for most outcomes, and many key outcomes were only reported in one publication or source.

The EAG consider that while the proposed conceptual model improves upon the approaches taken by existing studies, the proposed model (as with all other identified studies) fails to capture non-cash benefits associated with demand on dermatologist time. To capture these benefits, a more complex simulation approach would be required, capturing demand, capacity and temporal dimensions.

The EAG were unable to provide an assessment of the likely budget impact and resource use which was a stated objective of the project. This in part reflects the compressed timelines and late provision of materials by Skin Analytics. However, uncertainties in the applicable unit costing and underlying diagnostic accuracy associated with each technology would likely limit the strength of conclusions that could be drawn from such analysis.

Key limitations of the evidence base

Diagnostic accuracy

Only three studies of DERM and two studies of Moleanalyzer Pro that prospectively evaluated the diagnostic accuracy of AI in clinical practice were identified. Hence, the evidence base for the technologies is modest. The prospective

studies of Moleanalyzer Pro were conducted outside of the UK, were not explicitly in a teledermatology setting, and did not evaluate the accuracy of AI for detecting non-melanoma cancer.

The DERM versions (in particular, the set sensitivity/specificity thresholds) and the dermatoscopes used for clinical assessments were out of date; therefore, the applicability of the diagnostic accuracy results to current practice is uncertain.

Most patients included in diagnostic accuracy had lighter skin colours (Fitzpatrick types II–III). The restricted eligibility to DERM and Moleanalyzer Pro and the systematic exclusion of a significant proportion of participants who would normally be assessed in practice meant that the evidence base for both devices was considered to be at high risk of bias and raises concerns about its applicability to practice.

In all except one diagnostic accuracy study, only a subset of participants (those with suspected malignancy) had a reference standard test that included histopathology. Although this is reflective of practice, the risk of reference standard test misclassification in these studies cannot be excluded.

Clinical impact and benefit

There is no evidence on the impact of DERM or Moleanalyzer Pro on clinical morbidities, mortality and HRQoL. In particular, the EAG note that there is no substantive evidence on the benefits or harms AI use might have for patients.

Evidence from healthcare practitioners on their confidence in DERM and its clinical and broader impact on the pathway and patient management is limited, although initial evidence from limited samples suggested that patients and clinicians do not support the autonomous use of AI tools.

Use of DERM has been limited to smaller lesions and lesions that are easier to photograph (e.g. not concealed by hair or tattoos) and excluded atypical locations such as palm or soles. This may restrict its use in actual clinical practice.

Resource use

Evidence on resource use for DERM was mostly limited to some unpublished results. Much of this evidence compared DERM as part of the teledermatology pathway to face-to-face dermatology. Consequently, the impact on resource use attributable specifically to DERM is uncertain. In particular, how autonomous use of DERM might compare to DERM combined with dermatologist assessment is unclear.

The EAG found no evidence on the impact of Moleanalyzer Pro on resource use.

Cost-effectiveness

No evidence on the cost-effectiveness of either DERM or Moleanalyzer Pro was identified in the EAG's review of published evidence. Evidence on cost-effectiveness of DERM submitted by Skin Analytics and NHSE (Unity Insights and University of Surrey) was both preliminary and incomplete. Uncertainties in the main value drivers including diagnostic accuracy of both DERM and comparator technologies limit the conclusions that can be drawn from this evidence. A more complete understanding of the economic analysis commissioned by NHSE may address some of these uncertainties.

Patient and public inclusion

The short time frame of this assessment meant the EAG did not seek any independent public or patient involvement. Patient representatives were included on the scoping committee for this assessment and will be involved in the decision-making process based on this report.

At scoping, patient representatives identified several key issues for consideration:

- The need to ensure that use of AI does not lead to malignant lesions being missed.

- Concerns around equality due to difficulty in assessing lesions covered by tattooing, hair or scarring, or in hard-to-assess areas.
- Equality issues around diagnosis of skin cancer in people with darker skin or non-white ethnicity.
- The need to reduce anxiety created by the diagnostic process (e.g. due to long waits for diagnoses, or incorrect initial diagnoses).

The EAG note that this report was largely unable to resolve these issues; see discussion in *Key limitations of the evidence base* and *Suggested research priorities*.

Equality, diversity and inclusion

As this was a rapid review of existing evidence, the EAG could not consider equality issues beyond what was available in publications or supplied material.

The EAG notes several equality concerns arising from our review:

The evidence base for both technologies included few patients with non-white ethnicity or darker skin tones. Since skin cancer may be harder to detect in these people, this is of concern. It is unclear whether the AI tools have been properly validated in people with darker skin tone, and what is the resulting diagnostic accuracy. Differences in diagnostic accuracy could lead to inequalities due to different diagnostic pathways, such as if some people have to wait for a face-to-face appointment because an AI assessment was inconclusive.

DERM could not be used for a substantial number of patients, due to lesions being too large to assess; lesions being in areas with tattoos, scarring or hair covering; or lesions being on parts of the body unsuited to assessment with a dermatoscope. This could potentially cause inequalities due to resulting differences in diagnostic pathways and access to diagnostic services.

Use of AI could improve access to skin cancer diagnosis as it may reduce the need for face-to-face appointments, so reducing patient time commitment and need to travel to appointments.

Chapter 8 Conclusions

Implications for service provision

The high diagnostic accuracy of DERM suggests that it has potential for use as a triage and diagnostic tool for skin cancer in a post-referral setting. This could be either as part of a teledermatology pathway alongside assessment by dermatologists or as an autonomous diagnostic tool where it replaces some of the need for consultant-led teledermatology.

Although evidence on the clinical impact of DERM was limited, it did suggest that, in eligible lesions, autonomous use of DERM could reduce the need for human dermatology assessment, without substantially adversely affecting accuracy. The practical impact and clinical benefit of using DERM in combination with dermatologist assessment are currently unclear, particularly when compared to teledermatology without using DERM. Current economic evidence to support the cost-effectiveness of DERM is also limited, and it is unclear whether the plausible advantages of DERM represent value for money. On the basis of early modelling exercises, there is a reasonably high certainty that DERM has the potential to be used cost-effectively in the post-referral setting, compared to the traditional urgent skin cancer referral pathway. It is less clear whether DERM has potential to be cost-effective compared to teledermatology without DERM. DERM with a second read is less likely to generate cost savings versus conventional teledermatology, but may have non-cash-releasing benefits (e.g. reduced waiting times, quality of care improvements) associated with outsourcing of consultant review to Skin Analytics.

The EAG consider that the evidence on Moleanalyzer Pro is too limited to judge how it might be used in practice. Currently, prospective studies in clinical practice have only assessed its accuracy in diagnosing melanoma. It is unclear whether it could be adapted to detect all forms of skin cancer, or if not, how a melanoma-only AI tool would be used in practice. As Moleanalyzer Pro has not been tested in the UK as part of a teledermatology programme, it is currently unclear what clinical benefits it could have within NHS practice. There is, similarly, no economic evidence to support the use of Moleanalyzer Pro in an NHS setting. Assuming a similar use case to DERM and appropriate data collection, the value of Moleanalyzer Pro could be assessed using the conceptual model outlined by the EAG.

The substantial resistance from both patients and clinicians to using AI without any human dermatological assessment means that if AI is to be used autonomously in some way, more robust evidence that is applicable to current practice is needed to demonstrate that it has clear benefits to patients, without sacrificing accuracy.

Suggested research priorities

Diagnostic accuracy

Future diagnostic studies should, where possible, examine and compare the diagnostic accuracy of:

- AI as a standalone device
- AI in combination with human teledermatology (e.g. with a ‘second read’ for all AI-assessed lesions)
- Teledermatology without AI
- Face-to-face assessment without teledermatology

The setting of future studies should be clearly reported and include UK peri-referral (following referral from primary care, before secondary care investigations).

There is a need for further research on the diagnostic accuracy of AI compared to standard teledermatology in specific patient subgroups:

- in individuals with darker skin types (Fitzpatrick IV–VI) and a broad range of ethnicity groups
- for lesion types and lesions located in body sites and not currently covered by DERM and Moleanalyzer Pro evidence
- to identify rare skin cancers.

All future studies should use adequate blinding between AI and dermatologists, and use an appropriate and robust reference standard. Particularly, an independent and blinded 'ground truth' diagnosis from dermatologists not involved in the teledermatology process, and with appropriate follow-up, is needed for all lesions that are not assessed with histology. Future studies should use up-to-date dermatoscopes to address the limited applicability of existing studies.

Future studies should also follow relevant reporting guidance.[84] This includes clarity on the pathway and positioning of AI within it, clear documentation of reasons for test failures and exclusions (including eligibility assessment and exclusions from analysis), diagnostic accuracy cut-offs (and timing at which these are specified) and reference standard definitions. Versions of AI devices (including algorithms versions, whether used offline or online) and dermatoscopes where applicable should be reported clearly to inform applicability to practice. Diagnostic accuracy should preferably be reported with sufficient granularity (such as with detailed matrices) so as to evaluate sensitivity and specificity by type of cancer. For patients with multiple lesions, studies should specify whether and how the risk of within-patient correlation was addressed.

All DERM and most Moleanalyzer Pro studies were co-authored by staff affiliated to their respective device manufacturer. There is a need for independent evaluations of DERM and Moleanalyzer Pro in clinical practice, using commonly agreed, standardised interoperable systems and agreed standards of data collection. Evaluations of Moleanalyzer Pro (ideally, in head-to-head designs against DERM) within a UK post-referral context are required to assess whether Moleanalyzer Pro is a suitable, autonomous alternative to DERM, including for the detection of non-melanoma cancer.

Clinical impact

The overall impact of AI requires evaluation, including clinical output (such as referral types and waiting times) throughout the clinical pathway, and longer-term morbidity and mortality outcomes. There is a lack of prospective data available to inform progression of disease in patients with missed diagnosis, which is required to appropriately populate an economic model. Larger, independent prospective studies are needed that examine all clinical impact outcomes where evidence is currently absent. These studies should also examine the perceptions of DERM from healthcare professionals and patients, and the impact of DERM on the diagnostic pathway and patient care, to further understand potential barriers to implementation.

Further evaluations of DERM are ongoing across a range of centres in the UK, including the post-2WW referral pathway and in the pre-referral setting. Although few details were reported, it is hoped that this future evidence will address whether DERM can provide clear clinical benefit, perform consistently and be received positively across a range of local services with differing case mix and pathways.

Evidence on the clinical value of Moleanalyzer Pro is required. This should ideally be through prospective observational cohort studies where Moleanalyzer Pro is used within an NHS 2WW referral pathway setting, along similar lines to the existing studies of DERM.

Cost-effectiveness and resource use

The use of AI technologies to direct the discharge of patients with benign lesions following referral from primary care has a range of cost and resource consequences which have not been adequately characterised in existing models. Company-sponsored analyses suggested that DERM used autonomously and with a second read could be highly cost-effective compared to current urgent skin cancer referral models. However, much of this value is generated through potentially optimistic assumptions around the diagnostic accuracy of comparators, and of the surrounding pathway. The parameterisation of these analyses is not aligned with NICE precedent, which may overvalue the cost and health implications of DERM. A (confidential information has been removed). Notably, the magnitude of uncaptured non-cash-releasing benefits remains unquantified.

While a conventional cost–utility analytical approach is able to capture important direct cost and health implications of alternative diagnostic strategies, a lack of key comparative data means the relative clinical and cost-effectiveness of pathways incorporating AI technologies and teledermatology remains highly uncertain. Directly comparable evidence

on the diagnostic accuracy of AI technologies and teledermatology in a post-referral setting compared with unassisted teledermatology is required to assess the potential value of AI technologies.

A better understanding of the resource implications associated with the implementation of AI technologies will also require further research to establish the costs to the NHS associated with current pathways. Evidence submitted to the EAG demonstrated that the costs of both teledermatology and face-to-face assessments are key value drivers.

Where possible, future studies should seek to address these uncertainties by collecting appropriate data on resource implications including impacts on healthcare professionals' time, set-up and operational costs associated with both teledermatology and AI technologies in trusts without existing infrastructure, as well as the proportion of patients eligible (and the effect of characteristics determining ineligibility) for AI/teledermatology assessment. Further research must also be undertaken to quantify the benefits to population health within skin cancer and other dermatology indications associated with any release of NHS consultant dermatologist resource, and understand the effects of these technologies on waiting times for final diagnosis.

Additional information

CRediT contribution statement

Matthew Walton (https://orcid.org/0000-0003-1932-3689): Conceptualisation, Formal analysis (co-lead), Investigation (co-lead), Methodology (co-lead), Writing.

Alexis Llewellyn (https://orcid.org/0000-0003-4569-5136): Conceptualisation, Data curation, Formal analysis, Investigation, Writing.

Eleanora Uphoff (https://orcid.org/0000-0002-9759-2502): Data curation, Formal analysis, Investigation, Writing.

Joseph Lord (https://orcid.org/0000-0002-2131-782X): Investigation, Formal analysis, Writing.

Melissa Harden (https://orcid.org/0000-0003-2338-6869): Data curation, Investigation.

Rob Hodgson (https://orcid.org/0000-0001-6962-2893): Conceptualisation, Formal analysis, Writing.

Mark Simmonds (https://orcid.org/0000-0002-1999-8515): Conceptualisation (lead), Data curation (lead), Formal analysis (lead), Investigation (lead), Methodology (lead), Project administration, Writing (lead).

Acknowledgements

The EAG would like to acknowledge the help and advice from all members of the scoping committee for this assessment. In particular, we acknowledge the advice provided by Dr Rubeta Matin (Consultant Dermatologist, Oxford University Hospitals).

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. The authors are not able to supply any confidential or unpublished data.

Ethics statement

As this was a systematic review of existing published data, no ethics approval was required.

Information governance statement

All non-confidential data used in this paper were taken from published sources: no personal data were included.

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/GJMS0317.

Primary conflicts of interest: All authors have no conflicts of interest to declare.

Disclaimer

Every effort has been made to obtain the necessary permissions for reproduction, to credit original sources appropriately and to respect copyright requirements. However, despite our diligence, we acknowledge the possibility of unintentional omissions or errors and we welcome notifications of any concerns regarding copyright or permissions.

References

1. *The Two-week Wait Skin Cancer Pathway: Innovative Approaches to Support Early Diagnosis of Skin Cancer as Part of the NHS COVID-19 Recovery Plan. 4 April 2022 Version 1*. NHS and British Association of Dermatologists; 2022.
2. Levell N. *Dermatology. GIRFT Programme National Specialty Report*. GIRFT and NHS; 2021.
3. *Clinical Knowledge Summary – Melanoma*. National Institute of Health and Care Excllence (NICE); 2022. URL: https://cks.nice.org.uk/topics/melanoma/ (accessed 26 October 2023).
4. Kwiatkowska M, Ahmed S, Ardern-Jones M, Bhatti LA, Bleiker TO, Gavin A, *et al*. An updated report on the incidence and epidemiological trends of keratinocyte cancers in the United Kingdom 2013–2018. *Skin Health Dis* 2021;**1**:e61. https://doi.org/10.1002/ski2.61
5. *Clinical Knowledge Summary – Melanoma. How Should I Assess a Lesion?* National Institute of Health and Care Excellence (NICE); 2022. URL: https://cks.nice.org.uk/topics/melanoma/diagnosis/assessment/ (accessed 26 October 2023).
6. *Non-melanoma Skin Cancer*. NHS. URL: www.nhs.uk/conditions/non-melanoma-skin-cancer/ (accessed 26 October 2023).
7. *Clinical Knowledge Summary – Skin Cancers – Recognition and Referral*. National Institute of Health and Care Excellence (NICE); 2021. URL: https://cks.nice.org.uk/topics/skin-cancers-recognition-referral/ (accessed 26 October 2023).
8. *Faster Diagnosis*. NHS England. URL: www.england.nhs.uk/cancer/faster-diagnosis/ (accessed 26 October 2023).
9. National Institute for Health and Care Excellence. *Suspected Cancer: Recognition and Referral (NG12)*. London: NICE; 2023.
10. Bhave P, Wu Y, Fernandez-Peñas P. *Melanoma*. BMJ Best Practice; 2023. URL: https://bestpractice.bmj.com/topics/en-gb/268 (accessed 26 October 2023).
11. Marsden JR, Newton-Bishop JA, Burrows L, Cook M, Corrie PG, Cox NH, *et al*.; British Association of Dermatologists (BAD) Clinical Standards Unit. Revised UK guidelines for the management of cutaneous melanoma 2010. *J Plast Reconstr Aesthet Surg* 2010;**63**:1401–19. https://doi.org/10.1016/j.bjps.2010.07.006
12. National Institute of Health and Care Excellence. *Improving Outcomes for People with Skin Tumours Including Melanoma. Cancer Service Guideline [CSG8]*. NICE; 2010. URL: www.nice.org.uk/guidance/csg8 (accessed 26 October 2023).
13. Keohane SG, Botting J, Budny PG, Dolan OM, Fife K, Harwood CA, *et al*.; British Association of Dermatologists' Clinical Standards Unit. British Association of Dermatologists guidelines for the management of people with cutaneous squamous cell carcinoma 2020. *Br J Dermatol* 2021;**184**:401–14. https://doi.org/10.1111/bjd.19621
14. National Institute of Health and Care Excellence. *Melanoma: Assessment and Management. NICE Guideline [NG14]*. NICE; 2022. URL: www.nice.org.uk/guidance/ng14/chapter/Recommendations#assessing-melanoma (accessed 26 October 2023).
15. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. York: CRD, University of York; 2009.
16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* 2021;**10**:89. https://doi.org/10.1186/s13643-021-01626-4
17. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the P-DTAG, *et al*. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;**319**:388–96. https://doi.org/10.1001/jama.2017.19163

18. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol* 2016;**75**:40–6. https://doi.org/10.1016/j.jclinepi.2016.01.021

19. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.*; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

20. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, *et al.*; QUADAS-C Group†. QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med* 2021;**174**:1592–9. https://doi.org/10.7326/m21-2234

21. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, *et al*. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ* 2020;**368**:l6890. https://doi.org/10.1136/bmj.l6890

22. Ayiku L, Levay P, Hudson T, Craven J, Finnegan A, Adams R, Barrett E. The Embase UK filter: validation of a geographic search filter to retrieve research about the UK from OVID Embase. *Health Info Libr J* 2019;**36**:121–33. https://doi.org/10.1111/hir.12252

23. Ayiku L, Levay P, Hudson T, Craven J, Barrett E, Finnegan A, Adams R. The MEDLINE UK filter: development and validation of a geographic search filter to retrieve research about the UK from OVID MEDLINE. *Health Info Libr J* 2017;**34**:200–16. https://doi.org/10.1111/hir.12187

24. Marsden H, Morgan C, Austin S, DeGiovanni C, Venzi M, Kemos P, *et al*. Effectiveness of an image analyzing AI-based digital health technology to identify non-melanoma skin cancer and other skin lesions: results of the DERM-003 study. *Front Med (Lausanne)* 2023;**10**:1288521. https://doi.org/10.3389/fmed.2023.1288521

25. Marsden H, Kemos P, Venzi M, Noy M, Maheswaran S, Francis N, *et al*. Accuracy of an artificial intelligence as a medical device as part of a UK-based skin cancer teledermatology service. *Front Med (Lausanne)* 2024;**11**:1302363. https://doi.org/10.3389/fmed.2024.1302363

26. Thomas L, Hyde C, Mullarkey D, Greenhalgh J, Kalsi D, Ko J. Real-world post-deployment performance of a novel machine learning-based digital health technology for skin lesion assessment and suggestions for post-market surveillance. *Front Med (Lausanne)* 2023;**10**:1264846. https://doi.org/10.3389/fmed.2023.1264846

27. *An Evaluation of AI Powered Teledermatology for Skin Cancer 2WW Pathway*. East Midlands Academic Health Science Network; 2023.

28. MacLellan AN, Price EL, Publicover-Brouwer P, Matheson K, Ly TY, Pasternak S, *et al*. The use of noninvasive imaging techniques in the diagnosis of melanoma: a prospective diagnostic accuracy study. *J Am Acad Dermatol* 2021;**85**:353–9. https://doi.org/10.1016/j.jaad.2020.04.019

29. Winkler JK, Blum A, Kommoss K, Enk A, Toberer F, Rosenberger A, Haenssle HA. Assessment of diagnostic performance of dermatologists cooperating with a convolutional neural network in a prospective clinical study: human with machine. *JAMA Dermatol* 2023;**159**:621–7. https://doi.org/10.1001/jamadermatol.2023.0905

30. Fink C, Blum A, Buhl T, Mitteldorf C, Hofmann-Wellenhof R, Deinlein T, *et al*. Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas. *J Eur Acad Dermatol Venereol* 2020;**34**:1355–61. https://doi.org/10.1111/jdv.16165

31. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, *et al.*; Reader study level-I and level-II Groups. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018;**29**:1836–42. https://doi.org/10.1093/annonc/mdy166

32. Haenssle HA, Fink C, Toberer F, Winkler J, Stolz W, Deinlein T, *et al.*; Reader Study Level I and Level II Groups. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. *Ann Oncol* 2020;**31**:137–43. https://doi.org/10.1016/j.annonc.2019.10.013

33. Kommoss KS, Winkler JK, Mueller-Christmann C, Bardehle F, Toberer F, Stolz W, *et al*. Observational study investigating the level of support from a convolutional neural network in face and scalp lesions deemed diagnostically 'unclear' by dermatologists. *Eur J Cancer* 2023;**185**:53–60. https://doi.org/10.1016/j.ejca.2023.02.025

34. Sies K, Winkler JK, Fink C, Bardehle F, Toberer F, Buhl T, *et al*. Past and present of computer-assisted dermoscopic diagnosis: performance of a conventional image analyser versus a convolutional neural network in a prospective data set of 1,981 skin lesions. *Eur J Cancer* 2020;**135**:39–46. https://doi.org/10.1016/j.ejca.2020.04.043

35. Sies K, Winkler JK, Fink C, Bardehle F, Toberer F, Kommoss FKF, *et al*. Dark corner artefact and diagnostic performance of a market-approved neural network for skin cancer classification. *J Dtsch Dermatol Ges* 2021;**19**:842–51. https://doi.org/10.1111/ddg.14384

36. Sies K, Winkler JK, Fink C, Bardehle F, Toberer F, Buhl T, *et al*. Does sex matter? Analysis of sex-related differences in the diagnostic performance of a market-approved convolutional neural network for skin cancer detection. *Eur J Cancer* 2022;**164**:88–94. https://doi.org/10.1016/j.ejca.2021.12.034

37. Winkler JK, Sies K, Fink C, Toberer F, Enk A, Deinlein T, *et al*. Melanoma recognition by a deep learning convolutional neural network – performance in different melanoma subtypes and localisations. *Eur J Cancer* 2020;**127**:21–9. https://doi.org/10.1016/j.ejca.2019.11.020

38. Winkler JK, Sies K, Fink C, Toberer F, Enk A, Abassi MS, *et al*. Collective human intelligence outperforms artificial intelligence in a skin lesion classification task. *J Dtsch Dermatol Ges* 2021;**19**:1178–85. https://doi.org/10.1111/ddg.14510

39. Winkler JK, Sies K, Fink C, Toberer F, Enk A, Abassi MS, *et al*. Association between different scale bars in dermoscopic images and diagnostic performance of a market-approved deep learning convolutional neural network for melanoma recognition. *Eur J Cancer* 2021;**145**:146–54. https://doi.org/10.1016/j.ejca.2020.12.010

40. Winkler JK, Tschandl P, Toberer F, Sies K, Fink C, Enk A, *et al*. Monitoring patients at risk for melanoma: may convolutional neural networks replace the strategy of sequential digital dermoscopy? *Eur J Cancer* 2022;**160**:180–8. https://doi.org/10.1016/j.ejca.2021.10.030

41. Phillips M, Marsden H, Jaffe W, Matin RN, Wali GN, Greenhalgh J, *et al*. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Netw Open* 2019;**2**:e1913436. https://doi.org/10.1001/jamanetworkopen.2019.13436

42. Phillips M, Greenhalgh J, Marsden H, Palamaras I. Detection of malignant melanoma using artificial intelligence: an observational study of diagnostic accuracy. *Dermatol Pract Concept* 2020;**10**:e2020011. https://doi.org/10.5826/dpc.1001a11

43. Marsden H, Palamaras I, Kemos P, Greenhalgh J. Effectiveness of an image-analysing artificial intelligence-based digital health technology to diagnose nonmelanoma skin cancer and benign skin lesions. *Br J Dermatol* 2023;**188**:iv43. https://doi.org/10.1093/bjd/ljad113.091

44. Austin S, Palamaras I, Greenhalgh J, Kemos P, Marsden H. *Effectiveness of an image analysing AI-based digital health technology to diagnose non-melanoma skin cancer (NMSC) and benign skin lesions: results of the DERM-003 study (unpublished)*.

45. Kawsar A, Hussain A, Kalsi D, Kemos P, Marsden H, Thomas L. Patient perspectives of artificial intelligence as a medical device in a skin cancer pathway. *Front Med (Lausanne)* 2023;**10**:1259595. https://doi.org/10.3389/fmed.2023.1259595

46. DERM. *DERM Clinical Performance Report Q3 2023 (unpublished)*. DERM; 2023.

47. Skin Analytics. *Skin Analytics Case Study: Chelsea & Westminster Hospital NHS Foundation Hospital (unpublished)*. London: Skin Analytics; 2023.

48. Andrew K, Price T, Barlow N, Morris E, Ungureanu S, Zaki I. *Continued Improvement of Artificial Intelligence in Identifying Skin Cancer*. Paper presented at: EADV Congress, Berlin, 11–14 October 2023.
49. Skin Analytics. *Skin Analytics Case Study: University Hospitals Birmingham NHS Foundation Trust (unpublished)*. London: Skin Analytics; 2023.
50. Jenkins R, Brewer F, Kalsi D, Mullarkey D. *Clinical Performance of an Artificial Intelligence as a Medical Device Deployed within an Urgent Suspected Skin Cancer Pathway (unpublished)*. London: Skin Analytics; undated.
51. Baker KA, Roberts E, Harman K, Mullarkey D, Kalsi D. Using artificial intelligence to triage skin cancer referrals: outcomes from a pilot study. *Br J Dermatol* 2023;**188**:iv173. https://doi.org/10.1093/bjd/ljad113.372
52. *Case Study: Artificial Intelligence Helping to Speed Up Skin Cancer Diagnosis in Leicester, Leicestershire, and Rutland Integrated Care System [unpublished]*. NHS England; 2023. URL: https://www.england.nhs.uk/publication/case-study-artificial-intelligence-helping-to-speed-up-skin-cancer-diagnosis-in-leicester-leicestershire-and-rutland-integrated-care-system/
53. Baker KA, Roberts E, Kalsi D, Mullarkey D, Harman K. *Using Artificial Intelligence to Triage Skin Cancer Referrals: Outcomes from a Pilot Study (unpublished)*; undated.
54. Skin Analytics Limited. *DERM US and EU Validation Study*. ClinicalTrials.gov. Bethesda (MD): National Library of Medicine (US). URL: https://classic.clinicaltrials.gov/show/NCT05126173 (accessed 1 November 2023).
55. Skin Analytics. *A Study to Assess Whether an Artificial Intelligence (AI) Technology Can Be Used to Assess Images of Suspicious Skin Lesions to Determine the Likelihood of Skin Cancer; and Reduce the Number, and Cost Associated with, Biopsies of Skin Lesions in Primary Care*. Australia New Zealand Clinical Trials Registry; 2019. URL: https://anzctr.org.au/ACTRN12619000398101.aspx (accessed 1 November 2023).
56. Haenssle HA, Winkler JK, Fink C, Toberer F, Enk A, Stolz W, *et al.*; Reader study level-I and level-II Groups Christina Alt. Skin lesions of face and scalp – classification by a market-approved convolutional neural network in comparison with 64 dermatologists. *Eur J Cancer* 2021;**144**:192–9. https://doi.org/10.1016/j.ejca.2020.11.034
57. Tschandl P, Hofmann L, Fink C, Kittler H, Haenssle HA. Melanomas vs. nevi in high-risk patients under long-term monitoring with digital dermatoscopy: do melanomas and nevi already differ at baseline? *J Eur Acad Dermatol Venereol* 2017;**31**:972–7. https://doi.org/10.1111/jdv.14065
58. Chuchu N, Dinnes J, Takwoingi Y, Matin RN, Bayliss SE, Davenport C, *et al.*; Cochrane Skin Cancer Diagnostic Test Accuracy Group. Teledermatology for diagnosing skin cancer in adults. *Cochrane Database Syst Rev* 2018;**12**:CD013193. https://doi.org/10.1002/14651858.CD013193
59. Wilson EC, Emery JD, Kinmonth AL, Prevost AT, Morris HC, Humphrys E, *et al*. The cost-effectiveness of a novel SIAscopic diagnostic aid for the management of pigmented skin lesions in primary care: a decision-analytic model. *Value Health* 2013;**16**:356–66. https://doi.org/10.1016/j.jval.2012.12.008
60. Wilson ECF, Usher-Smith JA, Emery J, Corrie P, Walter FM. A modeling study of the cost-effectiveness of a risk-stratified surveillance program for melanoma in the United Kingdom. *Value Health* 2018;**21**:658–68. https://doi.org/10.1016/j.jval.2017.11.009
61. Edwards SJ, Mavranezouli I, Osei-Assibey G, Marceniuk G, Wakefield V, Karner C. VivaScope R 1500 and 3000 systems for detecting and monitoring skin lesions: a systematic review and economic evaluation. *Health Technol Assess* 2016;**20**:1–260. https://doi.org/10.3310/hta20580
62. Balch CM, Gershenwald JE, Soong SJ, Thompson JF, Atkins MB, Byrd DR, *et al*. Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol* 2009;**27**:6199–206. https://doi.org/10.1200/jco.2009.23.4799
63. Losina E, Walensky RP, Geller A, Beddingfield FC 3rd, Wolf LL, Gilchrest BA, Freedberg KA. Visual screening for malignant melanoma: a cost-effectiveness analysis. *Arch Dermatol* 2007;**143**:21–8. https://doi.org/10.1001/archderm.143.1.21

64. Hadley JC, Bendeck S, Brown K, Chen S. Quality of life impact by melanoma as measured by utilities. *J Am Acad Dermatol* 2004;**50**:P119. https://doi.org/10.1016/j.jaad.2003.10.402

65. Alarcon I, Carrera C, Palou J, Alos L, Malvehy J, Puig S. Impact of in vivo reflectance confocal microscopy on the number needed to treat melanoma in doubtful lesions. *Br J Dermatol* 2014;**170**:802–8. https://doi.org/10.1111/bjd.12678

66. Pellacani G, Pepe P, Casari A, Longo C. Reflectance confocal microscopy as a second-level examination in skin oncology improves diagnostic accuracy and saves unnecessary excisions: a longitudinal prospective study. *Br J Dermatol* 2014;**171**:1044–51. https://doi.org/10.1111/bjd.13148

67. Wilson ECF, Usher-Smith JA, Emery J, Corrie PG, Walter FM. Expert elicitation of multinomial probabilities for decision-analytic modeling: an application to rates of disease progression in undiagnosed and untreated melanoma. *Value Health* 2018;**21**:669–76. https://doi.org/10.1016/j.jval.2017.10.009

68. Veena P, Clare G, Prashanth P, Timothy D, Sajeda H, Virginia L, *et al*. Reducing vitamin D requests in a primary care cohort: a quality improvement study. *BJGP Open* 2020;**4**:bjgpopen20X. 101090. https://doi.org/10.3399/bjgpopen20X101090

69. Filby A, Lewis L, Taylor M. *An Economic Evaluation of Interventions to Improve the Uptake of Vitamin D Supplements in England and Wales*. London: National Institute for Health and Care Excellence; 2014.

70. Ethna M, Sarah D, Mark L. Skin in the game: the cost consequences of skin cancer diagnosis, treatment and care in Northern Ireland. *J Cancer Policy* 2024;**39**:100468. https://doi.org/10.1016/j.jcpo.2024.100468

71. Tromme I, Devleesschauwer B, Beutels P, Richez P, Leroy A, Baurain JF, *et al*. Health-related quality of life in patients with melanoma expressed as utilities and disability weights. *Br J Dermatol* 2014;**171**:1443–50. https://doi.org/10.1111/bjd.13262

72. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. *Med Decis Making* 2011;**31**:800–4. https://doi.org/10.1177/0272989X11401031

73. McDaniel B, Badri T, Steele RB. Basal cell carcinoma. In *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2023.

74. Thomson J, Hogan S, Leonardi-Bee J, Williams HC, Bath-Hextall FJ. Interventions for basal cell carcinoma of the skin. *Cochrane Database Syst Rev* 2020;**11**:CD003412. https://doi.org/10.1002/14651858.CD003412.pub3

75. *Basal Cell Carcinoma – An Overview*. The Primary Care Dermatology Society; 2023. URL: www.pcds.org.uk/clinical-guidance/basal-cell-carcinoma-an-overview (accessed 20 December 2023).

76. *Basal Cell Carcinoma PIL*. British Association of Dermatologists; 2022. URL: www.bad.org.uk/pils/basal-cell-carcinoma/ (accessed 20 December 2023).

77. Fijałkowska M, Bonczar M, Jastrzębski I, Ostrowski P, Antoszewski B, Koziej M. Growth rate of basal cell carcinoma: a meta-analysis and systematic review. *Postepy Dermatol Alergol* 2023;**40**:220–4. https://doi.org/10.5114/ada.2023.124795

78. National Institute for Health and Care Excellence. *Nivolumab for Treating Advanced (Unresectable or Metastatic) Melanoma [TA384]*. NICE; 2016. URL: www.nice.org.uk/guidance/ta384/ (accessed 20 December 2023).

79. Hernández Alava M, Pudney S, Wailoo A. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from a UK population study. *PharmacoEconomics* 2023;**41**:199–207. https://doi.org/10.1007/s40273-022-01218-7

80. Hernández Alava M, Pudney S, Wailoo A. *Estimating EQ-5D by Age and Sex for the UK*. London: NICE; Decision Support Unit; 2022.

81. Jones KC, Weatherly H, Birch S, Castelli A, Chalkley M, Dargan A, *et al*. *Unit Costs of Health and Social Care 2022 Manual*. Kent: Personal Social Services Research Unit (University of Kent) & Centre for Health Economics (University of York); 2023.

82. NHS Prescription Services. *Drug Tariff*. NHS Business Services Authority; 2023. URL: www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/drug-tariff (accessed 20 December 2023).

83. National Institute for Health and Care Excellence. *Olaparib for Maintenance Treatment of BRCA Mutation-positive Advanced Ovarian, Fallopian Tube or Peritoneal Cancer After Response to First-line Platinum-based Chemotherapy (review of TA598) [ID6191]*. URL: www.nice.org.uk/guidance/indevelopment/gid-ta11315 (accessed 10 March 2025).

84. Daneshjou R, Barata C, Betz-Stablein B, Celebi ME, Codella N, Combalia M, *et al*. Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR derm consensus guidelines from the international skin imaging collaboration artificial intelligence working group. *JAMA Dermatol* 2022;**158**:90–6. https://doi.org/10.1001/jamadermatol.2021.4915

Appendix 1 Literature search strategies

Named technology searches

MEDLINE(R) ALL

via Ovid http://ovidsp.ovid.com/

1946–26 October 2023

Date searched: 27 October 2023

Records retrieved: 65

1 "Deep Ensemble for the Recognition of Malignancy".af. (1)
2 (DERM and (Algorithm$ or Artificial Intelligen$ or AI)).tw. (12)
3 "Melanoma Image Analysis Algorithm".af. (0)
4 (Skin Analytics$ or SkinAnalytics$).af. (6)
5 (Moleanalyzer$ or Mole analyzer$ or Moleanalyser$ or Mole analyser$ or FotoFinder$).af. (63)
6 1 or 2 or 3 or 4 or 5 (79)
7 exp animals/ not humans.sh. (5163374)
8 6 not 7 (77)
9 limit 8 to yr="2015 -Current" (65)

EMBASE

via Ovid http://ovidsp.ovid.com/

1974–26 October 2023

Date searched: 27 October 2023

Records retrieved: 398 (NB – date limit 2015 onwards applied in EndNote)

1 "Deep Ensemble for the Recognition of Malignancy".af. (1)
2 DERM.dv. (114)
3 (DERM and (Algorithm$ or Artificial Intelligen$ or AI)).mp. (22)
4 "Melanoma Image Analysis Algorithm".af. (0)
5 (Skin Analytics$ or SkinAnalytics$).af. (8)
6 (Moleanalyzer$ or Mole analyzer$ or Moleanalyser$ or Mole analyser$ or FotoFinder$).af. (273)
7 or/1-6 (415)
8 Nonhuman/ not Human/ (5308649)
9 7 not 8 (398)

Cochrane Central Register of Controlled Trials (CENTRAL)

via Wiley http://onlinelibrary.wiley.com/

Issue 10 of 12, October 2023

Date searched: 27 October 2023

Records retrieved: 19

#1 "Deep Ensemble for the Recognition of Malignancy"0
#2 DERM653
#3 (Algorithm* or Artificial Intelligen* or AI)29320
#4 #2 and #3 21
#5 "Melanoma Image Analysis Algorithm"0
#6 (Skin next Analytics* or SkinAnalytics*)1
#7 (Moleanalyzer* or Mole next analyzer* or Moleanalyser* or Mole next analyser* or FotoFinder*)20
#8 #1 or #4 or #5 or #6 or #7 with Publication Year from 2015 to 2023, in Trials 19

ACM Digital Library

https://dl.acm.org/

Date searched: 27 October 2023

Records retrieved: 128 records

Search of The ACM Guide to Computing Literature

1. 35 Results for: [All: "deep ensemble for the recognition of malignancy"] OR [All: "melanoma image analysis algorithm"] OR [All: "skin analytics"] OR [All: "skin-analytics"] OR [All: "skinanalytics"] OR [All: moleanalyzer*] OR [All: "mole-analyzer"] OR [All: "mole analyzer"] OR [All: moleanalyser*] OR [All: "mole analyser"] OR [All: "mole-analyser"] OR [All: fotofinder*] AND [E-Publication Date: (01/01/2015 TO 12/31/2023)]
2. 93 Results for: [All: derm] AND [[All: algorithm* or] OR [All: "artificial intelligence"] OR [All: or ai]] AND [E-Publication Date: (01/01/2015 TO 12/31/2023)]

ClinicalTrials.gov

https://classic.clinicaltrials.gov/ct2/home

Date searched: 27 October 2023

Records retrieved: 50

Basic search screen used unless otherwise stated with terms entered into the 'other terms' search box.

1. 3 Studies found for: "Deep Ensemble for the Recognition of Malignancy"
2. 1 Study found for: "Melanoma Image Analysis Algorithm" OR "Melanoma Image Analysis Algorithms"
3. 21 Studies found for: "DERM" AND (Algorithm OR algorithms OR "Artificial Intelligence" OR AI)
4. 4 Studies found for: "Skin Analytics" OR "Skin-Analytics" OR "SkinAnalytics"
5. 4 Studies found for: "Skin Analytics" OR "Skin-Analytics" OR "SkinAnalytics" in sponsor field in advanced search screen
6. 2 Studies found for: Moleanalyzer OR "Mole-analyzer" OR "Mole analyzer" OR Moleanalyser OR "Mole analyser" OR "Mole-analyser"
7. 14 Studies found for: FotoFinder
8. 1 Study found for: FotoFinder in sponsor field in advanced search screen

WHO International Clinical Trials Registry Platform (WHO ICTRP)

https://trialsearch.who.int/

Date searched: 27 October 2023

Records retrieved: 37

1. Basic search screen:
   11 records for 11 trials found for: "Deep Ensemble for the Recognition of Malignancy" OR "Melanoma Image Analysis Algorithm" OR "Melanoma Image Analysis Algorithm" OR "Skin Analytics" OR "Skin-Analytics" OR "SkinAnalytics" OR Moleanalyzer* OR "Mole-analyzer" OR "Mole analyzer" OR Moleanalyser* OR "Mole analyser" OR "Mole-analyser" OR FotoFinder*
2. Basic search screen:
   4 records for 4 trials found for: "DERM" AND (Algorithm OR algorithms OR "Artificial Intelligence" OR AI)
3. Advanced search screen, recruitment status set to all:
   2 records for 2 trials found: Tile field - "Deep Ensemble for the Recognition of Malignancy" OR "Melanoma Image Analysis Algorithm" OR "Melanoma Image Analysis Algorithm" OR "Skin Analytics" OR "Skin-Analytics" OR "SkinAnalytics" OR Moleanalyzer* OR "Mole-analyzer" OR "Mole analyzer" OR Moleanalyser* OR "Mole analyser" OR "Mole-analyser" OR FotoFinder*
4. Advanced search screen, recruitment status set to all:
   10 records for 10 trials found: Intervention field - "Deep Ensemble for the Recognition of Malignancy" OR "Melanoma Image Analysis Algorithm" OR "Melanoma Image Analysis Algorithm" OR "Skin Analytics" OR "Skin-Analytics" OR "SkinAnalytics" OR Moleanalyzer* OR "Mole-analyzer" OR "Mole analyzer" OR Moleanalyser* OR "Mole analyser" OR "Mole-analyser" OR FotoFinder*
5. Advanced search screen, recruitment status set to all:
   5 records for 5 trials found: Primary Sponsor field - "Skin Analytics" OR "Skin-Analytics" OR SkinAnalytics* OR FotoFinder*
6. Advanced search screen, recruitment status set to all:
   4 records for 4 trials found: Title field -"DERM" AND (Algorithm OR algorithms OR "Artificial Intelligence" OR AI)
7. Advanced search screen, recruitment status set to all:
   1 records for 1 trials found: Intervention field - "DERM" AND (Algorithm OR algorithms OR "Artificial Intelligence" OR AI)

AI and dermoscopy search strategies

MEDLINE(R) ALL

via Ovid http://ovidsp.ovid.com/

1946–30 October 2023

Date searched: 31 October 2023

Records retrieved: 676

1 exp Skin Neoplasms/ (144404)
2 melanoma/ or hutchinson's melanotic freckle/ or melanoma, amelanotic/ (99075)
3 exp Carcinoma, Basal Cell/ (19781)
4 Carcinoma, Squamous Cell/ (141659)
5 Bowen's Disease/ (2003)
6 Carcinoma, Merkel Cell/ (3172)
7 Carcinoma, Neuroendocrine/ (5888)

8 exp Nevus/ (17238)
9 (skin adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (51808)
10 (cutaneous adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (18524)
11 melanoma$.ti,ab. (138894)
12 (nonmelanoma$ or non-melanoma$ or NMSC).ti,ab. (7225)
13 (melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab. (75689)
14 ((melanotic or malignan$ or Hutchinson$) adj2 freckle$).ti,ab. (66)
15 (lentigo$ adj2 maligna$).ti,ab. (1363)
16 ((basal adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)) or basalioma$ or BCC).ti,ab. (30190)
17 ((squamous cell adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)) or SCC or cSCC).ti,ab. (137121)
18 (Bowen$ adj3 (disease$ or lesion$ or cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (2466)
19 (Merkel cell adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (4168)
20 ((intra-epiderm$ or intraepiderm$ or intra-derm$ or intraderm$) adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (864)
21 ((neuroendocrine or neuro-endocrine) adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (28719)
22 ((skin or cutaneous or pigmented or nonpigmented) adj3 (lesion$ or nodul$ or macule$)).ti,ab. (59252)
23 (mole$1 or nevus or nevi or naevus or naevi).ti,ab. (43265)
24 ((acitinic or solar or senile) adj2 kerato$).ti,ab. (535)
25 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 (617501)
26 Artificial Intelligence/ (40908)
27 algorithms/ (306306)
28 exp Machine Learning/ (61112)
29 exp neural networks, computer/ (61671)
30 ((artificial$ or machine$ or computational$) adj2 intelligen$).ti,ab. (35799)
31 computer vision.ti,ab. (7427)
32 (AI or AIDHT or AIaMD).ti,ab. (47517)
33 (augment$ adj2 intelligen$).ti,ab. (209)
34 algorithm$.ti,ab. (366699)
35 deep learning.ti,ab. (46831)
36 machine learning.ti,ab. (85994)
37 ((supervised or unsupervised or semi-supervised) adj2 learning).ti,ab. (11807)
38 ((neural or convolutional) adj2 network$).ti,ab. (100150)
39 (CNN or CNNs or DCNN or DCNNs).ti,ab. (18024)
40 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 (727630)
41 25 and 40 (8447)
42 Dermoscopy/ (5910)
43 (dermoscop$ or dermascop$ or dermatoscop$).ti,ab. (7658)
44 ((skin or cutaneous or epidermis) adj3 (microscopy or microscopies)).ti,ab. (1062)
45 (epiluminescen$ adj3 (microscopy or microscopies)).ti,ab. (229)
46 (teledermoscop$ or teledermascop$ or teledermatoscop$).ti,ab. (150)
47 (videodermoscop$ or videodermascop$ or videodermatoscop$).ti,ab. (188)
48 (Dermlite Handyscope$ or “Medicam 1000”).ti,ab. (4)
49 (teledermatolog$ or tele-dermatolog$).ti,ab. (1283)
50 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 (11821)

51 41 and 50 (987)
52 exp animals/ not humans.sh. (5164446)
53 51 not 52 (984)
54 limit 53 to yr="2015 -Current" (676)

EMBASE

via Ovid http://ovidsp.ovid.com/

1974 –30 October 2023

Date searched: 31 October 2023

Records retrieved: 1035

1 exp skin tumor/ (242335)
2 exp "nevi and melanomas"/ (217266)
3 (skin adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (71175)
4 (cutaneous adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (24674)
5 melanoma$.ti,ab. (195684)
6 (nonmelanoma$ or non-melanoma$ or NMSC).ti,ab. (11666)
7 (melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab. (102721)
8 ((melanotic or malignan$ or Hutchinson$) adj2 freckle$).ti,ab. (73)
9 (lentigo$ adj2 maligna$).ti,ab. (1951)
10 ((basal adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)) or basalioma$ or BCC).ti,ab. (39985)
11 ((squamous cell adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)) or SCC or cSCC).ti,ab. (190640)
12 (Merkel cell adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (6086)
13 (Bowen$ adj2 (disease$ or lesion$ or cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (3026)
14 ((intra-epiderm$ or intraepiderm$ or intra-derm$ or intraderm$) adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (1092)
15 neuroendocrine carcinoma/ (4182)
16 ((neuroendocrine or neuro-endocrine) adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (48955)
17 skin defect/ (66760)
18 ((skin or cutaneous or pigmented or nonpigmented) adj3 (lesion$ or nodul$ or macule$)).ti,ab. (87454)
19 (mole$1 or nevus or nevi or naevus or naevi).ti,ab. (50346)
20 ((acitinic or solar or senile) adj2 kerato$).ti,ab. (623)
21 or/1-20 (870026)
22 exp artificial intelligence/ (88413)
23 exp algorithm/ (594577)
24 exp machine learning/ (425179)
25 convolutional neural network/ (26698)
26 ((artificial$ or machine$ or computational$) adj2 intelligen$).ti,ab. (42771)
27 computer vision.ti,ab. (7980)
28 (AI or AIDHT or AIaMD).ti,ab. (63751)

29 (augment$ adj2 intelligen$).ti,ab. (292)
30 algorithm$.ti,ab. (463852)
31 deep learning.ti,ab. (54317)
32 machine learning.ti,ab. (101443)
33 ((supervised or unsupervised or semi-supervised) adj2 learning).ti,ab. (13757)
34 ((neural or convolutional) adj2 network$).ti,ab. (117758)
35 (CNN or CNNs or DCNN or DCNNs).ti,ab. (21381)
36 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 (1085732)
37 21 and 36 (18333)
38 exp epiluminescence microscopy/ (14216)
39 (dermoscop$ or dermascop$ or dermatoscop$).ti,ab,mv,my. (10281)
40 ((skin or cutaneous or epidermis) adj3 (microscopy or microscopies)).ti,ab,mv,my. (1437)
41 (epiluminescen$ adj3 (microscopy or microscopies)).ti,ab,mv,my. (282)
42 (teledermoscop$ or teledermascop$ or teledermatoscop$).ti,ab,mv,my. (203)
43 (videodermoscop$ or videodermascop$ or videodermatoscop$).ti,ab,mv,my. (256)
44 (Dermlite Handyscope$ or "Medicam 1000").ti,ab. (2)
45 teledermatology/ (1803)
46 (teledermatolog$ or tele-dermatolog$).ti,ab. (1887)
47 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 (19385)
48 37 and 47 (1393)
49 limit 48 to yr="2015 -Current" (1035)

Cochrane Central Register of Controlled Trials (CENTRAL)

via Wiley http://onlinelibrary.wiley.com/

Issue 10 of 12, October 2023

Date searched: 31 October 2023

Records retrieved: 10

IDSearchHits

#1 MeSH descriptor: [Skin Neoplasms] explode all trees2152
#2 MeSH descriptor: [Melanoma] this term only2742
#3 MeSH descriptor: [Hutchinson's Melanotic Freckle] this term only14
#4 MeSH descriptor: [Melanoma, Amelanotic] this term only2
#5 MeSH descriptor: [Carcinoma, Basal Cell] explode all trees451
#6 MeSH descriptor: [Carcinoma, Squamous Cell] this term only3422
#7 MeSH descriptor: [Bowen's Disease] this term only41
#8 MeSH descriptor: [Carcinoma, Merkel Cell] this term only34
#9 MeSH descriptor: [Carcinoma, Neuroendocrine] this term only80
#10 MeSH descriptor: [Nevus] explode all trees104
#11 (skin near/3 (cancer* or carcinoma* or tumour* or tumor* or neoplas* or oncolog* or adenoma* or adenocarcinoma* or epithel* or malignan* or premalignan* or precancer*)):ti,ab,kw4751
#12 (cutaneous near/3 (cancer* or carcinoma* or tumour* or tumor* or neoplas* or oncolog* or adenoma* or adenocarcinoma* or epithel* or malignan* or premalignan* or precancer*)):ti,ab,kw454
#13 melanoma*:ti,ab,kw6573
#14 (nonmelanoma* or non-melanoma* or NMSC):ti,ab,kw844
#15 (melanocyt* or non-melanocyt* or nonmelanocyt* or keratinocyt*):ti,ab,kw1465
#16 ((melanotic or malignan* or Hutchinson*) near/2 freckle*):ti,ab,kw17

#17 (lentigo* near/2 maligna*):ti,ab,kw49
#18 ((basal near/3 (cancer* or carcinoma* or tumour* or tumor* or neoplas* or oncolog* or adenoma* or adenocarcinoma* or epithel* or malignan* or premalignan* or precancer*)) or basalioma* or BCC):ti,ab,kw1585
#19 (("squamous cell" near/2 (cancer* or carcinoma* or tumour* or tumor* or neoplas* or oncolog* or adenoma* or adenocarcinoma* or epithel* or malignan* or premalignan* or precancer*)) or SCC or cSCC):ti,ab,kw10125
#20 (Bowen* near/3 (disease* or lesion* or cancer* or carcinoma* or tumour* or tumor* or neoplas* or oncolog* or adenoma* or adenocarcinoma* or epithel* or malignan* or premalignan* or precancer*)):ti,ab,kw115
#21 ("Merkel cell" near/2 (cancer* or carcinoma* or tumour* or tumor* or neoplas* or oncolog* or adenoma* or adenocarcinoma* or epithel* or malignan* or premalignan* or precancer*)):ti,ab,kw108
#22 ((intra-epiderm* or intraepiderm* or intra-derm* or intraderm*) near/3 (cancer* or carcinoma* or tumour* or tumor* or neoplas* or oncolog* or adenoma* or adenocarcinoma* or epithel* or malignan* or premalignan* or precancer*)):ti,ab,kw46
#23 ((neuroendocrine or neuro-endocrine) near/2 (cancer* or carcinoma* or tumour* or tumor* or neoplas* or oncolog* or adenoma* or adenocarcinoma* or epithel* or malignan* or premalignan* or precancer*)):ti,ab,kw1248
#24 ((skin or cutaneous or pigmented or nonpigmented) near/3 (lesion* or nodul* or macule*)):ti,ab,kw2762
#25 (mole or moles or nevus or nevi or naevus or naevi):ti,ab,kw684
#26 ((acitinic or solar or senile) near/2 kerato*):ti,ab,kw38
#27 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #2624686
#28 MeSH descriptor: [Artificial Intelligence] this term only554
#29 MeSH descriptor: [Algorithms] this term only4515
#30 MeSH descriptor: [Machine Learning] explode all trees931
#31 MeSH descriptor: [Neural Networks, Computer] explode all trees540
#32 ((artificial* or machine* or computational*) near/2 intelligen*):ti,ab,kw1756
#33 "computer vision":ti,ab,kw140
#34 (AI or AIDHT or AIaMD):ti,ab,kw5476
#35 (augment* near/2 intelligen*):ti,ab,kw13
#36 algorithm*:ti,ab,kw17728
#37 "deep learning":ti,ab,kw1037
#38 "machine learning":ti,ab,kw2501
#39 ((supervised or unsupervised or semi-supervised) near/2 learning):ti,ab,kw207
#40 ((neural or convolutional) near/2 network*):ti,ab,kw1888
#41 (CNN or CNNs or DCNN or DCNNs):ti,ab,kw320
#42 #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #4126379
#43 #27 and #42356
#44 MeSH descriptor: [Dermoscopy] this term only103
#45 (dermoscop* or dermascop* or dermatoscop*):ti,ab,kw473
#46 ((skin or cutaneous or epidermis) near/3 (microscopy or microscopies)):ti,ab,kw78
#47 (epiluminescen* near/3 (microscopy or microscopies)):ti,ab,kw161
#48 (teledermoscop* or teledermascop* or teledermatoscop*):ti,ab,kw26
#49 (videodermoscop* or videodermascop* or videodermatoscop*):ti,ab,kw14
#50 (Dermlite next Handyscope* or "Medicam 1000"):ti,ab,kw0
#51 (teledermatolog* or tele-dermatolog*):ti,ab,kw110
#52 #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51660
#53 #43 and #52 with Publication Year from 2015 to 2023, in Trials10

ACM Digital Library

https://dl.acm.org/

Date searched: 31 October 2023

Records retrieved: 424 records

Search of the The ACM Guide to Computing Literature using advanced search interface.

1. 20 Results for: [[Title: skin] OR [Title: cutaneous] OR [Title: pigmented] OR [Title: nonpigmented] OR [Title: freckle*] OR [Title: lentigo*] OR [Title: basal] OR [Title: "squamous cell"] OR [Title: bowen*] OR [Title: "merkel cell"] OR [Title: intra-epiderm*] OR [Title: intraepiderm*] OR [Title: intra-derm*] OR [Title: intraderm*] OR [Title: neuroendocrine] OR [Title: neuro-endocrine]] AND [[Title: cancer*] OR [Title: carcinoma*] OR [Title: tumour*] OR [Title: tumor*] OR [Title: neoplas*] OR [Title: oncolog*] OR [Title: adenoma*] OR [Title: adenocarcinoma*] OR [Title: epithel*] OR [Title: maligna*] OR [Title: melanotic] OR [Title: premalignan*] OR [Title: precancer*] OR [Title: lesion*] OR [Title: nodul*] OR [Title: macule*]] AND [[Title: "artificial intelligence"] OR [Title: "machine intelligence"] OR [Title: "computational intelligence"] OR [Title: "computer vision"] OR [Title: ai] OR [Title: ai-dht] OR [Title: aidht] OR [Title: aiamd] OR [Title: "augmented intelligence"] OR [Title: algorithm*] OR [Title: "deep learning"] OR [Title: "machine learning"] OR [Title: "supervised learning"] OR [Title: "unsupervised learning"] OR [Title: "semi-supervised learning"] OR [Title: "neural network"] OR [Title: "neural networks"] OR [Title: "neural networking"] OR [Title: convolutional] OR [Title: cnn] OR [Title: cnns] OR [Title: dcnn] OR [Title: dcnns]] AND [[Title: dermoscop*] OR [Title: dermascop*] OR [Title: dermatoscop*] OR [Title: teledermoscop*] OR [Title: teledermascop*] OR [Title: teledermatoscop*] OR [Title: videodermoscop* or videodermascop* or videodermatoscop*] OR [Title: teledermatolog*] OR [Title: tele-dermatolog*] OR [Title: microscopy] OR [Title: microscopies] OR [Title: epiluminescen*] OR [Title: handyscope*] OR [Title: "medicam 1000"]] AND [E-Publication Date: (01/01/2015 TO 12/31/2023)]
2. 218 Results for: [[Abstract: skin] OR [Abstract: cutaneous] OR [Abstract: pigmented] OR [Abstract: nonpigmented] OR [Abstract: freckle*] OR [Abstract: lentigo*] OR [Abstract: basal] OR [Abstract: "squamous cell"] OR [Abstract: bowen*] OR [Abstract: "merkel cell"] OR [Abstract: intra-epiderm*] OR [Abstract: intraepiderm*] OR [Abstract: intra-derm*] OR [Abstract: intraderm*] OR [Abstract: neuroendocrine] OR [Abstract: neuro-endocrine]] AND [[Abstract: cancer*] OR [Abstract: carcinoma*] OR [Abstract: tumour*] OR [Abstract: tumor*] OR [Abstract: neoplas*] OR [Abstract: oncolog*] OR [Abstract: adenoma*] OR [Abstract: adenocarcinoma*] OR [Abstract: epithel*] OR [Abstract: maligna*] OR [Abstract: melanotic] OR [Abstract: premalignan*] OR [Abstract: precancer*] OR [Abstract: lesion*] OR [Abstract: nodul*] OR [Abstract: macule*]] AND [[Abstract: "artificial intelligence"] OR [Abstract: "machine intelligence"] OR [Abstract: "computational intelligence"] OR [Abstract: "computer vision"] OR [Abstract: ai] OR [Abstract: ai-dht] OR [Abstract: aidht] OR [Abstract: aiamd] OR [Abstract: "augmented intelligence"] OR [Abstract: algorithm*] OR [Abstract: "deep learning"] OR [Abstract: "machine learning"] OR [Abstract: "supervised learning"] OR [Abstract: "unsupervised learning"] OR [Abstract: "semi-supervised learning"] OR [Abstract: "neural network"] OR [Abstract: "neural networks"] OR [Abstract: "neural networking"] OR [Abstract: convolutional] OR [Abstract: cnn] OR [Abstract: cnns] OR [Abstract: dcnn] OR [Abstract: dcnns]] AND [[Abstract: dermoscop*] OR [Abstract: dermascop*] OR [Abstract: dermatoscop*] OR [Abstract: teledermoscop*] OR [Abstract: teledermascop*] OR [Abstract: teledermatoscop*] OR [Abstract: videodermoscop*] OR [Abstract: videodermascop*] OR [Abstract: videodermatoscop*] OR [Abstract: teledermatolog*] OR [Abstract: tele-dermatolog*] OR [Abstract: "epiluminescence microscopy"] OR [Abstract: "epiluminescence microscopies"] OR [Abstract: "dermlite handyscope"] OR [Abstract: "medicam 1000"]] AND [E-Publication Date: (01/01/2015 TO 12/31/2023)]
3. 11 Results for: [[Title: melanoma*] OR [Title: nonmelanoma*] OR [Title: non-melanoma*] OR [Title: nmsc] OR [Title: melanocyt*] OR [Title: non-melanocyt*] OR [Title: nonmelanocyt*] OR [Title: keratinocyt*] OR [Title: mole] OR [Title: moles] OR [Title: nevus] OR [Title: nevi] OR [Title: naevus] OR [Title: naevi] OR [Title: basalioma*] OR [Title: bcc or scc or cscc] OR [Title: "hutchinson freckle"] OR [Title: "hutchinson's freckle"] OR [Title: "solar keratosis"] OR [Title: "solar keratoses"] OR [Title: "acitinic keratosis"] OR [Title: "acitinic keratoses"] OR [Title: "senile keratosis"] OR [Title: "senile keratoses"]] AND [[Title: "artificial intelligence"] OR [Title: "machine intelligence"] OR [Title: "computational intelligence"] OR [Title: "computer vision"] OR [Title: ai] OR [Title: ai-dht] OR [Title: aidht] OR [Title: aiamd] OR [Title: "augmented intelligence"] OR [Title: algorithm*] OR [Title: "deep learning"] OR [Title: "machine learning"] OR [Title: "supervised learning"] OR [Title: "unsupervised learning"] OR [Title: "semi-supervised learning"] OR [Title: "neural network"] OR [Title: "neural networks"] OR [Title: "neural networking"] OR [Title: convolutional] OR [Title: cnn] OR [Title: cnns] OR [Title: dcnn] OR [Title: dcnns]] AND [[Title: dermoscop*] OR [Title: dermascop*] OR [Title: dermatoscop*] OR [Title: teledermoscop*] OR [Title: teledermascop*] OR [Title: teledermatoscop*] OR [Title: videodermoscop* or videodermascop* or videodermatoscop*] OR [Title: teledermatolog*] OR [Title: tele-dermatolog*] OR [Title: "epiluminescence microscopy"] OR [Title: "epiluminescence microscopies"] OR [Title: "dermlite handyscope"] OR [Title: "medicam 1000"]] AND [E-Publication Date: (01/01/2015 TO 12/31/2023)]

4. 175 Results for: [[Abstract: melanoma*] OR [Abstract: nonmelanoma*] OR [Abstract: non-melanoma*] OR [Abstract: nmsc] OR [Abstract: melanocyt*] OR [Abstract: non-melanocyt*] OR [Abstract: nonmelanocyt*] OR [Abstract: keratinocyt*] OR [Abstract: mole] OR [Abstract: moles] OR [Abstract: nevus] OR [Abstract: nevi] OR [Abstract: naevus] OR [Abstract: naevi] OR [Abstract: basalioma*] OR [Abstract: bcc or scc or cscc] OR [Abstract: "hutchinson freckle"] OR [Abstract: "hutchinson's freckle"] OR [Abstract: "solar keratosis"] OR [Abstract: "solar keratoses"] OR [Abstract: "acitinic keratosis"] OR [Abstract: "acitinic keratoses"] OR [Abstract: "senile keratosis"] OR [Abstract: "senile keratoses"]] AND [[Abstract: "artificial intelligence"] OR [Abstract: "machine intelligence"] OR [Abstract: "computational intelligence"] OR [Abstract: "computer vision"] OR [Abstract: ai] OR [Abstract: ai-dht] OR [Abstract: aidht] OR [Abstract: aiamd] OR [Abstract: "augmented intelligence"] OR [Abstract: algorithm*] OR [Abstract: "deep learning"] OR [Abstract: "machine learning"] OR [Abstract: "supervised learning"] OR [Abstract: "unsupervised learning"] OR [Abstract: "semi-supervised learning"] OR [Abstract: "neural network"] OR [Abstract: "neural networks"] OR [Abstract: "neural networking"] OR [Abstract: convolutional] OR [Abstract: cnn] OR [Abstract: cnns] OR [Abstract: dcnn] OR [Abstract: dcnns]] AND [[Abstract: dermoscop*] OR [Abstract: dermascop*] OR [Abstract: dermatoscop*] OR [Abstract: teledermoscop*] OR [Abstract: teledermascop*] OR [Abstract: teledermatoscop*] OR [Abstract: videodermoscop* or videodermascop* or videodermatoscop*] OR [Abstract: telederma-tolog*] OR [Abstract: tele-dermatolog*] OR [Abstract: "epiluminescence microscopy"] OR [Abstract: "epiluminescence microscopies"] OR [Abstract: "dermlite handyscope"] OR [Abstract: "medicam 1000"]] AND [E-Publication Date: (01/01/2015 TO 12/31/2023)]

ClinicalTrials.gov

https://classic.clinicaltrials.gov/ct2/

Date searched: 2 November 2023

Records retrieved: 270

1. 30 Studies found for: "artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR "computer vision" OR AI OR "augmented intelligence" OR algorithm | "skin cancer" OR "skin neoplasm" OR "skin tumor" OR "skin tumour" OR "skin carcinoma"
2. 7 Studies found for: "deep learning" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "semi-supervised learning" OR "neural network" OR "neural networks" OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs | "skin cancer" OR "skin neoplasm" OR "skin tumor" OR "skin tumour" OR "skin carcinoma"
3. 29 Studies found for: "artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR "computer vision" OR AI OR "augmented intelligence" OR algorithm | "cutaneous cancer" OR "cutaneous neoplasm" OR "cutaneous tumor" OR "cutaneous tumour" OR "cutaneous carcinoma"
4. 6 Studies found for: "deep learning" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "semi-supervised learning" OR "neural network" OR 'neural networks" OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs | "cutaneous cancer" OR "cutaneous neoplasm" OR "cutaneous tumor" OR "cutaneous tumour" OR "cutaneous carcinoma"
5. 55 Studies found for: "artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR "computer vision" OR AI OR "augmented intelligence" OR algorithm | melanoma OR nonmelanoma OR non-melanoma OR melanocytic OR non-melanocytic OR nonmelanocytic OR keratinocytic OR melanocyte OR non-melanocyte OR nonmelanocyte OR keratinocyte OR melanotic OR "lentigo maligna"
6. 11 Studies found for: "deep learning" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "semi-supervised learning" OR "neural network" OR "neural networks" OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs | melanoma OR nonmelanoma OR non-melanoma OR melanocytic OR non-melanocytic OR nonmelanocytic OR keratinocytic OR melanocyte OR non-melanocyte OR nonmelanocyte OR keratinocyte OR melanotic OR "lentigo maligna"
7. 38 Studies found for: "artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR "computer vision" OR AI OR "augmented intelligence" OR algorithm | "Basal cell cancer" OR "Basal cell neoplasm" OR "Basal cell tumor" OR "Basal cell tumour" OR "Basal cell carcinoma" OR "Squamous cell cancer" OR "Squamous cell neoplasm" OR "Squamous cell tumor" OR "Squamous cell tumour" OR "Squamous cell carcinoma"

8. 13 Studies found for: “deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs | “Basal cell cancer” OR “Basal cell neoplasm” OR “Basal cell tumor” OR “Basal cell tumour” OR “Basal cell carcinoma” OR “Squamous cell cancer” OR “Squamous cell neoplasm” OR “Squamous cell tumor” OR “Squamous cell tumour” OR “Squamous cell carcinoma”
9. 23 Studies found for: “artificial intelligence” OR “machine intelligence” OR “computational intelligence” OR “computer vision” OR AI OR “augmented intelligence” OR algorithm | (skin OR cutaneous OR dermatology OR dermal OR dermis OR epidermal OR epidermis) AND (“Neuroendocrine cancer” OR “Neuroendocrine neoplasm” OR “Neuroendocrine tumor” OR “Neuroendocrine tumour” OR “Neuroendocrine carcinoma”)
10. 5 Studies found for: “deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs | (skin OR cutaneous OR dermatology OR dermal OR dermis OR epidermal OR epidermis) AND (“Neuroendocrine cancer” OR “Neuroendocrine neoplasm” OR “Neuroendocrine tumor” OR “Neuroendocrine tumour” OR “Neuroendocrine carcinoma”)
11. 45 Studies found for: “artificial intelligence” OR “machine intelligence” OR “computational intelligence” OR “computer vision” OR AI OR “augmented intelligence” OR algorithm | Bowen OR Bowens OR Bowen’s OR “Merkel cell” OR mole OR moles OR nevus OR nevi OR naevus OR naevi OR “senile keratosis” OR “acitinic keratosis” OR “solar keratosis”
12. 8 Studies found for: “deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs | Bowen OR Bowens OR Bowen’s OR “Merkel cell” OR mole OR moles OR nevus OR nevi OR naevus OR naevi OR “senile keratosis” OR “acitinic keratosis” OR “solar keratosis”

WHO International Clinical Trials Registry Platform (WHO ICTRP)

https://trialsearch.who.int/

Date searched: 2 November 2023

Records retrieved: 177

1. Advanced search screen, recruitment status set to all:
   Condition field: “skin cancer” OR “skin neoplasm” OR “skin tumor” OR “skin tumour” OR “skin carcinoma” Intervention field: “artificial intelligence” OR “machine intelligence” OR “computational intelligence” OR “computer vision” OR AI OR “augmented intelligence” OR algorithm*
   8 records for 8 trials found
2. Advanced search screen, recruitment status set to all:
   Condition field: “skin cancer” OR “skin neoplasm” OR “skin tumor” OR “skin tumour” OR “skin carcinoma” Intervention field: “deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs
   0 records for 0 trials found
3. Advanced search screen, recruitment status set to all:
   Condition field: “cutaneous cancer” OR “cutaneous neoplasm” OR “cutaneous tumor” OR “cutaneous tumour” OR “cutaneous carcinoma” Intervention field: “artificial intelligence” OR “machine intelligence” OR “computational intelligence” OR “computer vision” OR AI OR “augmented intelligence” OR algorithm*
   0 records for 0 trials found
4. Advanced search screen, recruitment status set to all:
   Condition field: “cutaneous cancer” OR “cutaneous neoplasm” OR “cutaneous tumor” OR “cutaneous tumour” OR “cutaneous carcinoma” Intervention field: “deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs
   0 records for 0 trials found

5. Advanced search screen, recruitment status set to all:
   Condition field: melanoma* OR nonmelanoma* OR non-melanoma* OR melanocyt* OR non-melanocyt* OR non-melanocyt* OR keratinocyt* OR "lentigo maligna" Intervention field: "artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR "computer vision" OR AI OR "augmented intelligence" OR algorithm*
   10 records for 10 trials found
6. Advanced search screen, recruitment status set to all:
   Condition field: melanoma* OR nonmelanoma* OR non-melanoma* OR melanocyt* OR non-melanocyt* OR non-melanocyt* OR keratinocyt* OR "lentigo maligna" Intervention field: "deep learning" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "semi-supervised learning" OR "neural network" OR "neural networks" OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs
   0 records for 0 trials found
7. Advanced search screen, recruitment status set to all:
   Condition field: "Basal cell cancer" OR "Basal cell neoplasm" OR "Basal cell tumor" OR "Basal cell tumour" OR "Basal cell carcinoma" OR "Squamous cell cancer" OR "Squamous cell neoplasm" OR "Squamous cell tumor" OR "Squamous cell tumour" OR "Squamous cell carcinoma" Intervention field: "artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR "computer vision" OR AI OR "augmented intelligence" OR algorithm*
   6 records for 6 trials found
8. Advanced search screen, recruitment status set to all:
   Condition field: "Basal cell cancer" OR "Basal cell neoplasm" OR "Basal cell tumor" OR "Basal cell tumour" OR "Basal cell carcinoma" OR "Squamous cell cancer" OR "Squamous cell neoplasm" OR "Squamous cell tumor" OR "Squamous cell tumour" OR "Squamous cell carcinoma" Intervention field: "deep learning" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "semi-supervised learning" OR "neural network" OR "neural networks" OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs
   0 records for 0 trials found
9. Advanced search screen, recruitment status set to all:
   Condition field: (skin OR cutaneous OR dermatology OR dermal OR dermis OR epidermal OR epidermis) AND ("Neuroendocrine cancer" OR "Neuroendocrine neoplasm" OR "Neuroendocrine tumor" OR "Neuroendocrine tumour" OR "Neuroendocrine carcinoma") Intervention field: "artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR "computer vision" OR AI OR "augmented intelligence" OR algorithm*
   0 records for 0 trials found
10. Advanced search screen, recruitment status set to all:
    Condition field: (skin OR cutaneous OR dermatology OR dermal OR dermis OR epidermal OR epidermis) AND ("Neuroendocrine cancer" OR "Neuroendocrine neoplasm" OR "Neuroendocrine tumor" OR "Neuroendocrine tumour" OR "Neuroendocrine carcinoma") " Intervention field: "deep learning" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "semi-supervised learning" OR "neural network" OR "neural networks" OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs
    0 records for 0 trials found
11. Advanced search screen, recruitment status set to all:
    Condition field: Bowen OR Bowens OR Bowen's OR "Merkel cell" OR mole OR moles OR nevus OR nevi OR naevus OR naevi OR "senile keratosis" OR "acitinic keratosis" OR "solar keratosis" Intervention field: "artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR "computer vision" OR AI OR "augmented intelligence" OR algorithm*
    0 records for 0 trials found
12. Advanced search screen, recruitment status set to all:
    Condition field: Bowen OR Bowens OR Bowen's OR "Merkel cell" OR mole OR moles OR nevus OR nevi OR naevus OR naevi OR "senile keratosis" OR "acitinic keratosis" OR "solar keratosis" Intervention field: "deep learning" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "semi-supervised learning" OR "neural network" OR "neural networks" OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs
    0 records for 0 trials found
13. Basic search screen:
    17 records for 17 trials found for: ("skin cancer" OR "skin neoplasm" OR "skin tumor" OR "skin tumour" OR "skin carcinoma") AND ("artificial intelligence" OR "machine intelligence" OR "computational intelligence" OR 'computer vision" OR AI OR "augmented intelligence" OR algorithm*)

14 1 trial found for: (“skin cancer” OR “skin neoplasm” OR “skin tumor” OR “skin tumour” OR “skin carcinoma”) AND (“deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs)
15. 44 records for 31 trials found for: (melanoma* OR nonmelanoma* OR non-melanoma* OR melanocyt* OR non-melanocyt* OR nonmelanocyt* OR keratinocyt* OR “lentigo maligna”) AND (“artificial intelligence” OR “machine intelligence” OR “computational intelligence” OR “computer vision” OR AI OR “augmented intelligence” OR algorithm*)
16. 1 trial found for: (melanoma* OR nonmelanoma* OR non-melanoma* OR melanocyt* OR non-melanocyt* OR nonmelanocyt* OR keratinocyt* OR “lentigo maligna”) AND (“deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs)
17. 25 records for 25 trials found for: (“Basal cell cancer” OR “Basal cell neoplasm” OR “Basal cell tumor” OR “Basal cell tumour” OR “Basal cell carcinoma” OR “Squamous cell cancer” OR “Squamous cell neoplasm” OR “Squamous cell tumor” OR “Squamous cell tumour” OR “Squamous cell carcinoma”) AND (“artificial intelligence” OR “machine intelligence” OR “computational intelligence” OR “computer vision” OR AI OR “augmented intelligence” OR algorithm*)
18. 8 records for 8 trials found for: (“Basal cell cancer” OR “Basal cell neoplasm” OR “Basal cell tumor” OR “Basal cell tumour” OR “Basal cell carcinoma” OR “Squamous cell cancer” OR “Squamous cell neoplasm” OR “Squamous cell tumor” OR “Squamous cell tumour” OR “Squamous cell carcinoma”) AND (“deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs)
19 9 records for 9 trials found for: (Bowen OR Bowens OR Bowen’s OR “Merkel cell” OR mole OR moles OR nevus OR nevi OR naevus OR naevi OR “senile keratosis” OR “acitinic keratosis” OR “solar keratosis”) AND (“artificial intelligence” OR “machine intelligence” OR “computational intelligence” OR “computer vision” OR AI OR “augmented intelligence” OR algorithm*)
20. 1 trial found for: (Bowen OR Bowens OR Bowen’s OR “Merkel cell” OR mole OR moles OR nevus OR nevi OR naevus OR naevi OR “senile keratosis” OR ‘acitinic keratosis” OR “solar keratosis”) AND (“deep learning” OR “machine learning” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “neural network” OR “neural networks” OR convolutional OR CNN OR CNNs OR DCNN OR DCNNs)

Appendix 2 Excluded studies at full-text screening stage

Exclude population (n = 1)

Corbin A, Marques O. Exploring strategies to generate Fitzpatrick skin type metadata for dermoscopic images using individual typology angle techniques. *Multimed Tools Appl* 2022;**82**:23771–95.

Exclude intervention (n = 39)

Abbes W, Sellami D. Deep neural networks for melanoma detection from optical standard images using transfer learning. *Procedia Comput Sci.* 2021;**192**:1304–12.

Abhishek K, Kawahara J, Hamarneh G. Predicting the clinical management of skin lesions using deep learning. *Sci Rep* 2021;**11**:7769.

Abouche H, Jimi A, Zrira N, Benmiloud I. *Segmentation and Classification of Dermoscopic Skin Cancer on Green Channel*. Proceedings of the 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2023:347–54.

Arlette J, Wong A, Khodadad I, Kazemzadeh F. Deep tissue sequencing using augmented intelligence to probe melanocytic lesions. *J Cutan Med Surg* 2017;**21**:572.

Assuta Hospital Systems. *Artificial Intelligence-assisted Evaluation of Pigmented Skin Lesions*. NCT03362138

Babino G, Lallas A, Agozzino M, Alfano R, Apalla Z, Brancaccio G, *et al.* Melanoma diagnosed on digital dermoscopy monitoring: a side-by-side image comparison is needed to improve early detection. *J Am Acad Dermatol* 2021;**85**:619–25.

Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, *et al.* Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer* 2019;**119**:11–7.

Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, *et al.* A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur J Cancer* 2019;**111**:148–54.

Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, *et al.* Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer* 2019;**113**:47–54.

Camacho-Gutierrez Jos A, Solorza-Calderon S, Ivarez-Borrego J. Multi-class skin lesion classification using prism- and segmentation-based fractal signatures. *Expert Syst Appl* 2022;**197**:116671.

Cerminara SE, Cheng P, Kostner L, Huber S, Kunz M, Maul JT, *et al.* Diagnostic performance of augmented intelligence with 2D and 3D total body photography and convolutional neural networks in a high-risk population for melanoma under real-world conditions: a new era of skin cancer screening? *Eur J Cancer* 2023;**190**:112954.

Cinotti E, Haouas M, Grivet D, Perrot JL. In vivo and ex vivo confocal microscopy for the management of a melanoma of the eyelid margin. *Dermatol Surg* 2015;**41**:1437–40.

Cinotti E, Santi F, Perrot JL, Habougit C, Tognetti L, Rubegni P. Squamous cell carcinoma arising on acrodermatitis continua of Hallopeau: clinical and noninvasive skin imaging features. *Int J Dermatol* 2021;**60**:763–5.

Del Rosario F, Farahi JM, Drendel J, Buntinx-Krieg T, Caravaglio J, Domozych R, *et al.* Performance of a computer-aided digital dermoscopic image analyzer for melanoma detection in 1,076 pigmented skin lesion biopsies. *J Am Acad Dermatol* 2018;**78**:927–34.e6.

Pérez E, Reyes Ó. Performing melanoma diagnosis by an effective multi-view convolutional network architecture. *Int J Comput Vision* 2023;**131**:3094–117

Pérez-Perdomo E, Ventura S. An ensemble-based convolutional neural network model powered by a genetic algorithm for melanoma diagnosis. *Neural Comput Appl* 2022;**34**:10429–48

Francese R, Frasca M, Risi M, Tortora G. A mobile augmented reality application for supporting real-time skin lesion analysis based on deep learning. *J Real Time Image Process* 2021;**18**:1247–59.

Gu R, Wang L, Zhang L. DE-Net: a deep edge network with boundary information for automatic skin lesion segmentation. *Neurocomputing* 2022;**468**:71–84.

Gupta A, Tiwari D, Agarwal S, Jain M. *Fuzzy Based Support System for Melanoma Diagnosis*. Proceedings of the Third International Conference on Mining Intelligence and Knowledge Exploration 2015;**9468**:235–46.

Haugsten ER, Vestergaard T, Trettin B. Experiences regarding use and implementation of artificial intelligence-supported follow-up of atypical moles at a dermatological outpatient clinic: qualitative study. *MIR Dermatol* 2023;**6**:e44913.

He X, Wang Y, Zhao S, Chen X. Co-attention fusion network for multimodal skin cancer diagnosis. *Pattern Recogn* 2023;**133**:108990.

Huang K, Jiang Z, Li Y, Wu Z, Wu X, Zhu W, *et al.* The classification of six common skin diseases based on Xiangya-Derm: development of a Chinese database for artificial intelligence. *J Med Internet Res* 2021;**23**:e26025.

Inthiyaz S, Altahan BR, Ahammad Sk H, Rajesh V, Kalangi Ruth R, Smirani Lassaad K, *et al.* Skin disease detection using deep learning. *Adv Eng Softw* 2023;**175**:103361.

Jahn AS, Navarini AA, Cerminara SE, Kostner L, Huber SM, Kunz M, *et al.* Over-detection of melanoma-suspect lesions by a CE-certified smartphone App: performance in comparison to dermatologists, 2D and 3D convolutional neural networks in a prospective data set of 1204 pigmented skin lesions involving patients' perception. *Cancers* 2022;**14**:3829.

Jamil U, Khalid S, Akram MU, Ahmad A, Jabbar S. Melanocytic and nevus lesion detection from diseased dermoscopic images using fuzzy and wavelet techniques. *Soft Comput* 2018;**22**:1577–93.

Janda M, Horsham C, Vagenas D, Loescher LJ, Gillespie N, Koh U, *et al.* Accuracy of mobile digital teledermoscopy for skin self-examinations in adults at high risk of skin cancer: an open-label, randomised controlled trial. *Lancet Digit Health* 2020;**2**:e129–37.

Jiji GW, DuraiRaj PJ. Content-based image retrieval techniques for the analysis of dermatological lesions using particle swarm optimization technique. *Appl Soft Comput* 2015;**30**:650–62.

Karolinska University Hospital, Karolinska Institute, Medical University of Vienna, Stockholm School of Economics. *AI-Augmented Skin Cancer Diagnosis in Teledermatoscopy*. NCT06080711

Majumder S, Ullah Muhammad A. A computational approach to pertinent feature extraction for diagnosis of melanoma skin lesion. *Pattern Recognit Image Anal* 2019;**29**:503–14.

Menzies SW, Sinz C, Menzies M, Lo SN, Yolland W, Lingohr J, *et al.* Comparison of humans vs. 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.

Mery D, Romero P, Garib G, Pedro A, Salinas Maria P, Sepulveda J, *et al. On Skin Lesion Recognition Using Deep Learning: 50 Ways To & Choose Your Model*. Image and Video Technology: 10th Pacific-Rim Symposium, PSIVT 2022, Virtual Event, 12–14 November 2022, Proceedings 2023:103–16.

Oliveira Roberta B, Marranghello N, Pereira Aledir S, Tavares JMR. A computational approach for detecting pigmented skin lesions in macroscopic images. *Expert Syst Appl* 2016;**61**:53–63.

Salido Julie Ann A, Ruiz C. *Using Morphological Operators and Inpainting for Hair Removal in Dermoscopic Images*. Proceedings of the Computer Graphics International Conference.

Surowka G, Ogorzallek M. Segmentation of the melanoma lesion and its border. *Int J Appl Math Comput Sci* 2022;**32**:683–99.

University Hospital Basel. *Melanoma Detection in Switzerland with VECTRA*. NCT04605822

University Hospital Ghent. *Artificial Intelligence-based Smartphone Application for Skin Cancer Detection*. NCT06080711

Yang Y. *Privacy-preserving with Sonification for Training of Convolutional Deep Neural Networks for Melanoma Diagnosis*. Proceedings of the 2021 International Conference on Bioinformatics and Intelligent Computing 2021:106–15.

Zhang Y, Xie F, Chen J. TFormer: a throughout fusion transformer for multi-modal skin lesion diagnosis. *Comput Biol Med* 2023;**157**:106712.

Zhao XY, Wu X, Li FF, Li Y, Huang WH, Huang K, *et al.* The application of deep learning in the risk grading of skin tumors for patients using clinical images. *J Med Syst* 2019;**43**:1–7.

Exclude design (n = 13)

Beltrami EJ, Brown AC, Salmon PJM, Leffell DJ, Ko JM, Grant-Kels JM. Artificial intelligence in the detection of skin cancer. *J Am Acad Dermatol* 2022;**87**:1336–42.

Daneshjou R, Barata C, Betz-Stablein B, Celebi ME, Codella N, Combalia M, *et al.* Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR Derm consensus guidelines from the International Skin Imaging Collaboration Artificial Intelligence Working Group. *JAMA Dermatol* 2022;**158**:90–6.

Ferrante di Ruffano L, Takwoingi Y, Dinnes J, Chuchu N, Bayliss SE, Davenport C, *et al.* Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults. *Cochrane Database Syst Rev* 2018;**12**:CD013186.

Filho M, Ma Z, Tavares JMRS. A review of the quantification and classification of pigmented skin lesions: from dedicated to hand-held devices. *J Med Syst* 2015;**39**:1–12.

Groh M, Harris C, Daneshjou R, Badri O, Koochek A. Towards transparency in dermatology image datasets with skin tone annotations by experts, crowds, and an algorithm. *Proc ACM Hum Comput Interact* 2022;**6**.

Haggenmuller S, Maron RC, Hekler A, Utikal JS, Barata C, Barnhill RL, *et al.* Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. *Eur J Cancer* 2021;**156**:202–16.

Marchetti MA, Liopyris K, Navarrete-Dechent C. Net benefit and decision curve analysis of competing diagnostic strategies for cutaneous melanoma. *J Am Acad Dermatol* 2021;**85**:e87–8.

Micocci M, Borsci S, Thakerar V, Walne S, Manshadi Y, Edridge F, *et al*. Attitudes towards trusting artificial intelligence insights and factors to prevent the passive adherence of GPs: a pilot study. *J Clin Med* 2021;**10**:14.

Skin Analytics. *The Evidence behind Skin Analytics* (unpublished). London: Skin Analytics; undated.

Skin Analytics. *Patient Sentiment Towards AI Use in Dermatology Pathways* (unpublished). London: Skin Analytics; 2023.

Sydney Melanoma Diagnostic Centre. *The Assessment of Experimental Artificial Intelligence (AI) Algorithms for the Diagnosis of Skin Tumours against Human Performance*. ACTRN12620000695909

Winkler JK, Haenssle HA. [Artificial intelligence-based classification for the diagnostics of skin cancer]. *Dermatologie (Heidelberg, Germany)* 2022;**73**:838–44.

Zhelev *et al. Cost-effectiveness of DERM: A Modelling Study*. Report version 1.0, 16 November 2023. [unpublished]. London: Skin Analytics; 2023.

Appendix 3 Ongoing studies

TABLE 22 DERM registered, ongoing studies

Study	Start–completion	Location	Population	Design	Main outcomes
DERM-006 (NCT05126173)	March 2022–September 2022	USA, Italy	Skin biopsy *N* = 1111 (actual)	Prosp. cohort	DA
ACTRN12619000398101	March 2022–December 2022 (anticipated)	Australia	Primary care (GP) *N* = 750 (target)	Prosp. cohort	DA Biopsy referral QoL

DA, diagnostic accuracy; prosp., prospective.

TABLE 23 Ongoing evaluations of DERM in the UK – post-referral

Commercial model	Location	Total cases assessed	Outcome data available ± included in performance reports	Currently live
Post referral pathway (following urgent suspected cancer referral)				
Commercial Partnership	University Hospitals Birmingham NHS Foundation Trust (Birmingham)	31,453	Yes	Yes
Paid deployment as part of AI in Health and Care Award (commercial)	Chelsea and Westminster NHS Foundation Trust (Chelsea)	5601	Yes	Yes
Commercial Partnership	West Suffolk NHS Foundation Trust (W Suffolk)	6054	Yes	Yes
Commercial Partnership	University Hospitals of Leicester NHS Trust (Leicester)	11,745	Yes	Yes
Paid deployment as part of AI in Health and Care Award	University Hospitals Bristol and Weston NHS Foundation Trust (Bristol)	2885	Yes	No
Paid deployment as part of AI in Health and Care Award (commercial)	Ashford and St Peter's Hospitals NHS Foundation Trust (Ashford)	2362	Yes	Yes
Commercial Partnership	East Suffolk and North Essex NHS Foundation Trust (E Suffolk)	605	Expected Q4 2024	No
Paid deployment as part of AI in Health and Care Award	Mid Cheshire NHS Foundation Trust (Mid Cheshire)	72	Expected Q4 2024	Recently launched
Paid deployment as part of AI in Health and Care Award	Royal Devon University Healthcare NHS Foundation Trust Eastern Services (E Devon)	20	Expected Q4 2024	Recently launched
Paid deployment as part of AI in Health and Care Award	Royal Devon University Healthcare NHS Foundation Trust Northern Services (*N* Devon)	0	Expected Q4 2024	Recently launched
Commercial Partnership	University Hospitals of Morecambe Bay NHS Foundation Trust (Morecambe Bay)	887	Expected Q4 2024	Yes

TABLE 24 Ongoing evaluations of DERM in the UK – pre-referral

Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Confidential information has been removed				
Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed

Appendix 4 Data extraction tables

TABLE 25 Selection criteria of studies included in the synthesis

Study	Reported selection criteria
DERM-003[24]	Dermatology clinic patients with >= 1 suspicious skin lesion suitable for photographing. Include lesions < 15 mm diameter, on a site suitable for photographing, not in area of visible scarring or tattooing, and not previously biopsied, excised or otherwise traumatised.
DERM-005 Chelsea and Westminster[25]	Adult patients with at least one suspicious lesion being photographed as part of standard care teledermatology; lesions < 15 mm, in a location suitable for photography, no previous trauma including biopsy or excision, no visible scarring or tattooing.
UHBFT and WSFT[26]	Adults with 1–3 suspicious lesions not larger than 15 mm. Exclusions are lesions that are not potentially malignant, those requiring monitoring for treatment response or staging of disease, non-dermascopic images of lesions, open ulcerated lesions, obscured by hair, tattoos or scars, subungual or on mucosal, genital or palmoplantar surfaces, previously biopsied lesions.
UHL[27]	Exclude < 18 years, 2 + lesions, genital lesions. No further details.
MacLellan 2021	Exclude: recurrent lesions or metastases; previously biopsied or excised; lesions < 2 mm or > 2 cm in diameter; lesions not accessible to the devices; lesions located on scars, crusts, psoriasis, eczema, sunburn, or other skin condition; lesions covered by thick hair; inaccessible genital, mucosal, obscured by foreign matter, ulcerated, sole, palm, close to eye; Fitzpatrick skin > III.
Winkler 2023	Melanocytic lesions. No further details.

TABLE 26 Full diagnostic accuracy results (DERM studies)

Study	Index test	Outcome	Sensitivity[a]	Specificity[a]	AUROC[a]	PPV[a]	NPV[a]
DERM-003	DERM v3.0 (iPhone 11)	Melanoma	93.3 (66.0–99.7)	73.6 (69.6–77.1)	92.6 (84.3–100)	8.7 (5.0–14.4)	99.8 (98.4–100)
		SCC	93.2 (80.3–98.2)	45.7 (41.3–50.1)	90.1 (86.1–94.0)	12.8 (9.5–17.1)	98.7 (96–99.7)
		BCC	95.8 (91.7–98)	45 (39.5–50.6)	92.0 (89.7–94.3)	51.1 (45.8–56.4)	94.7 (89.5–97.5)
		Malignant	96.0 (92.6–98)	45 (39.5–50.6)	NR	58 (53.1–62.7)	93.5 (88.1–96.7)
		IEC	100 (67.9–100)	46.6 (41–52.3)	89.0 (84.2–93.8)	6.2 (3.3–11.2)	100 (96.8–100)
		AK	84.8 (72.5–92.4)	47.2 (40.9–53.6)	81.1 (75.0–87.2)	27.5 (21.3–34.7)	92.9 (86.6–96.5)
		Atypical	59.1 (36.7–78.5)	43.9 (37.4–50.6)	89.4 (82.7–96.2)	9.2 (5.2–15.6)	91.7 (84.5–95.9)
		Benign	43.9 (37.4–50.6)	93.3 (90.0–95.6)	80.9 (77.3–84.5)	81.3 (73.1–87.5)	71.4 (67.0–75.5)
	Clinicians	Melanoma	81.2 (53.7–95.0)	98.9 (97.6–99.6)	90.3 (80.4–100)	68.4 (43.5–86.4)	99.5 (98.3–99.9)
		SCC	63.6 (47.7–77.2)	89.1 (86–91.5)	76.9 (69.6–84.3)	32.9 (23.4–44.1)	96.7 (94.5–98.0)
		BCC	97.5 (93.9–99.1)	77.4 (72.4–81.8)	90.0 (87.3–92.7)	72.6 (66.7–77.7)	98 (95.2–99.3)
		Malignant	93.8 (90–96.3)	77.4 (72.4–81.8)	NR	77 (71.9–81.4)	94.3 (90.6–96.7)
		IEC	90.9 (57.1–99.5)	78.8 (73.8–83.2)	63.6 (49.8–77.4)	13.2 (6.8–23.3)	99.6 (97.4–100)
		AK	96.7 (87.6–99.4)	79.3 (73.6–84)	85.0 (79.2–90.8)	53.1 (43.5–62.6)	99 (96.1–99.8)
		Atypical	76.2 (52.5–90.9)	73.9 (67.6-79.4)	85.1 (75.1–95)	21 (12.9–32.2)	97.1 (93.1–98.9)
		Benign	73.9 (67.6–79.4)	93.7 (90.5-95.9)	82.1 (78.8–85.5)	88.5 (83.0–92.5)	84.6 (80.5–87.9)

continued

TABLE 26 Full diagnostic accuracy results (DERM studies) (*continued*)

Study	Index test	Outcome	Sensitivity[a]	Specificity[a]	AUROC[a]	PPV[a]	NPV[a]
DERM-005	DERM (confidential information has been removed)	Malignant	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
	Teledermatologist		Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
	DERM (post hoc analysis)[b]		Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed	Confidential information has been removed
Thomas (2023)	Derm vA (UHB)	Melanoma	95.0 (90–97.6)	58.80 (57.4–60.2)	NR	6.7 (5.7–7.9)	99.7 (99.5–99.9)
	Derm vA (WSFT)		97.0 (84.7–99.5)	63.20 (59.5–66.7)	NR	11.4 (8.2–15.6)	99.8 (98.7–100)
	Derm vB (UHB)		100.0 (93.8–100)	80.90 (79.3–82.4)	NR	10.7 (8.4–13.6)	100.0 (99.8–100.0)
	Derm vB (WSFT)		100.0 (82.4–100)	80.40 (77.2–83.4)	NR	12.9 (8.3–19.4)	100.0 (99.2–100)
	Derm vA (UHB)	Malignant	96.0 (94.4–97.2)	45.00 (43.4–46.6)	NR	25.3 (23.7–26.9)	98.3 (97.6–98.8)
	Derm vA (WSFT)		99.3 (96.3–99.9)	33.1 (29.3–37.1)	NR	28.5 (24.8–32.5)	99.5 (97–99.9)
	Derm vB (UHB)		98.9 (96–99.7)	64.8 (62.9–66.7)	NR	17.4 (15.2–19.8)	99.9 (99.5–100.0)
	Derm vB (WSFT)		100.0 (94.7–100)	60.6 (56.6–64.5)	NR	23.1 (18.7–28.3)	100.0 (98.9–100.0)

AK, Actinic keratoses; NR, not reported.
a All results expressed as % (95% CI).
b Target sensitivity changed to > 95% for melanoma and SCC and > 90% for BCC.

TABLE 27 Included studies reporting subtype, Breslow thickness and stage of melanoma

Type of cancer	Lesion characteristics	DERM-003	DERM-005	MacLellan 2021	Winkler 2023
Subtype of melanoma					
Superficial spreading		9	Confidential information has been removed	NR	NR
Lentigo melanoma		1	Confidential information has been removed	NR	NR
Other/not available/ambiguous		6	Confidential information has been removed	NR	NR
Breslow thickness					
In situ		2	Confidential information has been removed	27	12
< 1.0 mm		7	Confidential information has been removed	NR	NR
1.01–2.0 mm		2	Confidential information has been removed	Mean (SD) 0.72 (0.56)	Invasive: 26
> 2.0 mm		4	Confidential information has been removed	Median (range) 0.57 (0.19–2.9)	
> 4 mm		0	Confidential information has been removed		
Not available		1	Confidential information has been removed	NR	NR
TOTAL		**16**	**Confidential information has been removed**	**59**	**38**

NR, not reported; SD, standard deviation.

TABLE 28 Included studies reporting subtype and stage of SCC, BCC and other malignancies

Lesion characteristics	DERM-003	DERM-005
SCC		
Subtype		
Poorly differentiated	4	Confidential information has been removed
Moderately differentiated	15	Confidential information has been removed
Well differentiated	16	Confidential information has been removed
Other/unknown	8	Confidential information has been removed
Stage		
Tis	1	Confidential information has been removed
T1	38	Confidential information has been removed
T2	0	Confidential information has been removed
T3	NR	Confidential information has been removed
T4	3	Confidential information has been removed
Not available/other/unknown	2	Confidential information has been removed
TOTAL	*44*	*Confidential information has been removed*
BCC		
Subtype		
Superficial	13	Confidential information has been removed
Nodular	94	Confidential information has been removed
Infiltrative	17	Confidential information has been removed
Morphoeic	0	Confidential information has been removed
Micronodular	2	Confidential information has been removed
Basosquamous	1	Confidential information has been removed
Not available/other/unknown	70	Confidential information has been removed
Stage		
Tis	3	Confidential information has been removed
T1	141	Confidential information has been removed
T2	2	Confidential information has been removed
T3	NR	Confidential information has been removed
T4	0	Confidential information has been removed
Not available/unknown	51	Confidential information has been removed
TOTAL	*197*	*Confidential information has been removed*
Other malignancies		
TOTAL	*2*	*Confidential information has been removed*

NR, not reported.

TABLE 29 Diagnostic pathway outcomes for patients in Thomas (2023)

Total number of cases (patients)		DERM vA		DERM vB	
		Birmingham *N* = 7171	West Suffolk *N* = 1119	Birmingham *N* = 4800	West Suffolk *N* = 1410
Not assessed with DERM[a]		27.4%	15.6%	25%	17%
Assessed with DERM[a]		72.6%	84.5%	75%	83%
Referred to dermatologist by DERM[a]	Total	53.2%	69.4%	44%	62%
	Malignant lesions	48.8%	67.0%	7.5%	9.7%
Judged non-malignant by DERM[a]	Total	19.4%	15.0%	31%	21.6%
	Discharged at second read	12.4%	7.8%	18.7%	10.7%
	Discharged after referral	2.8%	0.8%	4.8%	2.7%
	Malignant lesions	4.3%	6.4%	0	0

a All % are out of total *n* of cases/patients, including those not assessed by DERM.

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library