Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma

Review information

Review type: Diagnostic test accuracy

Review number: #164e

Authors

Naomi Chuchu[1], Yemisi Takwoingi[1], Jacqueline Dinnes[1], Rubeta N Matin[2], Oliver Bassett[3], Jacqueline F Moreau[4], Susan E Bayliss[1], Clare Davenport[1], Kathie Godfrey[5], Abhilash Jain[6], Fiona M Walter[7], Jonathan J Deeks[1], Hywel C Williams[8], Cochrane Skin Cancer Diagnostic Test Accuracy Group[1]

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[2]Department of Dermatology, Churchill Hospital, Oxford, UK
[3]Plastic Surgery, Addenbrooke's Hospital, Cambridge, UK
[4]Internal Medicine, University of Pittsburgh Medical Center, Pittsburgh, USA
[5]c/o Cochrane Skin Group, The University of Nottingham, Nottingham, UK
[6]Department of Plastic and Reconstructive Surgery, Imperial College Healthcare NHS trust, St Mary's Hospital, London, UK
[7]Public Health & Primary Care, University of Cambridge, Cambridge, UK
[8]Centre of Evidence Based Dermatology, University of Nottingham, Nottingham, UK

Citation example: Chuchu N, Takwoingi Y, Dinnes J, Matin RN, Bassett O, Moreau JF, Bayliss SE, Davenport C, Godfrey K, Jain A, Walter FM, Deeks JJ, Williams HC, Cochrane Skin Cancer Diagnostic Test Accuracy Group. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. Cochrane Database of Systematic Reviews , Issue . Art. No.: . DOI: .

Contact person

Jacqueline Dinnes

Institute of Applied Health Research
University of Birmingham
Birmingham
B15 2TT
UK

E-mail: j.dinnes@bham.ac.uk

Dates

Assessed as Up-to-date:	29 August 2016
Date of Search:	29 August 2016
Next Stage Expected:	Not provided
Protocol First Published:	Not specified
Review First Published:	Not specified
Last Citation Issue:	Not specified

What's new

Date	Event	Description

History

Date	Event	Description

Abstract

Background

Melanoma accounts for a small proportion of all skin cancer cases but is responsible for the majority of skin cancer-related deaths. Early detection and treatment can improve survival. Smartphone applications are readily accessible and potentially offer an instant risk assessment of the likelihood of malignancy, so that the right people seek further medical attention from a clinician for more detailed assessment of the lesion. There is, however, a risk that melanomas will be missed and treatment delayed if the application reassures the user that their lesion is low risk.

Objectives

To determine the diagnostic accuracy of smartphone applications to rule out cutaneous invasive melanoma and intraepidermal melanocytic variants in adults with concerns about suspicious skin lesions.

Search methods

We undertook a comprehensive search of the following databases from inception up to August 2016: Cochrane Central Register of Controlled Trials; MEDLINE; Embase; CINAHL; CPCI; Zetoc; Science Citation Index; US National Institutes of Health Ongoing Trials Register; NIHR Clinical Research Network Portfolio Database; and the World Health Organization International Clinical Trials Registry Platform. We studied reference lists and published systematic review articles.

Selection criteria

Studies of any design evaluating smartphone applications intended for use by individuals in a community setting who have lesions that might be suspicious for melanoma or intraepidermal melanocytic variants compared with a reference standard of histological confirmation or clinical follow-up and expert opinion.

Data collection and analysis

Two review authors independently extracted all data using a standardised data extraction and quality assessment form (based on QUADAS-2). Due to scarcity of data and poor quality of studies, no meta-analysis was undertaken for this review. For illustrative purposes, estimates of sensitivity and specificity were plotted on coupled forest plots for each application under consideration.

Main results

This review reports on two cohorts of lesions published in two studies. Both studies were at high risk of bias from selective participant recruitment, and high rates of non-evaluable images. Concerns about applicability of findings were high due to inclusion only of lesions already selected for excision in a dermatology clinic setting, and image acquisition by clinicians rather than by smartphone app users.

Data for five mobile phone applications were reported for 332 suspicious skin lesions with 86 melanomas across the two studies. Across the four artificial intelligence-based applications which classified lesion images (photographs) as melanomas (one application) or as high risk or ‘problematic’ lesions (three applications) using a pre-programmed algorithm, sensitivities ranged from 7% (95% CI: 2%, 16%) to 73% (95% CI: 52%, 88%) and specificities from 37% (95% CI: 29% to 46%) to 94% (95% CI: 87%, 97%). The single application using store-and-forward review of lesion images by a dermatologist had a sensitivity of 98% (95% CI: 90%, 100%) and specificity 30% (95% CI: 22%, 40%).

The number of test failures (lesion images analysed by the applications but classed as ‘not evaluable’ and excluded by the study authors) ranged from 3 to 31 (or 2% to 18% of lesions analysed). The store-and-forward application had one of the highest rates of test failure (15%). At least one melanoma was classed as ‘not evaluable’ in three of the four application evaluations.

Authors' conclusions

Smartphone applications using artificial intelligence-based analysis have not yet demonstrated sufficient promise in terms of accuracy, and are associated with a high likelihood of missing melanomas. Applications based on store-and-forward images could have a potential role in the timely presentation of people with potentially malignant lesions by facilitating active self-management health practices and early engagement of those with suspicious skin lesions; however, they may incur a significant increase in resource and workload. Given the paucity of evidence and low methodological quality, no implications for practice can be drawn. Nevertheless, this is a rapidly advancing field and new and better applications with robust reporting of studies could change these conclusions substantially.

Plain language summary

What is the diagnostic accuracy of smartphone applications (“apps”) when used by the general public for the detection of melanoma in adults?

What is the aim of the review?

The aim of this Cochrane Review was to find out how good smartphone applications intended for use by the general public are at providing an accurate assessment of the risk of melanoma in lesions that are suspicious for skin cancer.

Why is improving the diagnosis of malignant melanoma skin cancer important?

Melanoma is one of the most dangerous forms of skin cancer. Not recognising a melanoma when it is present (a false negative test result) could delay seeking appropriate advice and surgery to remove it. This risks the cancer spreading to other organs in the body, and possibly death. Diagnosing a skin lesion as a melanoma when it is not present (a false positive result) may result in increased patient anxiety and unnecessary surgery and further investigations.

What was studied in the review?

Specialised applications (“apps”) that provide advice on skin lesions or moles that might cause people concern are widely available for smartphones. Some applications allow people to photograph any skin lesion that they might be worried about and then provide guidance on whether to get medical advice. Some applications automatically classify lesions as high or low risk while others act as store-and-forward devices where images are sent to an experienced professional, such as a dermatologist, who then makes a risk assessment based on the photographic image. Researchers in Cochrane included 2 studies that evaluated five applications that used automated analysis of images and one that used a 'store-and-forward' approach to evaluate suspicious skin lesions.

What are the main results of the review?

The review included two studies with 332 lesions analysed by at least one smartphone application, including 86 melanomas. Both of the studies used photographs of moles or skin lesions that were about to be removed because doctors had already decided that they could be melanomas. The photographs were also taken by doctors instead of by people using their own smartphones to photograph lesions that were worrying them. For these reasons, we are not able to make a reliable estimate about how well smartphone applications actually work.

Four applications that produce an immediate (automated) assessment of a skin lesion or mole that has been photographed by the smartphone missed between 7 and 55 melanomas.

One application that sends the photograph of a mole or skin lesion to a dermatologist for assessment missed only one melanoma. Another 6 melanomas examined by the dermatologist via the application were not classified as high risk, instead the dermatologist was not able to classify the lesion as either 'atypical' (possibly a melanoma) or 'typical' (definitely not a melanoma).

How reliable are the results of the studies of this review?

The small number and poor quality of included studies reduces the reliability of findings. The people included were not typical of those who could use the applications in real life. The final diagnosis of melanoma was made by histology which is likely to have been a reliable method for deciding whether patients really had melanoma. However, the studies excluded between 2% and 18% of images because the applications failed to produce a recommendation.

Who do the results of this review apply to?

Studies were conducted in the US and Germany. Key patient information such as age and gender were not reported. The percentage of people with a final diagnosis of melanoma was 18% and 35%, much higher than observed in community settings. The definition of eligible patients was narrow in comparison to likely users of the applications. The photographs used were taken by doctors rather than by smartphone users, which seriously impacts on the applicability of results.

What are the implications of this review?

Current Smartphone applications using automated analysis are observed to have a high chance of missing melanomas (false negatives). Store and forward image applications could have a potential role in the timely identification of people with potentially malignant lesions by facilitating early engagement of those with suspicious skin lesions, but have resource and workload implications.

The development of applications to help identify people who might have melanoma is a fast-moving field. The emergence of new applications, higher quality and better reported studies could change the conclusions of this review substantially.

How up to date is this review?

The review authors searched for and used studies published up to August 2016.

Background

Target condition being diagnosed

Melanoma arises from uncontrolled proliferation of melanocytes - the epidermal cells that produce pigment or melanin (Thompson 2003). Melanoma can occur in any organ that contains melanocytes, including mucosal surfaces, the back of the eye, and lining around the spinal cord and brain (McLaughlin 2005), but most commonly arises in the skin (Erdmann 2013; Ferlay 2015). Cutaneous melanoma refers to any skin lesion with malignant melanocytes present in the dermis, and includes superficial spreading, nodular, acral lentiginous, and lentigo maligna melanoma variants (see Figure 1). Melanoma *in situ* refers to malignant melanocytes that are contained within the epidermis and have not yet invaded the dermis, but are at risk of progression to melanoma if left untreated (Thompson 2003; SEER 2007). Lentigo maligna, a subtype of melanoma-in-situ in chronically sun-damaged skin, denotes another form of proliferation of abnormal melanocytes. Lentigo maligna can progress to invasive melanoma if its growth breaches the dermo-epidermal junction during a vertical growth phase (when it becomes known as 'lentigo maligna melanoma'); however, its malignant transformation is both lower and slower than for melanoma *in situ* (Kasprzak 2015). Melanoma *in situ* and lentigo maligna are both atypical intraepidermal melanocytic variants (Thompson 2003; SEER 2007). Melanoma is one of the most dangerous forms of skin cancer, with the potential to metastasise to other parts of the body via the lymphatic system and blood stream; it accounts for only a small percentage of all skin cancer cases but is responsible for up to 75% of skin cancer deaths (Boring 1994; Cancer Research UK 2017a).

The incidence of melanoma has risen to over 200,000 newly diagnosed cases worldwide in 2012 (Erdmann 2013; Ferlay 2015), with an estimated 55,000 deaths (Ferlay 2015). Despite rising incidence, melanoma mortality appears to be stable (Apalla 2017). In the UK, melanoma has one of the fastest rising incidence rates of any cancer, and has had the biggest projected increase in incidence between 2007 and 2030 (Mistry 2011). In the decade leading up to 2013, age standardised incidence increased by 46%, with 14,500 new cases in 2013 and 2,459 deaths in 2014 (Cancer Research UK 2017b). Rates are higher in women than in men; however, the rate of incidence in males is increasing faster than in females (Arnold 2014).

The rising incidence in melanoma is thought to be primarily related to an increase in recreational sun exposure and tanning bed use and an increasingly ageing population with higher lifetime recreational ultraviolet (UV) exposure, in conjunction with possible earlier detection (Linos 2009; Belbasis 2016). Putative risk factors are reviewed in detail elsewhere (Belbasis 2016); however, risk factors can be broadly divided into host or environmental factors. Host factors include pale skin and light hair or eye colour; older age (Geller 2002); male sex (Geller 2002); previous skin

cancer (Tucker 1985); predisposing skin lesions, e.g., high melanocytic naevus counts (Gandini 2005a), clinically atypical naevi (Gandini 2005a), or large congenital naevi (Swerdlow 1995); genetically inherited skin disorders, e.g., xeroderma pigmentosum (Lehmann 2011); and a family history of melanoma (Gandini 2005a). Environmental factors include recreational, occupational, and work-related exposure to sunlight (both cumulative and episodic burning) (Gandini 2005b; Armstrong 2017); artificial tanning (Boniol 2012); and immunosuppression, e.g., as seen in organ transplant recipients or human immunodeficiency virus (HIV)-positive individuals (DePry 2011). Lower socioeconomic class may be associated with delayed presentation and thus more advanced disease at diagnosis (Reyes-Ortiz 2006).

Five-year survival for stage I melanoma is reported to be 91% to 95%, falling to 27% to 69% in stage III disease (Balch 2009). Tumour thickness, the presence of tumour ulceration and age are the main determinants of melanoma prognosis and prognostic tools have been developed that include such features (Mahar 2016). Before the advent of targeted and immuno-therapies, metastatic melanoma (involving distant sites and visceral organs) resulted in median survival of six to nine months with a three-year survival of 15% (Balch 2009; Korn 2008). Between 1975 and 2010, five-year relative survival for melanoma in the US increased from 80% to 94% but mortality rates showed little change, at 2.1 per 100,000 deaths in 1975 and 2.7 per 100,000 in 2010 (Cho 2014). Increasing incidence in localised disease over the same period (from 5.7 to 21 per 100,000) suggests that the observed survival benefits may be due to earlier detection and heightened vigilance (Cho 2014); however, targeted therapies for stage IV melanoma (e.g. BRAF inhibitors) have improved survival expectation, and immunotherapies are demonstrating potential for long-term survival (see below).

Treatment of melanoma

For primary melanoma, the mainstay of definitive treatment is wide local excision of the lesion, to remove both the tumour and any malignant cells that might have spread into the surrounding skin (Sladden 2009; Marsden 2010; NICE 2015; Garbe 2016; SIGN 2017). Recommended surgical margins vary according to tumour thickness (Garbe 2016) and stage of disease at presentation (NICE 2015). The role of narrower (e.g. 1 cm healthy tissue) excisions margins for thinner lesions is still debated (Sladden 2009; Wheatley 2016). Following histological confirmation of diagnosis, the lesion is staged according to the American Joint Committee on Cancer (AJCC) Staging System to guide treatment (Balch 2009). Stage 0 refers to melanoma *in situ*; stages I to II indicate localised melanoma; stage III occurs where there is regional metastasis; and stage IV indicates distant metastasis (Balch 2009). The main prognostic indicators can be divided into histological and clinical factors. Histologically, Breslow thickness is the single most important predictor of survival, as it is a quantitative measure of tumour invasion which correlates with the propensity for metastatic spread (Balch 2001). Microscopic ulceration, mitotic rate, microscopic satellites, regression, lymphovascular invasion, and nodular (rapidly growing) or amelanotic (lacking in melanin pigment) subtypes are also associated with worse prognosis (Shaikh 2012; Moreau 2013). Independent of tumour thickness, prognosis is worse in older people (Geller 2002), males (Geller 2002), those with recurrent lesions (Dong 2000), and in those with distant lymph node involvement (micro or macroscopic) and/or metastatic disease at the time of primary presentation (Balch 2009). There is debate regarding the prognostic effect from primary lesion site, with some evidence suggesting a worse prognosis for truncal lesions or those on the scalp or neck (Zemelman 2014).

Index test(s)

Smartphones are rapidly evolving from being devices for communication and entertainment to include specialised applications (“apps”) that are intimately involved in many aspects of daily life (Kassianos 2015). The processing powers of modern smartphones allow their use in more demanding tasks such as image analysis (Massone 2007). Melanoma risk assessment tools are recent additions and include applications such as “Mel App” and “Skin Scan” (Robson 2012).

Once downloaded to a user's mobile phone (both android and Apple iOS platforms), the applications can act as an information resource about melanoma or other skin cancer, can provide guidance on whether medical advice should be sought for a particular lesion that has been photographed by the mobile phone, or can be used to monitor skin lesions to identify any changes over time (Kassianos 2015).

Applications that provide guidance on particular skin lesions can use internally programmed algorithms (or 'artificial intelligence') to catalogue and classify the lesion images, or can act as store-and-forward devices whereby the lesion photograph is forwarded to an experienced professional such as a dermatologist for review and a recommendation regarding the nature of the lesion is communicated to the user via the application (essentially allowing members of the public direct access to a teledermatology-type service) (Kassianos 2015).

The artificial intelligence-based applications use algorithms to compare the acquired image against a bank of exemplar images of malignant and benign lesions or compare the image against a host of benign and malignant lesion characteristics learned from analysing thousands of other images to assess the likelihood of melanoma. These algorithms are generally based on fractal analysis. A fractal, in biology, is a natural phenomenon that exhibits a repeating pattern at every scale (Landini 2011). Fractal analysis can provide a quantitative measure of irregularity where regularity is expected. With regard to melanoma, this includes irregularities in a lesion’s physical characteristics, such as those used in established algorithms to assist in the diagnosis of melanoma (e.g. the ‘ABCs’ of melanoma (Friedman 1985)), as well as texture, patterns, and other geometric features. Fractal analysis has been used for diagnosis of other cancers (for example, mammography for breast cancer) (Rangayyan 2007; Raguso 2010), but has not historically been made available to consumers for assessment of their own malignancy risk. A major benefit of fractal analysis is that it is automated and thus observer-independent.

A recent review by Kassianos 2015 identified 39 available smartphone applications related to melanoma; most

were found to be multifunctional in that they provided information about melanoma in addition to lesion classification or a means of monitoring a given lesion. Just under half of the applications (46%; 18/39) provided some form of image analysis, and a quarter (23%; 9/39) were reported to provide a 'store-and-forward' lesion image review by a dermatologist. Those providing image analysis often did not describe how the photographic images were processed and analysed to provide advice on the likelihood of melanoma (Kassianos 2015). Four applications were described as providing an assessment of the likelihood of melanoma: two used an artificial intelligence-based algorithm based on the ABCDE method, one provided a risk approximation based on the completion of a visual analogue scale by the application user, and one provided insufficient information regarding the method involved (Kassianos 2015). Between 2014 and 2018, dermatology smartphone applications increased by 80.8% (an increase of 235), including an increase in teledermatology applications from 32 to 106 (Flaten 2018).

Clinical Pathway

Individuals or their relatives are often best placed to recognise suspicious or changing skin lesions and may use a range of resources to become better informed about their concerns. Smartphone applications could have a role very early on in the clinical pathway, as they are readily accessible and potentially offer an instant risk assessment of the likelihood of malignancy, reassuring those with benign appearing lesions and effectively triaging those who need to seek further medical attention from a clinician for more detailed assessment of the lesion (Figure 2).

In the UK, people with concerns about a new or changing lesion (either based on skin self-examination alone or with the aid of a mobile phone application) will then present to their general practitioner rather than directly to a specialist in secondary care (Figure 3). If the general practitioner has concerns then he/she usually refers the patient to a specialist in secondary care – usually a dermatologist but sometimes to a plastic surgeon or an ophthalmologist. Other systems may be in place in other countries with the possibility of presenting directly to a skin specialist. Suspicious skin lesions may also be identified in a specialist setting, for example, by a general surgeon or other specialist surgeon (including ear, nose, and throat (ENT) specialist (Figure 3) and referred for a specialist consultation with a dermatologist or plastic surgeon. Current UK guidelines recommend that all suspicious pigmented lesions presenting in primary care should be assessed by taking a clinical history and visual inspection using the revised seven-point checklist (MacKie 1990); suspicious pigmented skin lesions should be urgently referred for specialist assessment within two weeks (Marsden 2010; Chao 2013; NICE 2015). Evidence is emerging, however, to suggest that excision of melanoma by GPs is not associated with increased risk compared with outcomes in secondary care (Murchie 2017). The specialist clinician will use history-taking, visual inspection of the lesion (in comparison with other lesions on the skin), and usually dermoscopy to inform a clinical decision. If melanoma is suspected, then urgent excision is recommended. Other lesions such as suspected dysplastic naevi or pre-malignant lesions such as lentigo maligna may also be referred for a diagnostic biopsy, further surveillance, or reassurance and discharge.

Role of index test(s)

Advances in smartphone technology have provided innovative platforms, where people can become more educated about their medical conditions, leading to better engagement with healthcare professionals (Robertson 2014). The use of smartphone technology can facilitate active self-management health practices and provide patients information related to their condition (Tyagi 2012). As they are self-initiated, psychological barriers to seeking medical advice can potentially be reduced, as assessments are performed outside clinical settings and are often interactive and personalised (Tyagi 2012). The advances in smartphone technology provide new strategies of engaging patients in the management of potentially suspicious skin lesions, increasing the likelihood of detecting melanoma earlier in the progression of the disease. Early detection of melanoma is crucial for patients as survival is dramatically improved and morbidity is reduced (Balch 2009). There is increased value not only to the users but healthcare professionals alike, as more educated patients can better engage with their doctors making consultations more effective and efficient (Robertson 2014).

The greatest concern for the use of mobile phone applications in this context is their ability to accurately stratify lesions by level of risk of development of melanoma, particularly given the potential for falsely reassuring people that their lesion is benign. There is real concern that people could be dissuaded from accessing healthcare advice if their lesion were deemed by the app to be low risk (Robson 2012). The most useful applications will therefore be those that maximise sensitivity for the detection of melanoma over specificity. There is however a concern that those who use such applications may be the ‘worried well’ rather than those who might actually have melanoma, which in turn could flood limited healthcare resources with unnecessary referrals, or simply generate profits for private providers who may take advantage of public cancer anxiety.

Alternative test(s)

For the purposes of our series of reviews, each component of the diagnostic process, including visual inspection or clinical examination, whether delivered in-person or remotely via teledermatology, for example, is considered a diagnostic or index ‘test', the accuracy of which can be established in comparison with a reference standard of diagnosis, either alone or in combination with other available technologies that may assist the diagnostic process.

Once a suspicious lesion has been identified by an individual or their relatives, the only alternative to the use of a mobile phone application is to immediately seek medical advice from their general practitioner or specialist clinician. The clinician will then use history-taking, visual inspection of the lesion (in comparison with other lesions on the skin), and usually dermoscopy to inform a clinical decision (Figure 2). The accuracy of visual inspection alone (Dinnes 2018a) and dermoscopy added to visual inspection (Dinnes 2018b) have also been reviewed as part of our series of systematic

reviews.

A review of the accuracy of teledermatology has also been conducted (Chuchu 2018). Teledermatology, whereby clinical photographs or dermoscopic images of a skin lesion are taken and forwarded to a dermatologist, has traditionally been used by non-specialist clinicians to obtain a specialist opinion on a suspicious lesion. This can be done on a store-and-forward basis, using digital cameras or mobile phones to acquire photographs or dermoscopic images of a lesion, or via a live videolink. According to UK guidelines, both clinical and dermoscopic images must be sent for 'full dermatology', i.e. as a replacement for a face-to-face consultation, whereas for 'triage teledermatology' dermoscopic images should be sent where facilities permit (BAD 2013).

Teledermatology not only allows clinicians rapid access to expert opinion but may lead to a reduction in waiting times and limit unnecessary referrals (Ndegwa 2010; Warshaw 2010; Bashshur 2015). In rural areas, where access to speciality services can have significant and potentially off-putting travel and time implications for the patient, teledermatology has the potential to widen access to specialist opinion.

Teledermatology is also becoming available in a community setting, especially within community or "high street" pharmacies (for example, the Boots 'Mole Scanning Service', www.boots.com/health-pharmacy-advice/skin-services/mole-scanning-service), and is therefore a potential alternative to use of a smartphone application. Due to their extended opening hours, ease of access, presence of healthcare professionals and availability of consultation rooms, community pharmacies are increasingly used for the provision of early detection services (Kjome 2016), such as mole scanning by trained pharmacy staff. In theory, using pharmacies as the first-line identifier to separate those with skin lesions requiring follow-up from those who do not, gives general practitioners and specialists more time and resources for those who require intervention (Kjome 2016).

A number of other tests which may have a role in the diagnosis of melanoma in a specialist setting have been reviewed as part of our series of systematic reviews, including reflectance confocal microscopy, optical coherence tomography, computer-aided diagnosis or artificial intelligence-based techniques, and high frequency ultrasound (New Reference).

Evidence permitting, the accuracy of available tests will be compared in an overview review, exploiting within-study comparisons of tests and allowing the analysis and comparison of commonly used diagnostic strategies where tests may be used singly or in combination.

Rationale

Our series of reviews of diagnostic tests used to assist clinical diagnosis of melanoma aims to identify the most accurate approaches to diagnosis and provide clinical and policy decision-makers with the highest possible standard of evidence on which to base decisions. With increasing rates of melanoma incidence and the push towards the use of dermoscopy and other high-resolution image analysis in primary care without adequate evidence of effectiveness or safety, the anxiety around missing early cases needs to be balanced against the risk of over referrals, to avoid sending too many people with benign lesions for a specialist opinion. It is questionable whether all skin cancers picked up by sophisticated techniques, even in specialist settings, help to reduce morbidity and mortality or whether newer technologies run the risk of increasing false-positive results. It is also possible that use of some technologies, e.g., widespread use of dermoscopy in primary care with no training, could actually result in harm by missing melanomas if they are used as replacement technologies for traditional history-taking and clinical examination of the entire skin; many branches of medicine have noted the danger of such "gizmo idolatry" amongst doctors (Leff 2008).

Smartphone applications in general are already widely available and used by consumers, and the popularity of such platforms to offer clinical and diagnostic advice is ever increasing. Given the rapidly changing evidence base and lack of available systematic reviews on the topic, there is a need for an up-to-date analysis of the accuracy of smartphone applications.

This is one of a series of Cochrane Diagnostic Test Accuracy (DTA) reviews on the diagnosis and staging of melanoma and keratinocyte skin cancers as part of the National Institute for Health Research (NIHR) Cochrane Systematic Reviews Programme. Appendix 1 shows the content and structure of the programme. As several reviews for each topic area followed the same methodology, generic protocols were prepared in order to avoid duplication of effort: one for diagnosis of melanoma (New Reference) and one for diagnosis of keratinocyte skin cancers (New Reference). The Background and Methods sections of this review therefore use some text that was originally published in the protocol concerning the evaluation of tests for the diagnosis of melanoma (New Reference) and text that overlaps some of our other reviews (Dinnes 2018b). Appendix 2 provides a glossary of terms used.

Objectives

To determine the diagnostic accuracy of smartphone applications to rule out cutaneous invasive melanoma and intraepidermal melanocytic variants in adults with concerns about suspicious skin lesions.

Secondary objectives

No secondary objectives, including formal investigations of heterogeneity, were investigated.

Methods

Criteria for considering studies for this review

Types of studies

We included test accuracy studies that allow comparison of the result of the index test with that of a reference standard, including the following:

- studies where all participants receive a single index test and a reference standard;
- studies where all participants receive more than one index test(s) and reference standard;
- studies where participants are allocated (by any method) to receive different index tests or combinations of index tests and all receive a reference standard (between-person comparative studies (BPC));
- studies that recruit series of participants unselected by true disease status (referred to as case series for the purposes of this review);
- diagnostic case-control studies that separately recruit diseased and non-diseased groups (see Rutjes 2005);
- both prospective and retrospective studies; and
- studies where previously acquired clinical or dermoscopic images were retrieved and prospectively interpreted for study purposes.

We excluded studies from which we could not extract 2x2 contingency data or small studies with less than five disease positive or less than five disease negative participants. Although the size threshold of five is arbitrary, such small studies are likely to give unreliable estimates of sensitivity or specificity, and may be biased like small randomised controlled trials of treatment effects.

Participants

We included studies in adults with pigmented skin lesions or lesions suspicious for melanoma. These could include those at high risk of developing melanoma, including those with a family history or previous history of melanoma skin cancer, atypical or dysplastic naevus syndrome, or genetic cancer syndromes. Ideally, participants should be recruited from community settings; however, due to an anticipated paucity of data, participants recruited from any setting were considered eligible. We excluded studies that recruited only participants with malignant diagnoses and studies that compared test results in participants with malignancy compared with test results based on 'normal' skin as controls, due to the bias inherent in such comparisons (Rutjes 2006). We excluded studies with more than 50% of participants aged 16 and under.

Index tests

Studies evaluating smartphone applications intended for use by any individual (or member of the public) with a smartphone in a community setting who has a skin lesion that they are concerned about were included. Applications intended for use by smartphone users were considered to be those using standard photographs acquired by the mobile phone. Applications that were intended for use with dermoscopic or other microscopic attachments for the acquisition of magnified images were considered to be intended for use by clinicians such as GPs, as a means of accessing a specialist dermatologist opinion (i.e. for store-and-forward teledermatology assessments).

Studies developing new mobile phone applications (i.e. derivation studies) were **included** if they:

- used a separate independent 'test set' of participants or images to evaluate the new approach.

Studies were **excluded** if they:

- used a statistical model to produce a data driven equation, or algorithm based on multiple diagnostic features, with no separate test set;
- used cross-validation approaches such as 'leave-one-out' cross-validation (Efron 1983); or
- evaluated the accuracy of the presence or absence of individual lesion characteristics or morphological features, with no overall diagnosis of malignancy.

Target conditions

The target condition was defined as the detection of cutaneous melanoma and atypical melanocytic intraepidermal variants (i.e., including melanoma in situ, or lentigo maligna, which has a risk of progression to invasive melanoma).

Reference standards

The preferred reference standard for establishing the final diagnosis of a skin lesion is histopathological diagnosis of the excised lesion or biopsy sample in all eligible lesions. Histopathlogical assessment is not a perfect reference standard because it only samples lesions for examination and may therefore miss tumour cells in non-sampled portions. Also being a subjective assessment, there is some degree of inter-observer variation especially for borderline lesions. A qualified pathologist or dermatopathologist should perform histopathology. Ideally, reporting should be standardised detailing a minimum dataset to include the histopathological features of melanoma to determine the American Joint Committee on Cancer (AJCC) Staging System (e.g. Slater 2014). We did not apply the reporting standard as a necessary inclusion criterion, but extracted any pertinent information.

Due to the potential for partial verification (with lesion excision or biopsy unlikely to be carried out for all benign-appearing lesions within a representative population sample), we also accepted clinical follow-up of benign-appearing lesions, cancer registry follow-up and 'expert opinion' with no histology or clinical follow-up as eligible reference standards. The risk of differential verification bias (as misclassification rates of histopathology and follow-up will differ) was considered in our quality assessment of studies.

All of the above were considered eligible reference standards for establishing lesion final diagnoses with the following caveats:

- all study participants with a final diagnosis of the target disorder must have a histological diagnosis, either subsequent to the application of the index test or after a period of clinical follow-up, and
- at least 50% of all participants with benign lesions must have either a histological diagnosis or clinical follow-up to confirm benignity.

The ability of a smartphone application to correctly trriage those who need further assessment of suspicious skin lesionsis not the only outcome of interest for this type of test however. The 'referral accuracy' (or ability of the smartphone application to approximate an in-person lesion assessment) can be estimated by comparing the action recommended by the smartphone application with the management recommendation from face to face assessment by an appropriately qualified clinician. To this end, 'expert opinion' as the sole reference standard is an eligible reference standard for our reviews of both mobile phone applications and for teledermatology.

Search methods for identification of studies

Electronic searches

The Information Specialist (SB) carried out a comprehensive search for published and unpublished studies. A single large literature search was conducted to cover all topics in the programme grant (see Appendix 1 for a summary of reviews included in the programme grant). This allowed for the screening of search results for potentially relevant papers for all reviews at the same time. A search combining disease-related terms with terms related to the test names, using both text words and subject headings was formulated. The search strategy was designed to capture studies evaluating tests for the diagnosis or staging of skin cancer. As the majority of records were related to the searches for tests for staging of disease, a filter using terms related to cancer staging and to accuracy indices was applied to the staging test search, to try to eliminate irrelevant studies, for example, those using imaging tests to assess treatment effectiveness. A sample of 300 records that would be missed by applying this filter was screened and the filter adjusted to include potentially relevant studies. When piloted on MEDLINE, inclusion of the filter for the staging tests reduced the overall numbers by around 6000. The final search strategy, incorporating the filter (Appendix 3), was subsequently applied to all bibliographic databases as listed below. The final search result was cross-checked against the list of studies included in five systematic reviews; our search identified all but one of the studies, and this study is not indexed on MEDLINE. The Information Specialist devised the search strategy, with input from the Information Specialist from Cochrane Skin. No additional limits were used.

We searched the following bibliographic databases to 29 August 2016 for relevant published studies:

- MEDLINE via OVID (from 1946);
- MEDLINE In-Process & Other Non-Indexed Citations via OVID; and
- EMBASE via OVID (from 1980).

We searched the following bibliographic databases to 30 August 2016 for relevant published studies:

- the Cochrane Central Register of Controlled Trials (CENTRAL) Issue 7, 2016, in the Cochrane Library;
- the Cochrane Database of Systematic Reviews (CDSR) Issue 8, 2016 in the Cochrane Library;
- Cochrane Database of Abstracts of Reviews of Effects (DARE) Issue 2, 2015;
- CRD HTA (Health Technology Assessment) database Issue 3, 2016;
- CINAHL (Cumulative Index to Nursing and Allied Health Literature via EBSCO from 1960).

We searched the following databases for relevant unpublished studies:

- CPCI (Conference Proceedings Citation Index) via Web of Science™ (from 1990);
- Zetoc (from 1993)
- SCI Science Citation Index Expanded™ via Web of Science™ (from 1900, using the "Proceedings and Meetings Abstracts" Limit function).

We searched the following trials registers:

- The US National Institutes of Health Ongoing Trials Register (www.clinicaltrials.gov);
- NIHR Clinical Research Network Portfolio Database (http://www.nihr.ac.uk/research-and-impact/nihr-clinical-research-network-portfolio/);
- The World Health Organization International Clinical Trials Registry Platform (apps.who.int/trialsearch/).

We aimed to identify all relevant studies regardless of language or publication status (published, unpublished, in press, or in progress). No date limits were applied. Update searches will be time and resource dependent.

Searching other resources

We have included information about potentially relevant ongoing studies in the 'Characteristics of ongoing studies' tables. We have screened relevant systematic reviews identified by the searches for their included primary studies, and included any missed by our searches. We have checked the reference lists of all included papers, and subject experts within the author team have reviewed the final list of included studies. No citation searching has been conducted.

Data collection and analysis

Selection of studies

Titles and abstracts were screened by at least one author (JDi or NC), with any queries discussed and resolved by consensus. A pilot screen of 539 MEDLINE references showed good agreement (89% with a kappa of 0.77)

between screeners. Primary test accuracy studies and test accuracy reviews (for scanning of reference lists) of any test used to investigate suspected melanoma, basal cell carcinoma (BCC), or cutaneous squamous cell carcinoma (cSCC) were included at initial screening. Inclusion criteria (Appendix 4) were applied independently by both a clinical reviewer (from one of a team of twelve clinician reviewers) and a methodologist reviewer (JDi or NC) to all full text articles, disagreements were resolved by consensus or by a third party (JDe, CD, HW, and RM). Authors of eligible studies were contacted when insufficient data were presented to allow for the construction of 2x2 contingency tables.

Data extraction and management

One clinical (as detailed above) and one methodological reviewer (JDi, NC or LFR) independently extracted data concerning details of the study design, participants, index test(s) or test combinations and criteria for index test positivity, reference standards, and data required to populate a 2x2 diagnostic contingency table for each index test using a piloted data extraction form. Data were extracted at all available index test thresholds. Disagreements were resolved by consensus or by a third party (JDe, CD, HW, and RM).

Authors of included studies were contacted where information related to the target condition (in particular to allow the differentiation of invasive cancers from '*in situ*' variants) or diagnostic threshold were missing. Authors of conference abstracts published from 2013 to 2015 were contacted to ask whether full data were available. If no full paper was identified, we marked conference abstracts as 'pending' and will revisit them in a future review update.

Dealing with multiple publications and companion papers

Where multiple reports of a primary study were identified, we maximised yield of information by collating all available data. Where there were inconsistencies in reporting or overlapping study populations, we contacted study authors for clarification in the first instance. If contact with authors was unsuccessful, we used the most complete and up-to-date data source where possible.

Assessment of methodological quality

We assessed risk of bias and applicability of included studies using the QUADAS-2 checklist (Whiting 2011), tailored to the topic of skin cancer diagnosis (see Appendix 5). The modified QUADAS-2 tool was piloted on a small number of full text articles included across the full series of diagnostic test accuracy reviews (New Reference). One clinical and one methodological reviewer (JDi, NC or LFR) independently assessed quality for the remaining studies; any disagreement was resolved by consensus or by a third party where necessary (JDe, CD, HW, and RM).

Statistical analysis and data synthesis

Due to scarcity of data and the poor quality of studies, no meta-analysis was undertaken for this review. For illustrative purposes, estimates of sensitivity and specificity were plotted on coupled forest plots for each application under consideration. Our unit of analysis was the lesion rather than the patient. This is because (i) in skin cancer initial treatment is directed to the lesion and not systemically to a patient (thus it is important to be able to correctly identify cancerous lesions within each patient), and (ii) it is the most common way in which the primary studies reported data. Although there is a theoretical possibility of correlations of test errors when multiple lesions are included from the same patients, most studies include very few patients with multiple lesions and any potential impact on findings is likely to be very small, particularly in comparison with other concerns regarding risk of bias and applicability. For each analysis, only one dataset was included per study to avoid multiple counting of lesions.

Investigations of heterogeneity

We examined heterogeneity between studies by visually inspecting the forest plots of sensitivity and specificity and summary ROC plots. Insufficient numbers of studies were identified to allow meta-regression to investigate potential sources of heterogeneity.

Sensitivity analyses

No sensitivity analyses were done.

Assessment of reporting bias

Due to uncertainty about the determinants of publication bias for diagnostic accuracy studies and the inadequacy of tests for detecting funnel plot asymmetry (Deeks 2005), no tests to detect publication bias were performed.

Results

Results of the search

A total of 34,347 unique references were identified and screened for inclusion. Of these, 1051 full text papers were reviewed for eligibility and 203 publications were included in at least one of the suite of reviews of tests to assist in the diagnosis of melanoma or keratinocyte skin cancer. Figure 4 provides a PRISMA flow diagram of search and eligibility results. A total of 16 studies were tagged as potentially eligible for this review of smartphone applications; ultimately, 2 publications were included. The reasons for exclusion are listed in Figure 4, but included: sample includes less than five melanoma cases (n=2; Massone 2007; Robson 2012); index test (n=3; including two studies where mobile phones were used to capture dermoscopic (Massone 2007) or otherwise magnified (Diniz 2016) images in a specialist clinic setting); derivation studies that did not separate data for training and test sets (n=2; Ramlakhan 2011; Wadhawan 2011) and duplicate or related publication (n=1; Von Braunmühl 2015 reported data for same patients as Maier 2015). A list of the 14 studies excluded from this review with reasons for exclusion is provided in Characteristics of excluded studies, with a list of all studies excluded from the full

series of reviews available as a supplementary file.

Across all of our reviews, the corresponding authors of 84 studies were contacted and asked to supply further information to allow study inclusion (37) to clarify diagnostic thresholds (18) or target condition definition (29). Responses were received from 39 authors allowing the inclusion of four studies across various reviews, and data clarifications were provided by 23 authors. One author contacted in relation to this review provided the additional data needed to allow the study to be included (Wolf 2013).

This review reports on a total of two cohorts of lesions published in two studies, providing 6 datasets (Wolf 2013; Maier 2015). A total of 332 lesions were successfully analysed by the applications, with a total of 86 melanomas. The number of participants with lesions included in the studies was not reported

Wolf 2013 evaluated photographs of lesions retrospectively selected from their dermatology database of lesions that had been scheduled for excision using a case-control type design. These lesion images were routinely captured from participants who had presented with suspicious skin lesions in a dermatological rather than community setting where these applications are intended to be used and included only lesions with a final diagnosis of melanoma, melanoma *in situ*, lentigo, benign nevi (including compound, junctional and low grade dysplastic naevi), dermatofibroma, sebhorrhoeic keratosis and haemangioma. The study included lesions with good quality photographs (as assessed by one or two dermatologists) and with a clear histological diagnosis. Lesions that were uncommon or had an equivocal diagnosis (such as “melanoma cannot be ruled out” or “atypical melanocytic proliferation”) and lesions with moderate or high grade atypia were all excluded. Over half of images reviewed for inclusion in the study (52%; 202/390) were excluded due to poor image quality, the presence of identifiable patient features or insufficient clinical or histological information. Between 3 and 29 additional lesions were analysed by the applications but considered 'not evaluable' or test failures (see Findings).

Maier 2015 conducted a prospective case series of patients with melanocytic skin lesions seen routinely at the department of dermatology for “skin cancer screening”. It is unclear whether participants were referred or could access the dermatology clinic directly. Up to three smartphone images (photographs) per lesion were described as having been acquired prior to lesion excision, and were presumably taken by the dermatologist however, the image acquisition process was not clearly described. Twenty lesions (10%; 20/195) were excluded due to poor image quality or incompletely imaged. An additional 31 lesions excluded by the study authors were considered as test failures for the purposes of this review (see Findings), including 13 excluded due to ‘two-point differences’ (explained as non-consecutive risk classes, presumably for different images of the same lesion) and 18 were classed as ‘tie-cases’ (defined as having ‘an equal number of results in two consecutive risk classes, e.g. 1 high risk, 1 medium risk and 1 low risk result) (Maier 2015).

The studies were conducted in the US (Wolf 2013) and Germany (Maier 2015). No information was reported on the number of patients recruited in either study; total numbers of included lesions were 188 (Wolf 2013) and 144 (Maier 2015). Patient characteristics such as age and gender were not reported. Both studies reported on the accuracy of smartphone applications for the detection of melanoma and its melanocytic intraepidermal variants. The prevalence of melanoma was 18% (Maier 2015) and 35% (Wolf 2013).

Wolf 2013 evaluated four different smartphone applications. No names of the applications were given to avoid consumer bias; therefore, the applications were numbered one to four to allow an assessment of accuracy. Three applications were artificial intelligence-based classifications of lesions as “problematic” vs “okay” (App 1), “melanoma vs not melanoma” (App 2) and “high risk vs medium/low-risk” (App 3) as part of the assessment of the images. Data for App 3 could also be extracted considering high and medium risk as test positive vs low risk as test negative. Application 4 used a store-and-forward approach with remote lesion assessment by a qualified dermatologist. This application could be run on either a smartphone or a website with lesion images uploaded and transmitted remotely to a dermatologist to make an assessment and return it to the user within 24 hours. The output given was “atypical” vs “typical”.

Maier 2015 evaluated an automated risk assessment algorithm using the SkinVision App; this relied on fractal image analysis of three images per lesion; lesions were classified by the application as “high risk” versus “medium” or “low risk”. The diagnostic accuracy of face-to-face clinical diagnosis by a dermatologist was also reported for the same lesions.

In both studies the reference standard diagnosis was made by histology alone (all lesions were either biopsied or excised).

Methodological quality of included studies

The overall methodological quality of included studies is summarised in (Figure 5 and Figure 6).

Both studies were assessed as being at high risk of bias for participant selection due to the inappropriate exclusion of lesions that would have otherwise being eligible for assessment with the applications. Wolf 2013 also used a case-control design, including only lesions with particular final diagnoses. Similarly, both had high concern regarding included participants and setting, due to unclear reporting of patient samples and whether there was inclusion of multiple lesions per patient. Wolf 2013 excluded lesions that were common or with equivocal diagnoses. All studies included only lesions selected for excision. This is not a representative spectrum of lesions that would be observed in daily life as they are already a highly selected sample of participants. These participants would have already presented to a doctor with concerns about a particular lesion and therefore likely represent a more severe spectrum of abnormality which will artificially increase the sensitivity of the test in comparison to use by smart phone users in general.

Both studies were assessed as being at low risk of bias in the index test domain, with the artificial intelligence-based assessment made without knowledge of the histological diagnosis. All had a pre-specified threshold. However, both studies had high concern about applicability of the index test as neither was conducted in the manner in which the tests are intended to be used in practice. Wolf 2013 used archived photographs of lesions rather than images taken using a phone and did not

report the real names of the applications used. Maier 2015 did use smartphone images; however, it was unclear how the imaging process was undertaken.

All studies reported the use of an acceptable reference standard; however, blinding of histology to the index test result was not reported in Maier 2015. Only Wolf 2013 reported histopathological interpretation by an experienced dermatopathologist.

Both studies were at high risk of bias for flow and timing due to exclusion of non-evaluable images from further analysis (the numbers excluded are stated allowing computation of test failure rates). Both were unclear on the interval between image capture and performance of the reference standard.

Findings

Detection of invasive melanoma or melanoma in situ

Across the 5 different applications assessed in the two studies, sensitivities for the detection of invasive melanoma or intraepidermal melanocytic variants ranged from 7% (95% CI: 2%, 16%) to 98% (95% CI: 90%, 100%) and specificities from 30% (95% CI: 22%, 40%) to 94% (95% CI: 87%, 97%) (Figure 7).

One of the four artificial intelligence-based applications attempted to correctly identify lesions as melanomas versus not. The resulting sensitivity and specificity were 69% (95% CI: 55%, 80%) and 37% (95% CI: 29%, 46%) (App 2 in Wolf 2013; 185 lesions and 58 melanoma cases).

The remaining three artificial intelligence-based algorithms attempted to categorise lesions as high risk or 'problematic' versus not (Figure 7). Sensitivities were around 70% for App 1 in Wolf 2013 (70%, 95% CI: 57%, 81%) and for the SkinVision app in Maier 2015 (73%, 95% CI: 52%, 88%), but was only 7% (95% CI: 2%, 16%) for App 3 in Wolf 2013. The corresponding specificities for the three applications were 39% (95%CI: 31, 49%) (Wolf 2013; 182 lesions and 60 melanomas), 83% (95% CI: 75%, 89%) (Maier 2015; 144 lesions and 26 melanomas), and 94% (95% CI: 87%, 97%) (Wolf 2013; 170 lesions and 59 melanomas). Decreasing the threshold for considering lesions as test positive to include both high and medium risk lesions for App 3 in Wolf 2013 increased sensitivity from 7% to 54% (95%CI 41%, 67%), with a fall in specificity from 94% to 61% (95%CI: 52%, 70%) (denoted App 3b in Figure 7).

The final application (App 4 in Wolf 2013) was a store-and-forward system, with lesion images classified as atypical or typical by a dermatologist. The sensitivity for this application was 98% (95% CI: 90%, 100%) and specificity 30% (95% CI: 22%, 40%) (159 lesions and 54 melanoma cases).

This application however, recorded the highest percentage of 'test failures' for this study (i.e. eligible lesions analysed by the applications but recorded as 'unevaluable') (Wolf 2013). The test failure rates were: 3% (App 1), 2% (App 2), 6% (App 3) and 15% (App 4; designated by the dermatologist as "send another photograph" or "unable to categorise") (Table 1). At least one melanoma was classed as 'unevaluable' for three of the four applications, with 6 (10%) of melanomas considered not evaluable by the dermatologist conducting the assessment for the store-and-forward application (App 4).

A total of 31 lesions (18%; 31/175) analysed with the SkinVision application (Maier 2015) could not be classified as high, medium or low risk (considered 'unevaluable' results) (Table 1). The number of melanomas classed as 'unevaluable' was not reported, however 35% of all melanomas originally eligible for the study (14/40) were excluded by the study authors either due to poor image quality or because the images were classed as not evaluable.

Maier 2015 also reported the accuracy of face-to-face clinical diagnosis of the same lesions by a dermatologist. This data is not directly comparable to the accuracy of the application as the in-person assessment relates to the diagnosis of melanoma whereas the SkinVision application was developed to identify lesions at high risk of melanoma. The sensitivity of the face-to-face assessment was 85% (95% CI: 65%, 96%) and specificity 97% (95% CI: 93%, 99%) (Figure 8; 144 lesions and 26 melanomas).

Investigations of heterogeneity

We were unable to undertake investigations of heterogeneity listed in the protocol due to insufficient number of studies.

Discussion

Summary of main results

This review aimed to assess the accuracy of smartphone applications for the detection of invasive melanoma or intraepidermal melanocytic variants. We included two studies with a total of 332 lesions, 86 of which were melanomas (Summary of findings table 1).

Studies were generally of poor methodological quality. Risk of bias was low for both studies only for the index test domain. Poor reporting did not always allow the quality of the reference standard to be adequately judged. Study participants were highly selected in comparison to those who might choose to use a smartphone application to check a skin lesion that was causing them concern. Both of the studies used photographs of skin lesions that were scheduled for excision in a dermatology clinic setting, and the images were acquired by clinicians instead of by people using their own smartphones, potentially leading to the acquisition of higher quality images. Index test interpretation was blinded, pre-specified test thresholds were used and adequate reference standards were used. Blinding of the reference standard to the lesion images was not reported in one study and interpretation by an experienced histopathologist was not reported in one. We are therefore unable to make a reliable estimate of the accuracy of smartphone applications for the detection of melanoma or intra-epidermal melanocytic variants.

Across the four artificial intelligence-based applications which classified lesion images (photographs) as melanomas (one

application) or as high risk or 'problematic' lesions (three applications), sensitivities ranged from 7% (95% CI: 2%, 16%) to 73% (95% CI: 52%, 88%) and specificities from 37% (95% CI: 29% to 46%) to 94% (95% CI: 87%, 97%). This means that between 27% and 93% of invasive melanoma or intraepidermal melanocytic variants were not picked up as requiring further assessment by a clinician by the automated applications (or as melanomas by one of the four applications). With a prevalence of melanoma ranging between 18% and 37% for these evaluations, the number of melanomas missed was between 7 and 55.

The single application using store-and-forward review of lesion images by a dermatologist had a sensitivity of 98% (95% CI: 90%, 100%) and specificity 30% (95% CI: 22%, 40%); one melanoma was missed by the dermatologist.

The number of test failures (lesion images analysed by the applications but classed as 'not evaluable' and excluded by the study authors) ranged from 3 to 31 (or 2% to 18% of lesions analysed). The store-and-forward application had one of the highest rates of test failure (15%). At least one melanoma was classed as 'not evaluable' in three of the four application evaluations, the highest number of melanomas excluded by the dermatologist evaluating the store-and-forward images (6/60 melanomas assessed).

Strengths and weaknesses of the review

The strengths of this review include an in-depth and comprehensive electronic literature search, systematic review methods including double extraction of papers by both clinicians and methodologists, and contact with authors to allow study inclusion or clarify data. A clear analysis structure was planned to allow test accuracy in different study populations to be estimated and a detailed and replicable analysis of methodologic quality was undertaken.

No other systematic reviews of smartphone applications were identified during the preparation of this review. Kassianos 2015 however systematically attempted to identify all available smartphone applications as of July 2014 by searching the online stores of smartphone providers (Apple and Android) and then systemically extracted data about the applications from their online descriptions. No attempt was made to identify any diagnostic test accuracy research underlying the applications. It is notable that Kassianos 2015 identified 39 applications and we were only able to identify test accuracy evaluations for five. We did not contact developers of commercially available smartphone applications for any further accuracy data; however, this could be undertaken for an update of the review.

The main concerns for the review are the clinical applicability of the findings, and exclusion of non evaluable test results, with likely over-estimation of sensitivity.

Applicability of findings to the review question

The data included in this review is unlikely to be generally applicable to the intended setting. Study participants were those with skin lesions that were already scheduled for excision in a dermatology clinic setting rather than smartphone users with concerns about a new or changing mole or skin lesion, and the lesion photographs used with the applications were likely taken by dermatologists in the clinic setting in both studies. One study also excluded equivocal lesions or those with moderate or high-grade atypia both of which could potentially be more likely to produce not evaluable results or to be misclassified by the application.

Authors' conclusions

Implications for practice

No summary estimates of test accuracy could be produced to answer the research question for this review. Smartphone applications using artificial intelligence-based analysis have not yet demonstrated sufficient promise in terms of accuracy, and are associated with a high likelihood of missing melanomas. Available data have limited applicability in practice due to selective participant recruitment from secondary referral settings and images were not acquired by the intended users of the smartphone applications (i.e. members of the public). Applications based on store-and-forward images could have a potential role in the timely presentation of people with potentially malignant lesions by facilitating active self-management health practices and early engagement of those with suspicious skin lesions; however, there are resource and workload implications from the use of a store-and-forward approach.

Given the paucity of evidence and low methodological quality, no implications for practice can be drawn. Nevertheless, this is a fast-moving field, and new and better apps and better reported studies could change these conclusions substantially.

Implications for research

Prospective evaluation of the ability of smartphone applications to correctly identify people with suspicious skin lesions who should seek further medical advice from a suitably qualified clinician is required for a full and proper evaluation of accuracy. Studies should be conducted in a clinically relevant community or primary care setting, recruiting smartphone users who may have concerns about their risk of developing melanoma or concerns about a new or changing skin lesion. The recommendation from the smartphone could be compared with the recommendation from a GP following a face-to-face clinical diagnosis of the same lesion. In such a study it is important that all lesions examined by the smartphone are assessed in the same way by the GP with blinding to the smartphone recommendation. Although histological confirmation of melanoma versus not melanoma is the ideal reference standard, it is not a practical or ethical one for study participants with lesions at low risk of malignancy. Systematic follow-up of non-excised lesions over a five year period would avoid over-reliance on a histological reference standard and would further allow results to be more generalisable to routine practice. Although a pragmatic evaluation amongst the general population of smartphone users would be challenging, studies could be undertaken in those most at risk of developing melanoma, in whom the prevalence of disease would be higher. Use of the test by

smartphone users themselves rather than healthcare professionals or equipment experts is also key to ensuring the clinical applicability of study findings and to determine the true test failure rate, which could seriously inhibit the use of smartphone applications in practice. Any future research study should conform to the updated Standards for Reporting of Diagnostic Accuracy (STARD) guideline (Bossuyt 2015).

Acknowledgements

Members of the Cochrane Skin Cancer Diagnostic Test Accuracy Group include:

- the full project team (Susan Bayliss, Naomi Chuchu, Clare Davenport, Jonathan Deeks, Jacqueline Dinnes, Lavinia Ferrante di Ruffano, Kathie Godfrey, Rubeta Matin, Colette O'Sullivan, Yemisi Takwoingi, Hywel Williams);
- our 12 clinical reviewers (Rachel Abbott, Ben Aldridge, Oliver Bassett, Sue Anne Chan, Alana Durack, Monica Fawzy, Abha Gulati, Jacqui Moreau, Lopa Patel, Daniel Saleh, David Thompson, Kai Yuen Wong) and 2 methodologists (Lavinia Ferrante di Ruffano and Louise Johnston) who assisted with full text screening, data extraction and quality assessment across the entire suite of reviews of diagnosis and staging and skin cancer;
- our expert advisors and co-authors Abhilash Jain and Fiona Walter; and
- all members of our Advisory Group (Jonathan Bowling, Colin Fleming, Matthew Gardiner, Abhilash Jain, Susan O'Connell, Pat Lawton, John Lear, Mariska Leeflang, Richard Motley, Paul Nathan, Julia Newton-Bishop, Miranda Payne, Rachael Robinson, Simon Rodwell, Julia Schofield, Neil Shroff, Hamid Tehrani, Zoe Traill, Fiona Walter).

Cochrane Skin editorial base wishes to thank Urbà González, who was the Dermatology Editor for this review; and the clinical referees, Saul Halpern and David de Berker. We also wish to thank the Cochrane DTA editorial base and colleagues.

Contributions of authors

NC was the contact person with the editorial base.
NC co-ordinated contributions from the co-authors and wrote the final draft of the review.
NC, JD, OB, JM screened papers against eligibility criteria.
NC obtained data on ongoing and unpublished studies.
NC, JD, OB appraised the quality of papers.
NC, OB extracted data for the review and sought additional information about papers.
NC entered data into RevMan.
NC analysed and interpreted data.
NC, JD, SB, CD, YT, JJD worked on the methods sections.
NC, HW, RM, OB, JM, AJ, FW drafted the clinical sections of the background and responded to the clinical comments of the referees.
JD responded to the methodology and statistics comments of the referees.
KG was the consumer co-author and checked the review for readability and clarity, as well as ensuring outcomes are relevant to consumers.
JD is the guarantor of the update.

Disclaimer

This project was supported by the National Institute for Health Research, via Cochrane Infrastructure funding to the Cochrane Skin Group and Cochrane Programme Grant funding. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health.

Declarations of interest

Naomi Chuchu: nothing to declare.
Yemisi Takwoingi: nothing to declare.Susan E Bayliss: nothing to declare.
Jac Dinnes: I am employed by the University of Birmingham under a NIHR Cochrane Programme Grant to produce the reviews.
Rubeta N Matin: My institution received a grant for a BARCO NV commercially sponsored study to evaluate digital dermoscopy in the skin cancer clinic; I have received payment from Public Health England for 'Be Clear on Cancer Skin Cancer' report and royalties for Oxford Handbook of Medical Dermatology (Oxford University Press). I have no conflicts of interest to declare that directly relate to the publication of this work.
Oliver Bassett: nothing to declare.
Jacqueline F Moreau: I helped draft the paper and performed data analysis for a study that is included (Wolf 2013).
Susan E Bayliss: nothing to declare.
Clare Davenport: My employer (The University of Birmingham) received funding for my participation in this review as part of an NIHR programme grant awarded to Jac Dinnes, the PI.
Kathie Godfrey: I have received reimbursement of travel expenses incurred by attending meetings.
Abhilash Jain: nothing to declare.
Fiona M Walter: nothing to declare.
Jonathan J Deeks: Funding was provided to the University of Birmingham from a Cochrane Programme Grant to complete this review and other linked reviews.
Hywel C Williams: I am director of the NIHR HTA Programme. HTA is part of the NIHR which also supports the NIHR systematic reviews programme from which this work is funded.
Clinical referee David de Berker: I am Principal investigator for a single site in a multicentre study for assessment of images

in pigmented lesions. The sponsor is Skin Analytics. I receive no payment for this from Skin Analytics, although they pay the hospital for participation of our site.

Differences between protocol and review

Primary objectives and primary target condition have been changed from detection of cutaneous invasive melanoma alone, to the detection of cutaneous invasive melanoma and intraepidermal melanocytic variants, as the latter is more clinically relevant in practice.

The primary objective was also amended from "To determine the diagnostic accuracy of smartphone applications for the detection of cutaneous invasive melanoma and intraepidermal melanocytic variants in adults when used by consumers" to: "To determine the diagnostic accuracy of mobile phone (or smartphone applications) to rule out cutaneous invasive melanoma and intraepidermal melanocytic variants in adults with concerns about suspicious skin lesions" in order to better reflect the intended role of smartphone applications.

Secondary objectives related to the detection of any skin cancer or skin lesion with a high risk of progression to melanoma and the original primary objective related to detection of invasive melanoma alone were not investigated due to lack of data.

To improve clarity of methods, this text from the protocol "We will include studies developing new algorithms or methods of diagnosis (i.e., derivation studies) if they use a separate independent 'test set' of participants or images to evaluate the new approach. We will also include studies using other forms of cross validation, such as 'leave-one-out' cross-validation (Efron 1983). We will note for future reference (but not extract) any data on the accuracy of lesion characteristics individually, e.g., the presence or absence of a pigment network or detection of asymmetry" has been replaced with "Studies developing new mobile phone applications (i.e. derivation studies) were **included** if they:

- used a separate independent 'test set' of participants or images to evaluate the new approach.

Studies were **excluded** if they:

- used a statistical model to produce a data driven equation, or algorithm based on multiple diagnostic features, with no separate test set;
- used cross-validation approaches such as 'leave-one-out' cross-validation (Efron 1983); or
- evaluated the accuracy of the presence or absence of individual lesion characteristics or morphological features, with no overall diagnosis of malignancy."

As per the secondary objectives above, the target conditions of invasive melanoma alone and of any skin cancer or skin lesion with a high risk of progression to melanoma have been removed from the review due to lack of data.

Clarification added to index test section that smartphone app use by clinicians for specialists second opinion is covered in teledermatology whereas intended use by general public is for this review.

We proposed to supplement the database searches by searching the annual meetings of appropriate organisations (e.g., British Association of Dermatologists Annual Meeting, American Academy of Dermatology Annual Meeting, European Academy of Dermatology and Venereology Meeting, Society for Melanoma Research Congress, World Congress of Dermatology, European Association of Dermato Oncology); however, due to volume of evidence retrieved from database searches and time restrictions we were unable to do this.

For quality assessment, the QUADAS-2 tool was further tailored according to the review topic.

In terms of analysis, restriction to analysis of per patient data was not performed due to lack of data. Heterogeneity investigations and sensitivity analyses were not performed as planned due to lack of data.

Published notes

Characteristics of studies

Characteristics of included studies

Maier 2015

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection**: Not reported **Country** Germany
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients seen routinely for skin cancer screening at the Department of Dermatology **Setting:** Secondary (general dermatology) **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** Poor quality index test image; (elements in the image not belonging to the lesion e.g. hair, images containing more than one lesion, incomplete imaged lesions); non-melanocytic lesions. Also excluded "two-point differences cases" mainly due to inappropriate imaging angle or distance (we assume this to mean lesions with results in non-consecutive risk classes, e.g. 1 high risk and 2 low risk); and tie cases (described as cases with an equal number of results in two consecutive risk classes, e.g. 1 high risk, 1 medium risk and 1 low risk result). **Sample size (patients):** Not reported **Sample size (lesions):** No. eligible: 195; No. included: 175 (at least 3 images included per lesion) **Participant characteristics:** Not reported **Lesion characteristics:** Not reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Mobile phone application** **Acquisition and transmission of images:** Secondary care **Nature of images used:** Clinical photographs **Any additional patient information provided:** Unclear the clinical and dermoscopic diagnosis was independently documented **Diagnostic threshold** the SkinVision application evaluates lesions to be of high risk (red), medium risk (yellow) and low risk (green). We classified histologically proven naevi (benign and dysplastic) as low or medium risk **Diagnosis based on:** artificial intelligence-based diagnosis # **In person assessment** **Method of diagnosis:** Visual inspection and Dermosocpy **Prior test data:** Not reported **Diagnostic threshold:** Diagnosis of melanoma **Diagnosis based on:** single observer Number of examiners: two **Observer qualifications:** Dermatologist **Experience in practice:** Unclear-not specified **Experience with index test:** Unclear-not specified

Index test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **Details:** *Histology (excision)* - 195 eligible lesions including 40 melanomas; 20 lesions excluded due to image quality (lesion types not reported), leaving 175 lesions analysed by the application (number of melanomas remaining not reported) **Target condition (Final diagnoses)** For the sample of 195: Melanoma (in situ or invasive): 40; Dysplastic naevi (mild/moderate) 42; benign naevi 113 For the analysed sample of 175: Lesion diagnoses not reported For the final sample of 144: Melanoma (in situ or invasive): 26; Dysplastic naevi (mild/moderate) 34; Benign naevus: 84
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the reference standard results interpreted without knowledge of the referral diagnosis?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**1. Excluded participants**: Twenty lesions (10%) excluded due to poor image quality (significant amount of hair, lesion out of focus or multiple lesions in the focus). An additional 31 were excluded as not evaluable (13 lesions (6%) based on two-point-differences and 18 (9%) tie cases with an equal number of results in different risk classes). **2. Time interval to reference test:** not reported - assume it is < 1 month as images of the lesions were taken prior to excision
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
Could the patient flow have introduced bias?	High risk

Notes

Notes	-

Wolf 2013

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control study **Data collection:** Retrospective image selection/Prospective interpretation **Period of data collection** Not reported **Country** USA
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Lesion images selected from the institution image database with specific and clear histologic diagnoses including: melanoma, melanoma in situ, lentigo, benign naevi (including compun, junctionla and low grade dysplastic naevi), dermatofibroma, sebhorrhoeic keratosis and haemangioma. **Setting:** Unspecified **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Unspecified **Exclusion criteria:** Poor quality index test image Images that contained any identifiable features, such as facial features, tattoos, or labels with patient information, were excluded or cropped to remove the identifiable features or information Lesions with specific diagnoses including: Spitz naevi, Reed nevus, uncommon or equivocal lesions; and lesions with moderate or high-grade atypia. **Sample size (patients):** Not reported **Sample size (lesions):** No. eligible 390; No. included 188 **Participant characteristics:** Not reported **Lesion characteristics:** Not reported **Other:** (Note: Von Braunmühl 2015 extrapolates sensitivity and specificity to the whole dataset of 195 lesions by including the poor quality index test images- excluded as overlapping populations with Maier 2015)
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**1. Mobile phone application** **Acquisition and transmission of images:** Secondary care **Nature of images used:** Not reported **Any additional patient information provided:** No further information used **Diagnostic threshold:** Application 1-The application analyses the image and gives an assessment of “problematic” (positive test result) or “okay” (negative test result) Application 2- The output given is “melanoma” (positive test result) or “looks good” (negative test result). Application 3-The output given is “high risk” (positive test result) or “medium risk” or “low risk,” both of which we considered to be a negative test result. Application 4- The dermatologist assigns an output of “atypical” (positive test result) or “typical” (negative test result); images classified as "send another photograph" or "unable to categorise" were considered test failures and excluded by study authors. **Observer qualifications (remote diagnosis):** Application 4 only: images interpreted by a board-certified dermatologist (n=NR)

Index test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** - Histological diagnosis alone **Details:** Reference standard details *Histology (not further described)* - No. patients/lesions: A total of 188 lesions - Disease positive: 60 melanomas - Disease negative: 128 benign **Target condition (Final diagnoses)** Malignant - Melanoma (in situ and invasive): 60 Benign - 'Benign' diagnoses: 128
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Were the reference standard results interpreted without knowledge of the referral diagnosis?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	1. **Excluded participants**: 202/390 lesion images excluded due to poor image quality, containing identifiable patient information or features, or lacking sufficient clinical or histological information. Between 3 and 29 additional lesions were analysed by the applications but considered 'not evaluable' or test failures. The test failure rates were: 3% (n=6; App 1), 2% (n=3; App 2), 6% (n=12; App 3) and 15% (n=29; App 4). 2. **Time interval to reference test**: N/A
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
Could the patient flow have introduced bias?	High risk

Notes

Notes	-

Footnotes

n – number; N/A – not applicable; No. – number; NR – not reported.

Characteristics of excluded studies

Braun 2015

Reason for exclusion	EXCLUDE on 2x2 data; case report

Burki 2013

Reason for exclusion	EXCLUDE not a primary study

Diniz 2016

Reason for exclusion	EXCLUDE on index test (*mobile phone used along with amplifying microscopic lens)*

Jahan-Tigh 2016

Reason for exclusion	EXCLUDE on index test *(Telediagnosis of ex vivo pathology specimens)*

Karargyris 2012

Reason for exclusion	EXCLUDE on study population EXCLUDE on target condition

Lai 2015

Reason for exclusion	EXCLUDE conference abstract EXCLUDE on 2x2 data

Massone 2007

Reason for exclusion	EXCLUDE on sample size (*<5 cases of melanoma as final diagnosis);* EXCLUDE on index test *mobile phone used to capture dermoscopic images*

Ramlakhan 2011

Reason for exclusion	EXCLUDE if derivation study (*results of the training and test sets not differentiated Table II)*

Robson 2012

Reason for exclusion	EXCLUDE on sample size

Varma 2011

Reason for exclusion	EXCLUDE not a primary study (*Editorial)*

Von Braunmühl 2015

Reason for exclusion	duplicate or related publication (Maier 2015)

Wadhawan 2011

Reason for exclusion	EXCLUDE if derivation study

Yu 2011

Reason for exclusion	EXCLUDE on study population EXCLUDE on 2x2 data

Zouridakis 2015

Reason for exclusion	EXCLUDE not a primary study (*book chapter)*

Footnotes

Characteristics of studies awaiting classification

Footnotes

Characteristics of ongoing studies

Footnotes

Summary of results tables

1 Summary of findings

Question	What is the diagnostic accuracy of Smartphone applications for the diagnosis of cutaneous melanoma in adults
Participants	Adults with suspicious skin lesions.
Prior testing and prevalence	Studies did not report the basis for participant selection. One selected a sample of lesions previously imaged during routine care just before excision of the lesion. The second study evaluated the test on patients who had been referred for further screening of the lesion by a specialist. Prevalence of melanoma was 18% and 35%.
Settings	Secondary care
Target condition(s)	Invasive melanoma and melanocytic intraepidermal variants
Index test	Smartphone applications intended for use by the general public. Lesions not visualised by applications were excluded.
Reference standard	Histology
Action:	If accurate, positive results of smartphone applications will help to highlight lesions of concern to the lay public promoting earlier diagnosis of melanoma and reducing consultations for benign lesions.
Limitations	
Risk of bias:	Patient selection methods at high risk of bias due to selective inclusion of lesion types (2/2) and use of a case-control design (1/2). Test interpretation was blinded to reference standard and pre-specified for artificial intelligence-based diagnosis (2/2). Reference standard blinding was not described. Timing of index and reference standards was not reported. Exclusions due to test failures were not reported (1/2) or their final diagnoses were not described (2/2)
Applicability of evidence to question:	High concerns about applicability due to unrepresentative participant samples with high disease prevalences (2/2). Test not applied and interpreted by the intended user of the application (2/2). Reference standard interpretation by experienced histopathologists was not described (1/2).
Total number of studies: 2	
Detection of melanoma	

Quantity of evidence	Number of studies	2	Total participants with test results	332	Total with target condition	86*

Findings	Across the four artificial intelligence-based applications which classified lesion images (photographs) as melanomas (one application) or as high risk or 'problematic' lesions (three applications), sensitivities ranged from 7% (95% CI: 2%, 16%) to 73% (95% CI: 52%, 88%) and specificities from 37% (95% CI: 29% to 46%) to 94% (95% CI: 87%, 97%). This means that between 27% and 93% of invasive melanoma or intraepidermal melanocytic variants were not picked up as high risk by the automated applications (or as melanomas by one of the four applications). With a prevalence of melanoma ranging between 18% and 37% for these evaluations, the number of melanomas missed was between 7 and 55. The single application using store-and-forward review of lesion images by a dermatologist had a sensitivity of 98% (95% CI: 90%, 100%) and specificity 30% (95% CI: 22%, 40%); one melanoma was missed by the dermatologist. The number of test failures (lesion images analysed by the applications but classed as 'not evaluable' and excluded by the study authors) ranged from 3 to 31 (or 2% to 18% of lesions analysed). The store-and-forward application had one of the highest rates of test failure (15%). At least one melanoma was classed as 'not evaluable' in three of the four application evaluations, the highest number of melanomas excluded by the dermatologist evaluating the store-and-forward images (6/60 melanomas assessed).

Footnotes

* of the 60 melanomas included in one study, between 54 and 60 were successfully analysed by the four applications evaluated

Additional tables

1 Index test failures – lesions not evaluable by smartphone application

Study	Total number of lesions (melanomas) successfully assessed by the each 'app'		Lesions not evaluable (%)	Number (%) of melanomas 'not evaluable'
Wolf 2013 188 lesions (60 melanomas)[1]	Application 1	182 (60)	6 (3%)	0 (0%)
	Application 2	185 (58)	3 (2%)	2 (3%)
	Application 3	170 (59)	12 (6%)	1 (2%)
	Application 4	159 (54)	29 (15%)	6 (10%)
Maier 2015 175 lesions (number of melanomas: not reported; between 26 and 40)[2]	SkinVision	144 (26)	31 (18%)	Not reported (<=14)

Footnotes

[1] 52% (202/390) of all images reviewed for inclusion in the study were excluded prior to analysis by the applications.

[2] 10% (20/195) of all images reviewed for inclusion in the study were excluded due to poor image quality or incomplete imaging. The original sample of 195 lesions included 60 melanomas. The number of melanomas excluded on the basis of image quality and the number analysed by the application but not included by the study authors (considered as test failures for the purposes of this review) were not separately reported.

References to studies

Included studies

Maier 2015

* Maier T, Kulichova D, Schotten K, Astrid R, Ruzicka T, Berking C, et al. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. Journal of the European Academy of Dermatology and Venereology : JEADV 2015;29(4):663-7. [Other: ER4:25012308; PubMed: 25087492]

Wolf 2013

* Wolf JA, Moreau JF, Akilov O, Patton T, English JC 3rd, Ho J, et al. Diagnostic inaccuracy of smartphone applications for melanoma detection. JAMA Dermatology 2013;149(4):422-6. [Other: ER4:15466167; PubMed: 23325302]

Excluded studies

Braun 2015

Braun RP, Marghoob A. High-dynamic-range dermoscopy imaging and diagnosis of hypopigmented skin cancers. JAMA Dermatology 2015;151(4):456-7. [PubMed: 25535875]

Burki 2013

Burki TK. Diagnostic accuracy of smartphone applications. Lancet Oncology 2013;14(3):e90. [PubMed: 23580957]

Diniz 2016

Diniz LE, Ennser K. Melanoma detection using a mobile phone app. In: Levitz D, Ozcan A, Erickson D, editors(s). Proceedings of SPIE. Optics and Biophotonics in Low-Resource Settings II edition. Vol. 9699. March 7, 2016. [DOI: 10.1117/12.2212446]

Jahan-Tigh 2016

Jahan-Tigh RR, Chinn GM, Rapini RP. A comparative study between smartphone-based microscopy and conventional light microscopy in 1021 dermatopathology specimens. Archives of Pathology & Laboratory Medicine 2016;140(1):86-90. [PubMed: 26717060]

Karargyris 2012

Karargyris A, Karargyris O, Pantelopoulos A. DERMA/care: An advanced image-processing mobile application for monitoring skin cancer. In: Tools with Artificial Intelligence (ICTAI), 2012 IEEE 24th International Conference on. Vol. Vol 2. 2012:1-7. [DOI: 10.1109/ICTAI.2012.180]

Lai 2015

Lai I, Ko J, Pathipati A. DermLens: Device for mobile teledermatology. Journal of the American Academy of Dermatology 2015;72(5, Suppl 1):AB88. [EMBASE: 71895108]

Massone 2007

Massone C, Hofmann-Wellenhof R, Ahlgrimm-Siess V, Gabler G, Ebner C, Soyer HP. Melanoma screening with cellular

phones. PloS ONE 2007;2(5):e483. [PubMed: 17534433]

Ramlakhan 2011

Ramlakhan K, Shang Y. A mobile automated skin lesion classification system. In: Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on. 23rd IEEE, 2011:138-41. [DOI: 10.1109/ICTAI.2011.29]

Robson 2012

Robson Y, Blackford S, Roberts D. Caution in melanoma risk analysis with smartphone application technology. British Journal of Dermatology 2012;167(3):703-4. [PubMed: 22762381]

Varma 2011

Varma S. Mobile teledermatology for skin tumour screening. British Journal of Dermatology 2011;164(5):939-40. [PubMed: 21518326]

Von Braunmühl 2015

Von Braunmühl T. Smartphone apps for skin cancer diagnosis? The Munich study [Smartphone Apps für die Hautkrebs-Diagnose? – die Münchner Studie]. Kosmetische Medizin 2015;36(4):152-157.

Wadhawan 2011

Wadhawan T, Situ N, Rui H, Lancaster K, Yuan X, Zouridakis G. Implementation of the 7-point checklist for melanoma detection on smart handheld devices. IEEE Engineering in Medicine and Biology Magazine - Conference Proceedings 2011; 2011:3180-3. [PubMed: 22255015]

Yu 2011

Yu LS, Joseph AONR, Lindsley EH, Farkas DL. Polarization-sensitive digital dermoscopy for image processing-assisted evaluation of atypical nevi: towards step-wise detection of melanoma. In: Farkas DL, Nicolau DV, Leif RC, editors(s). Proceedings of SPIE. Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues IX edition. Vol. 7902. February 28, 2011. [DOI: 10.1117/12.891083]

Zouridakis 2015

Zouridakis G, Wadhawan T, Situ N, Hu R, Yuan X, Lancaster K, et al. Melanoma and other skin lesion detection using smart handheld devices. Methods in Molecular Biology 2015;1256:459-96. [PubMed: 25626557]

Studies awaiting classification

Ongoing studies

Other references

Additional references

Altamura 2008

Altamura D, Avramidis M, Menzies SW. Assessment of the optimal interval for and sensitivity of short-term sequential digital dermoscopy monitoring for the diagnosis of melanoma. Archives of Dermatology 2008;144(4):502-6. [PubMed: 18427044]

Apalla 2017

Apalla Z, Lallas A, Sotiriou E, Lazaridou E, Ioannides D. Epidemiological trends in skin cancer. Dermatology Practical & Conceptual 2017;7(2):1. [DOI: 10.5826/dpc.0702a01]

Armstrong 2017

Armstrong BK, Cust AE. Sun exposure and skin cancer, and the puzzle of cutaneous melanoma: A perspective on Fears et al. Mathematical models of age and ultraviolet effects on the incidence of skin cancer among whites in the United States. American Journal of Epidemiology 1977; 105: 420-427. Cancer Epidemiology 2017;48:147-56. [PubMed: 28478931]

Arnold 2014

Arnold M, Holterhues C, Hollestein LM, Coebergh JW, Nijsten T, Pukkala E, et al. Trends in incidence and predictions of cutaneous melanoma across Europe up to 2015. Journal of the European Academy of Dermatology and Venereology : JEADV 2014;28(9):1170-8. [PubMed: 23962170]

BAD 2013

British Association of Dermatology. Quality standards for Teledermatology using ‘store and forward’ images. www.bad.org.uk/shared/get-file.ashx?itemtype=document&id=794 (accessed prior to 16 May 2018).

Balch 2001

Balch CM, Soong SJ, Gershenwald JE, Thompson JF, Reintgen DS, Cascinelli N, et al. Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system. Journal of Clinical Oncology 2001;19(16):3622-34. [PubMed: 11504744]

Balch 2009

Balch CM, Gershenwald JE, Soong SJ, Thompson JF, Atkins MB, Byrd DR, et al. Final Version of 2009 AJCC Melanoma Staging and Classification. Journal of Clinical Oncology 2009;27(36):6199-206. [PubMed: 19917835]

Bashshur 2015

Bashshur RL, Shannon GW, Tejasvi T, Kvedar JC, Gates M. The empirical foundations of teledermatology: a review of the research evidence. Telemedicine Journal and E-Health 2015;21(12):953-79. [PubMed: 26394022]

Belbasis 2016

Belbasis L, Stefanaki I, Stratigos AJ, Evangelou E. Non-genetic risk factors for cutaneous melanoma and keratinocyte skin cancers: an umbrella review of meta-analyses. Journal of Dermatological Science 2016;84(3):330-339. [PubMed: 27663092]

Boniol 2012

Boniol M, Autier P, Boyle P, Gandini S. Cutaneous melanoma attributable to sunbed use: systematic review and meta-analysis. BMJ 2012;345:e4757. [PubMed: 22833605]

Boring 1994

Boring CC, Squires TS, Tong T, Montgomery S. Cancer statistics, 1994. CA: a Cancer Journal for Clinicians 1994; 44(1):7-26. [PubMed: 8281473]

Bossuyt 2015

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527. [DOI: 10.1136/bmj.h5527; PubMed: 26511519]

Cancer Research UK 2017a

Cancer Research UK. Skin cancer statistics. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer#heading-One. (accessed prior to 21 July 2017).

Cancer Research UK 2017b

Skin cancer incidence statistics. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer/incidence (accessed prior to 30 May 2018).

Chao 2013

Chao D, London Cancer (North and East). Guidelines for Cutaneous Malignant Melanoma Management August 2013. www.londoncancer.org/media/76373/london-cancer-melanoma-guidelines-2013-v1.0.pdf (accessed 25 February 2015).

Cho 2014

Cho H, Mariotto AB, Schwartz LM, Luo J, Woloshin S. When do changes in cancer survival mean progress? The insight from population incidence and mortality. Journal of the National Cancer Institute. Monographs 2014;2014(49):187-97. [PubMed: 25417232]

Chuchu 2018

Chuchu N, Dinnes J, Takwoingi Y, Matin RN, Bayliss SE, Davenport C, et al. The use of teledermatology for the diagnosis of skin cancer in adults. Cochrane Database of Systematic Reviews (in press).

Deeks 2005

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of Clinical Epidemiology 2005;58(9):882-93. [PubMed: 16085191]

DePry 2011

DePry JL, Reed KB, Cook-Norris RH, Brewer JD. Iatrogenic immunosuppression and cutaneous malignancy [Review]. Clinics in Dermatology 2011;29(6):602-13. [PubMed: 22014982]

Dinnes 2018a

Dinnes J, Deeks JJ, Grainge MJ, Chuchu N, Ferrante di Ruffano L, Matin RN, et al. Visual inspection for the diagnosis of cutaneous melanoma in adults. Cochrane Database of Systematic Reviews (in press).

Dinnes 2018b

Dinnes J, Deeks JJ, Chuchu N, Ferrante di Ruffano L, Matin RN, Thomson DR, et al. Dermoscopy, with and without visual inspection, for the diagnosis of melanoma in adults. Cochrane Database of Systematic Reviews (in press).

Dong 2000

Dong XD, Tyler D, Johnson JL, DeMatos P, Seigler JF. Analysis of prognosis and disease progression after local recurrence of melanoma. Cancer 2000;88(5):1063-71. [DOI: 10.1002/(SICI)1097-0142(20000301)88:5<1063::AID-CNCR17>3.0.CO;2-E; PubMed: 10699896]

Efron 1983

Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical

Association 1983;78(382):316-31. [DOI: 10.1080/01621459.1983.10477973]

Erdmann 2013

Erdmann F, Lortet-Tieulent J, Schuz J, Zeeb H, Greinert R, Breitbart E W, et al. International trends in the incidence of malignant melanoma 1953-2008--are recent generations at higher or lower risk? International Journal of Cancer 2013; 132(2):385-400. [PubMed: 22532371]

Ferlay 2015

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer 2015;136(5):E359-86. [PubMed: 25220842]

Flaten 2018

Flaten HK, St Claire C, Schlager E, Dunnick CA, Dellavalle RP. Growth of mobile applications in dermatology - 2017 update. Dermatology Online Journal 2018;24(2):-. [PubMed: 29630159]

Friedman 1985

Friedman RJ, Rigel DS, Kopf AW. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA: a Cancer Journal for Clinicians 1985;35(3):130-51. [PubMed: 3921200]

Gandini 2005a

Gandini S, Sera F, Cattaruzza MS, Pasquini P, Abeni D, Boyle P, et al. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi [Review]. European Journal of Cancer 2005;41(1):28-44. [PubMed: 15617989]

Gandini 2005b

Gandini S, Sera F, Cattaruzza MS, Pasquini P, Picconi O, Boyle P, et al. Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure [Review]. European Journal of Cancer 2005;41(1):45-60. [PubMed: 15617990]

Garbe 2016

Garbe C, Peris K, Hauschild A, Saiag P, Middleton M, Bastholt L, et al. Diagnosis and treatment of melanoma. European consensus-based interdisciplinary guideline - Update 2016. European Journal of Cancer 2016;63:201-17. [PubMed: 27367293]

Geller 2002

Geller AC, Miller DR, Annas GD, Demierre MF, Gilchrest BA, Koh HK. Melanoma incidence and mortality among US whites, 1969-1999. JAMA 2002;288(14):1719-20. [PubMed: 12365954]

Kasprzak 2015

Kasprzak JM, Xu YG. Diagnosis and management of lentigo maligna: a review. Drugs in Context 2015;4:212281. [PubMed: 26082796]

Kassianos 2015

Kassianos AP, Emery JD, Murchie P, Walter FM. Smartphone applications for melanoma detection bycommunity, patient and generalist clinician users: a review. British Journal of Dermatology 2015;172(6):1507–18. [PubMed: 25600815]

Kjome 2016

Kjome RL, Wright DJ, Bjaaen AB, Garstad KW, Valeur M. Dermatological cancer screening: Evaluation of a new community pharmacy service. Research in Social and Administrative Pharmacy 2016;16:30581-2. [DOI: 10.1016/j.sapharm.2016.12.001; PubMed: 27964893]

Korn 2008

Korn EL, Liu PY, Lee SJ, Chapman JA, Niedzwiecki D, Suman VJ, et al. Meta-analysis of phase II cooperative group trials in metastatic stage IV melanoma to determine progression-free and overall survival benchmarks for future phase II trials. Journal of Clinical Oncology 2008;26(4):527-34. [PubMed: 18235113]

Landini 2011

Landini G. Fractals in microscopy. Journal of Microscopy 2011;241(1):1-8. [PubMed: 21118245]

Leff 2008

Leff B, Finucane TE. Gizmo idolatry. JAMA 2008;299(15):1830-2. [PubMed: 18413879]

Lehmann 2011

Lehmann AR, McGibbon D, Stefanini M. Xeroderma pigmentosum. Orphanet Journal Of Rare Diseases 2011;6:70. [PubMed: 22044607]

Linos 2009

Linos E, Swetter SM, Cockburn MG, Colditz GA, Clarke CA. Increasing burden of melanoma in the United States. Journal of Investigative Dermatology 2009;129(7):1666-74. [PubMed: 19131946]

MacKie 1990

MacKie RM. Clinical recognition of early invasive malignant melanoma. BMJ 1990;301(6759):1005-6. [PubMed: 2249043]

Mahar 2016

Mahar AL, Compton C, Halabi S, Hess KR, Gershenwald JE, Scolyer RA, et al. Critical assessment of clinical prognostic tools in melanoma. Annals of Surgical Oncology 2016;23(9):2753-61. [PubMed: 27052645]

Marsden 2010

Marsden JR, Newton-Bishop JA, Burrows L, Cook M, Corrie PG, Cox NH, et al. BAD Guidelines: Revised UK guidelines for the management of cutaneous melanoma 2010. British Journal of Dermatology 2010;163(2):238-56. [PubMed: 20608932]

McLaughlin 2005

McLaughlin CC, Wu XC, Jemal A, Martin HJ, Roche LM, Chen VW. Incidence of noncutaneous melanomas in the U.S. Cancer 2005;103(5):1000-7. [PubMed: 15651058]

Mistry 2011

Mistry M, Parkin DM, Ahmad AS, Sasieni P. Cancer incidence in the United Kingdom: projections to the year 2030. British Journal of Cancer 2011;105(11):1795-803. [DOI: 10.1038/bjc.2011.430; PubMed: 22033277]

Moreau 2013

Moreau JF, Weissfeld JL, Ferris LK. Characteristics and survival of patients with invasive amelanotic melanoma in the USA. Melanoma Research 2013;23(5):408-13. [PubMed: 23883947]

Murchie 2017

Murchie P, Amalraj Raja E, Brewster DH, Iversen L, Lee AJ. Is initial excision of cutaneous melanoma by General Practitioners (GPs) dangerous? Comparing patient outcomes following excision of melanoma by GPs or in hospital using national datasets and meta-analysis. European Journal of Cancer 2017;86:373-84. [PubMed: 29100192]

Ndegwa 2010

Ndegwa S, Prichett-Pejic W, McGill S, Murphy G, Severn M. Teledermatology services: rapid review of diagnostic, clinical management, and economic outcomes. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH), 2010.

NICE 2015

National Institute for Health and Care Excellence. Melanoma: assessment and management. www.nice.org.uk/guidance/ng14 (accessed prior to 21 July 2017).

Raguso 2010

Raguso G, Ancona A, Chieppa L, L'Abbate S, Pepe ML, Mangieri F, et al. Application of fractal analysis to mammography. In: Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE. 2010:3182-5. [PubMed: 21096599]

Rangayyan 2007

Rangayyan RM, Nguyen TM. Fractal analysis of contours of breast masses in mammograms. Journal of Digital Imaging 2007;20(3):223-37. [PubMed: 17021926]

Reyes-Ortiz 2006

Reyes-Ortiz CA, Goodwin JS, Freeman JL, Kuo YF. Socioeconomic status and survival in older patients with melanoma. Journal of the American Geriatrics Society 2006;54(11):1758-64. [PubMed: 17087705]

Robertson 2014

Robertson N, Polonsky M, McQuilken L. Are my symptoms serious Dr Google? A resource-based typology of value co-destruction in online self-diagnosis. Australasian Marketing Journal 2014;22(3):246-56. [DOI: 10.1016/j.ausmj.2014.08.009]

Rutjes 2005

Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clinical Chemistry 2005;51(8):1335-41. [PubMed: 15961549]

Rutjes 2006

Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;174(4):469-76. [PubMed: 16477057]

SEER 2007

SEER. Cutaneous melanoma equivalent terms, definitions and illustrations. C440-C449 with histology 8720-8780. seer.cancer.gov/tools/mphrules/2007/melanoma/terms_defs.pdf (accessed 28 February 2018).

Shaikh 2012

Shaikh WR, Xiong M, Weinstock MA. The contribution of nodular subtype to melanoma mortality in the United States, 1978

to 2007. Archives of Dermatology 2012;148(1):30-6. [PubMed: 21931016]

SIGN 2017

Scottish Intercollegiate Guidelines Network. Cutaneous Melanoma. www.sign.ac.uk/sign-146-melanoma.html (accessed prior to 21 July 2017).

Sladden 2009

Sladden MJ, Balch C, Barzilai DA, Berg D, Freiman A, Handiside T, et al. Surgical excision margins for primary cutaneous melanoma. Cochrane Database of Systematic Reviews 2009, Issue 10. Art. No.: CD004835 DOI: 10.1002/14651858.CD004835.pub2.

Slater 2014

Slater D, Walsh M. Standards and datasets for reporting cancers: Dataset for the histological reporting of primary cutaneous malignant melanoma and regional lymph nodes, May 2014. www.rcpath.org/Resources/RCPath/Migrated%20Resources/Documents/G/G125_DatasetMaligMelanoma_May14.pdf (accessed 29 July 2015).

Swerdlow 1995

Swerdlow AJ, English JS, Qiao Z. The risk of melanoma in patients with congenital nevi: a cohort study. Journal of the American Academy of Dermatology 1995;32(4):595-9. [PubMed: 7896948]

Thompson 2003

Thompson JF, Morton DL, Krin BBR. Textbook of Melanoma: Pathology, Diagnosis and Management. CRC Press, 2003. [Other: ISBN 9781901865653]

Tucker 1985

Tucker MA, Boice JD Jr, Hoffman DA. Second cancer following cutaneous melanoma and cancers of the brain, thyroid, connective tissue, bone, and eye in Connecticut, 1935-82. National Cancer Institute Monographs 1985;68:161-89. [PubMed: 4088297]

Tyagi 2012

Tyagi A, Miller K, Cockburn M. e-Health tools for targeting and improving melanoma screening: a review. Journal of Skin Cancer 2012;2012:437502. [DOI: 10.1155/2012/437502; PubMed: 23304515]

Warshaw 2010

Warshaw EM, Gravely AA, Nelson DB. Accuracy of teledermatology/teledermoscopy and clinic-based dermatology for specific categories of skin neoplasms. Journal of the American Academy of Dermatology 2010;63(2):348-52. [PubMed: 20633809]

Wheatley 2016

Wheatley K, Wilson JS, Gaunt P, Marsden JR. Surgical excision margins in primary cutaneous melanoma: A meta-analysis and Bayesian probability evaluation. Cancer Treatment Reviews 2016;42:73-81. [PubMed: 26563920]

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine 2011;155(8):529-36. [PubMed: 22007046]

Zemelman 2014

Zemelman VB, Valenzuela CY, Sazunic I, Araya I. Malignant melanoma in Chile: different site distribution between private and state patients. Biological Research 2014;47(1):34. [PubMed: 25204018]

Other published versions of this review

Dinnes 2015a

Dinnes J, Matin RN, Moreau JF, Patel L, Chan SA, Wong KY, et al. Tests to assist in the diagnosis of cutaneous melanoma in adults: a generic protocol. Cochrane Database of Systematic Reviews 2015, Issue 10. Art. No.: CD011902 DOI: 10.1002/14651858.CD011902.

Dinnes 2015b

Dinnes J, Wong KY, Gulati A, Chuchu N, Leonardi-Bee J, Bayliss SE, Takwoingi Y, Davenport C, Matin RN, Bath-Hextall FJ, Jain A, Lear JT, Motley R, O'Sullivan C, Deeks JJ, Williams HC. Tests to assist in the diagnosis of keratinocyte skin cancers in adults: a generic protocol. Cochrane Database of Systematic Reviews 2015, Issue 10. Art. No.: CD011901 DOI: 10.1002/14651858.CD011901.

Classification pending references

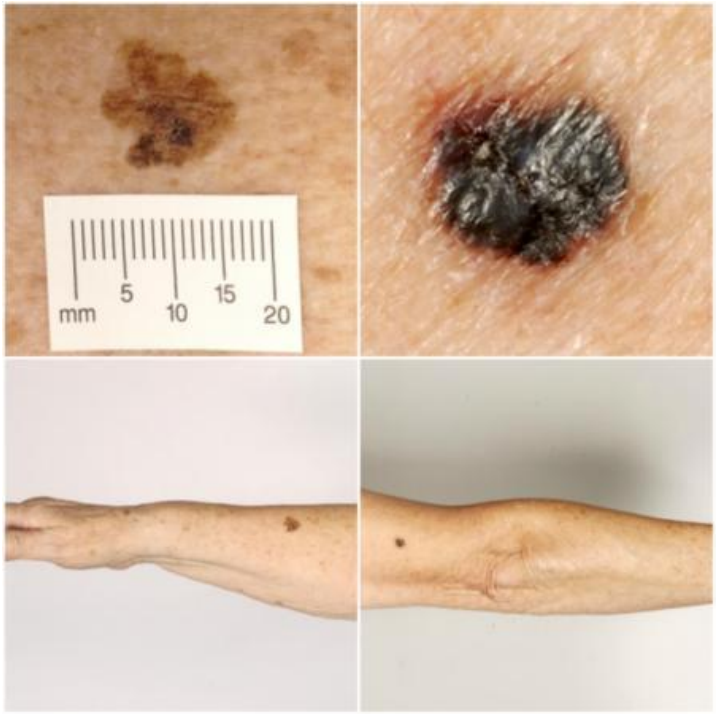
Data and analyses

Data tables by test

Test	Studies	Participants
1 App 1 [decision: problematic vs okay]	1	182
2 App 2 [decision: melanoma vs not melanoma]	1	185
3 App 3(a) [decision: high risk vs medium+low risk]	1	170
4 App 3(b) [decision: high+medium risk vs low risk]	1	170
5 App 4 (remote diagnosis) [decision: atypical vs typical]	1	159
6 SkinVision [decision: high risk vs medium/low risk]	1	144
7 Face-to-Face dermatologist diagnosis [decision: melanoma vs not melanoma]	1	144

Figures

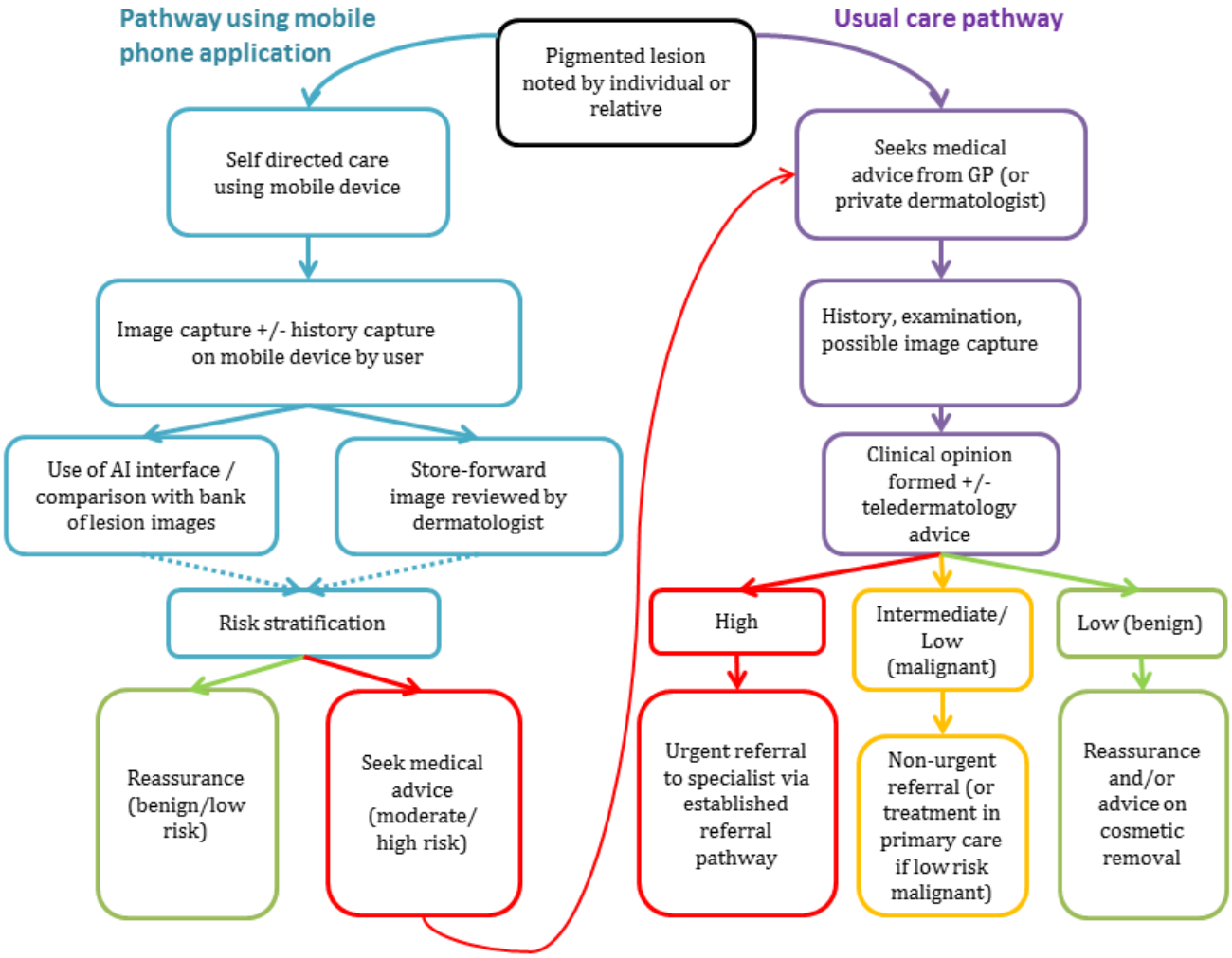
Figure 1



Caption

Sample photographs of superficial spreading melanoma (left) and nodular melanoma (right)

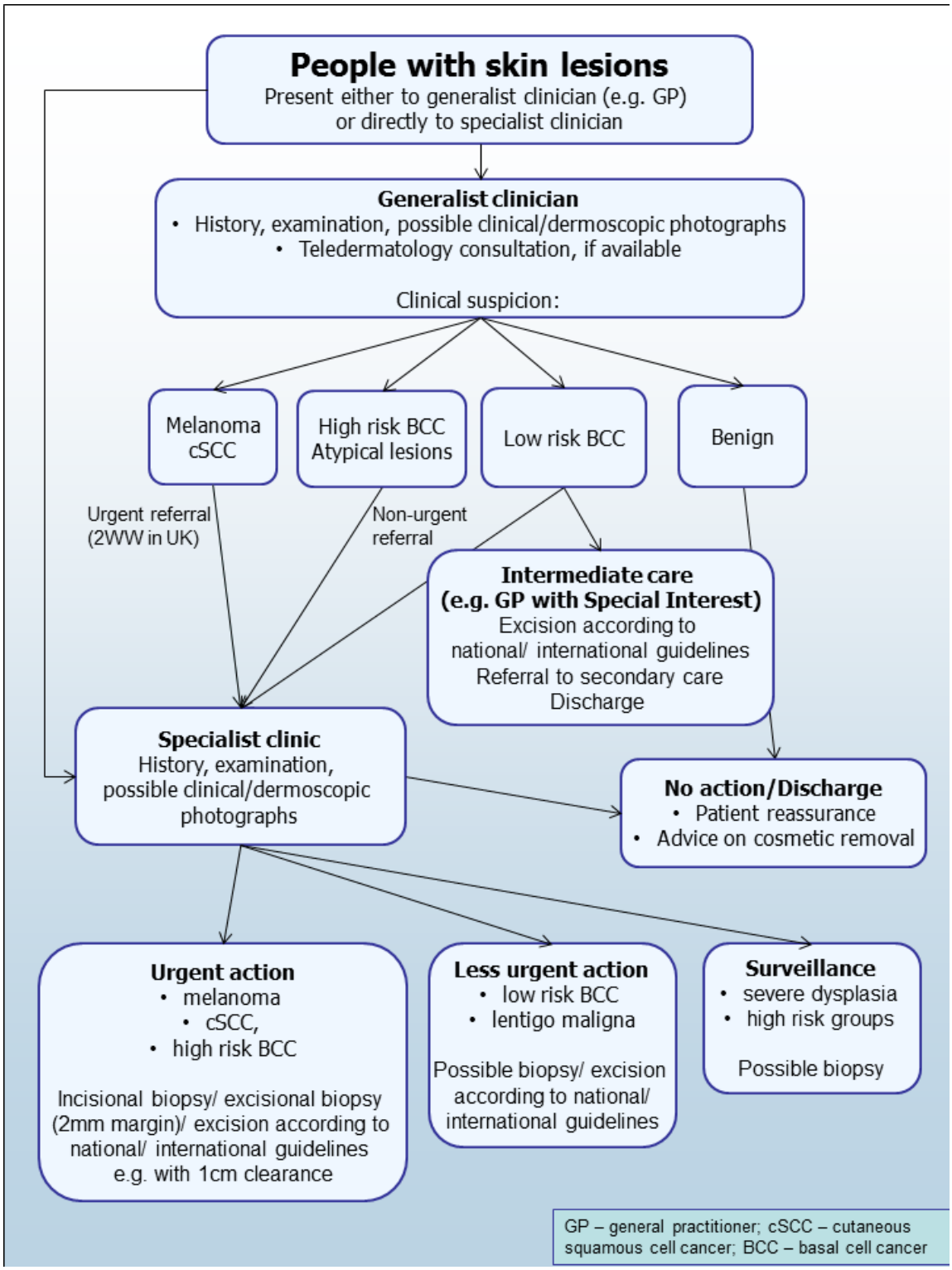
Figure 2



Caption

Example pathway for an individual using a smartphone application to examine a suspicious mole in resource settings with smartphones

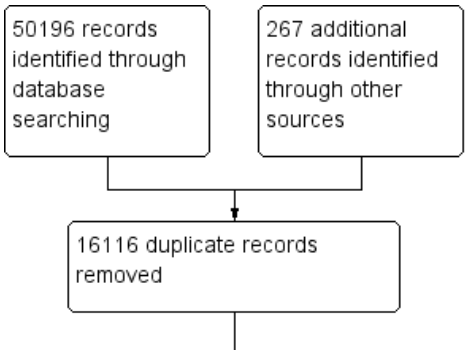
Figure 3

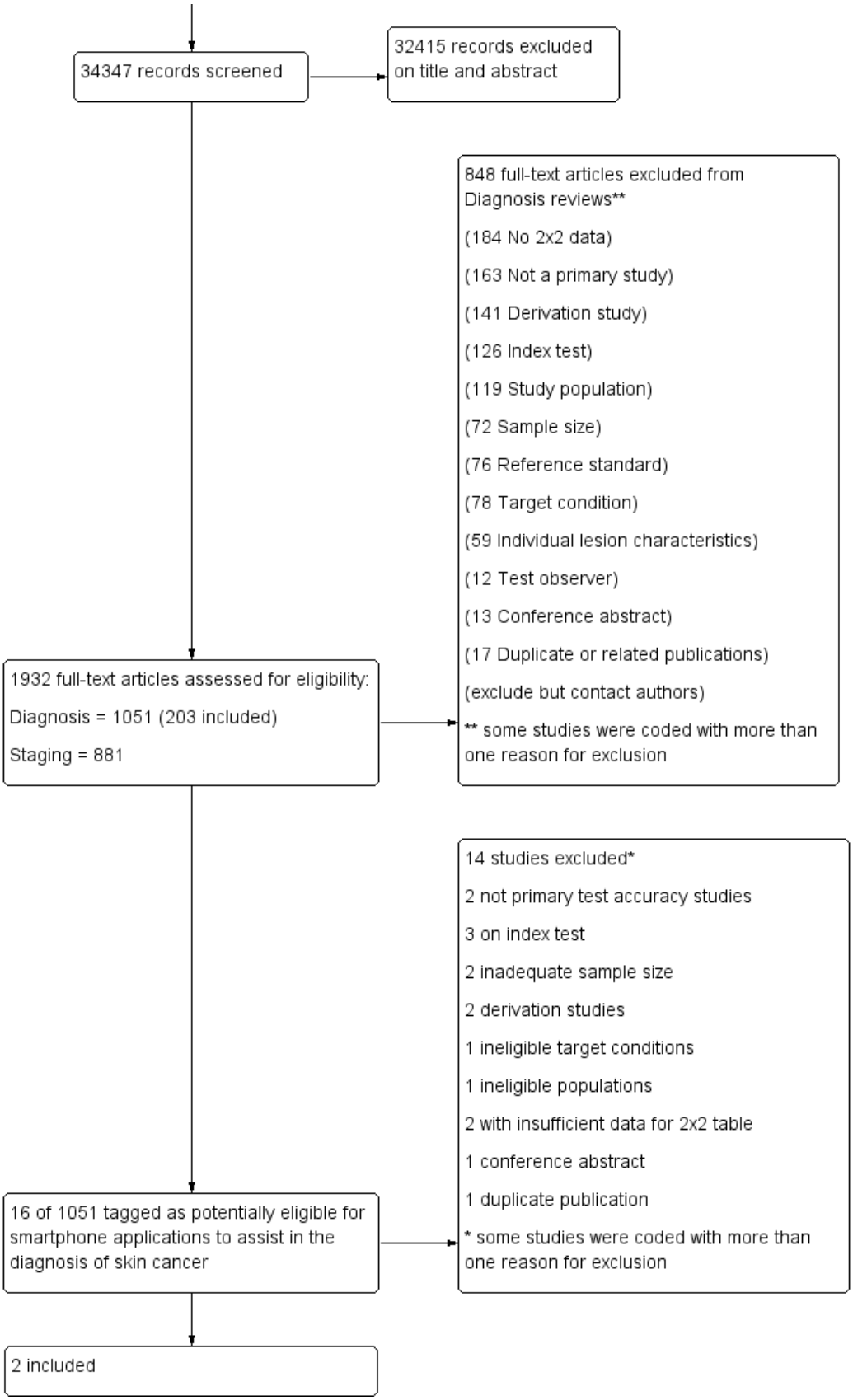


Caption

Current clinical pathway for people with suspicious skin lesions (based on current UK practice)

Figure 4

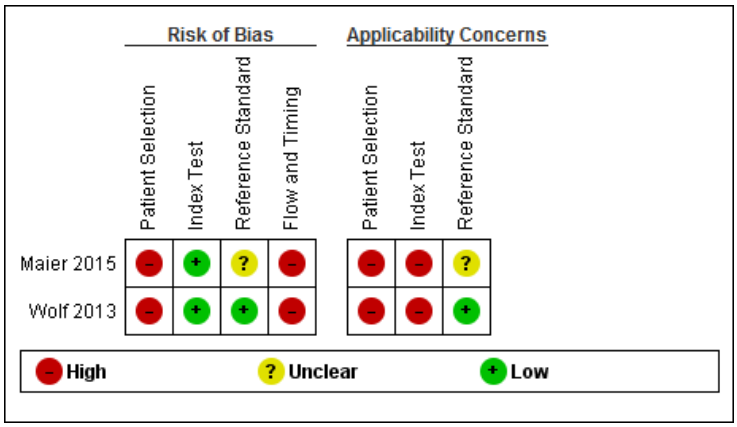




Caption

PRISMA flow diagram.

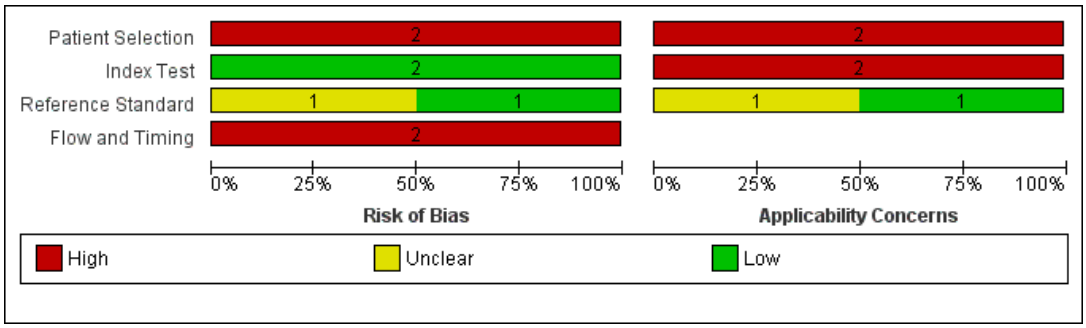
Figure 5



Caption

Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study

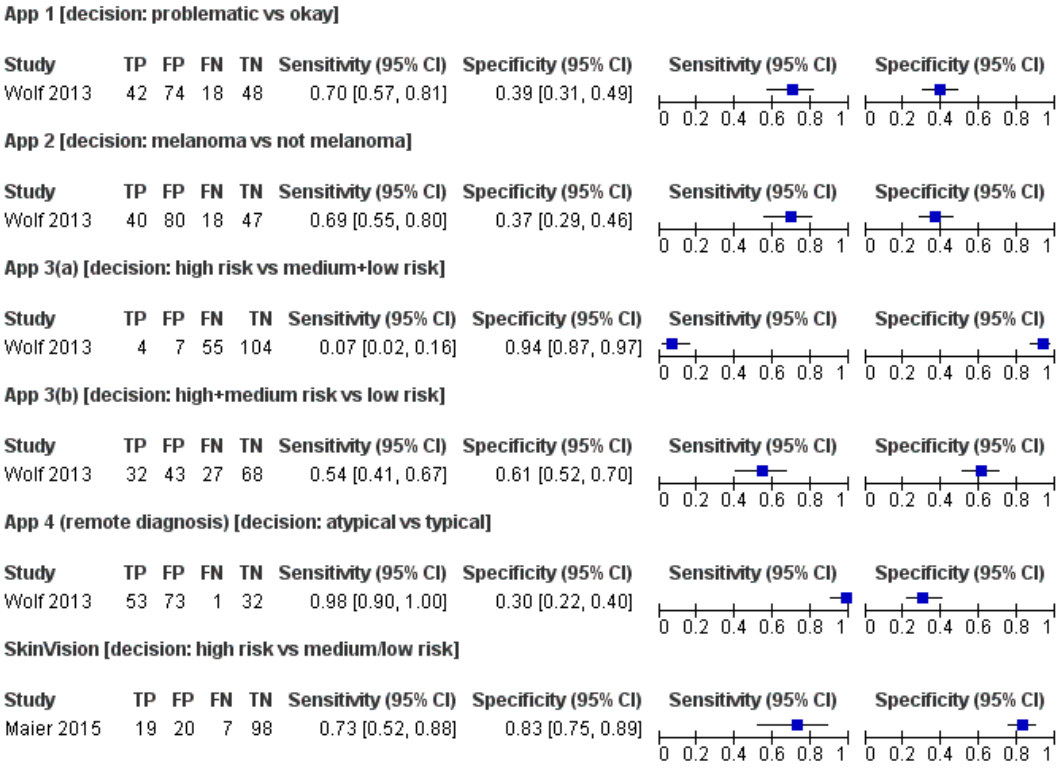
Figure 6



Caption

Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies

Figure 7 (Analysis 3)

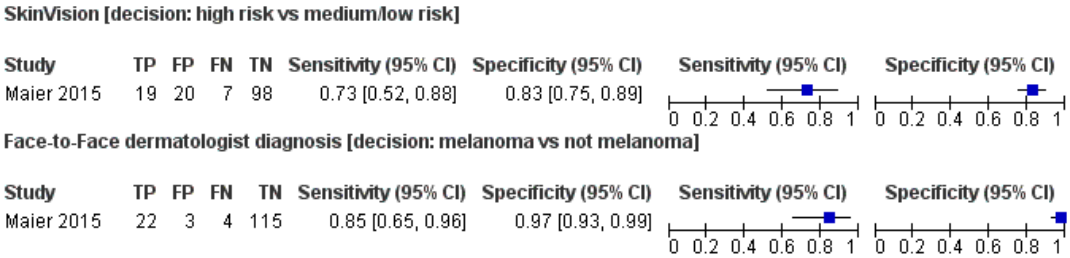


Caption

Forest plot of tests: **Showing sensitivity and specificity of all the applications for the detection of cutaneous melanoma and intraepidermal variants**

1 App 1[problematic vs okay], 2 App 2 [mel vs not mel], 3 App 3(a) [high risk vs medium+low risk], 4 App 3(b) [high+medium risk vs low risk], 5 App 4 (remote diagnosis) [atypical vs typical], 6 SkinVision [high risk vs medium/low risk].

Figure 8 (Analysis 4)



Caption

Forest plot of tests: **SkinVision automated diagnosis compared to face to face clinical diagnosis by a dermatologist**

6 SkinVision [high risk vs medium/low risk], 7 Face to Face clinical diagnosis [high risk vs medium/low risk].

Sources of support

Internal sources

- No sources of support provided

External sources

- NIHR Systematic Review Programme, UK
- The National Institute for Health Research (NIHR), UK
  The NIHR, UK, is the largest single funder of the Cochrane Skin Group

Feedback

Appendices

1 Current content and structure of the Programme Grant

List of reviews	Estimated number of studies
Diagnosis of melanoma	
1. Visual inspection versus visual inspection plus dermoscopy	120
2. Teledermatology	12
3. Mobile phone applications	2
4. Computer-aided diagnosis: dermoscopy based and spectroscopy based techniques	37
5. Reflectance confocal microscopy	19
6. High frequency ultrasound	3
7. Overview: comparing the accuracy of tests for which sufficient evidence was identified either alone or in combination	–
Diagnosis of keratinocyte skin cancer (basal cell carcinoma and cutaneous squamous cell carcinoma)	
8. Visual inspection ± dermoscopy	22
9. Computer aided diagnosis: dermoscopy based and spectroscopy based techniques	3
10. Optical coherence tomography	6
11. Reflectance confocal microscopy	9
12. High frequency ultrasound	1
13. Exfoliative cytology	5
14. Overview: comparing the accuracy of tests for which sufficient evidence was identified either alone or in combination	–
Staging of melanoma	
15. Ultrasound	25 to 30
16. Computer tomography	5 to 10
17. Positron emission tomography or positron emission tomography-computer tomography	20 to 25
18. Magnetic resonance imaging	5
19. Sentinel lymph node biopsy ± high frequency ultrasound	70
20. Overview: comparing the accuracy of tests for which sufficient evidence was identified either alone or in combination	–
Staging of cutaneous squamous cell carcinoma	
21. Imaging tests review	10 to 15
22. Sentinel lymph node biopsy ± high frequency ultrasound	15 to 20

2 Glossary of terms

Term	Definition
Atypical intraepidermal melanocytic variant	Unusual area of darker pigmentation contained within the epidermis that may progress to an invasive melanoma; includes melanoma *in situ* and lentigo maligna
Atypical naevi	Unusual looking but noncancerous mole or area of darker pigmentation of the skin
BRAF V600 mutation	BRAF is a human gene that makes a protein called B-Raf which is involved in the control of cell growth. BRAF mutations (damaged DNA) occur in around 40% of melanomas, which can then be treated with particular drugs.
BRAF inhibitors	Therapeutic agents which inhibit the serine-threonine protein kinase BRAF mutated metastatic melanoma.
Breslow thickness	A scale for measuring the thickness of melanomas by the pathologist using a microscope, measured in mm from the top layer of skin to the bottom of the tumour.
Congenital naevi	A type of mole found on infants at birth

Term	Definition
Dermoscopy	Whereby a handheld microscope is used to allow more detailed, magnified, examination of the skin compared to examination by the naked eye alone
False negative	An individual who is truly positive for a disease, but whom a diagnostic test classifies them as disease-free.
False positive	An individual who is truly disease-free, but whom a diagnostic test classifies them as having the disease.
Histopathology/Histology	The study of tissue, usually obtained by biopsy or excision, for example under a microscope.
Incidence	The number of new cases of a disease in a given time period.
Index test	A diagnostic test under evaluation in a primary study
Lentigo maligna	Unusual area of darker pigmentation contained within the epidermis which includes malignant cells but with no invasive growth. May progress to an invasive melanoma
Lymph node	Lymph nodes filter the lymphatic fluid (clear fluid containing white blood cells) that travels around the body to help fight disease; they are located throughout the body often in clusters (nodal basins).
Melanocytic naevus	An area of skin with darker pigmentation (or melanocytes) also referred to as ‘moles’
Meta-analysis	A form of statistical analysis used to synthesise results from a collection of individual studies.
Metastases/metastatic disease	Spread of cancer away from the primary site to somewhere else through the bloodstream or the lymphatic system.
Micrometastases	Micrometastases are metastases so small that they can only be seen under a microscope.
Mitotic rate	Microscopic evaluation of number of cells actively dividing in a tumour.
Morbidity	Detrimental effects on health.
Mortality	Either (1) the condition of being subject to death; or (2) the death rate, which reflects the number of deaths per unit of population in relation to any specific region, age group, disease, treatment or other classification, usually expressed as deaths per 100, 1000, 10,000 or 100,000 people.
Multidisciplinary team	A team with members from different healthcare professions and specialties (e.g. urology, oncology, pathology, radiology, and nursing). Cancer care in the National Health Service (NHS) uses this system to ensure that all relevant health professionals are engaged to discuss the best possible care for that patient.
Prevalence	The proportion of a population found to have a condition.
Prognostic factors/indicators	Specific characteristics of a cancer or the person who has it which might affect the patient’s prognosis.
Receiver operating characteristic (ROC) plot	A plot of the sensitivity and 1 minus the specificity of a test at the different possible thresholds for test positivity; represents the diagnostic capability of a test with a range of binary test results
Receiver operating characteristic (ROC) analysis	The analysis of a ROC plot of a test to select an optimal threshold for test positivity
Recurrence	Recurrence is when new cancer cells are detected following treatment. This can occur either at the site of the original tumour or at other sites in the body.
Reference Standard	A test or combination of tests used to establish the final or ‘true’ diagnosis of a patient in an evaluation of a diagnostic test
Reflectance confocal microscopy (RCM)	A microscopic technique using infrared light (either in a handheld device or a static unit) that can create images of the deeper layers of the skin
Sensitivity	In this context the term is used to mean the proportion of individuals with a disease who have that disease correctly identified by the study test
Specificity	The proportion of individuals without the disease of interest (in this case with benign skin lesions) who have that absence of disease correctly identified by the study test
Staging	Clinical description of the size and spread of a patient’s tumour, fitting into internationally agreed categories.

Term	Definition
Subclinical (disease)	Disease that is usually asymptomatic and not easily observable, e.g. by clinical or physical examination.
Systemic treatment	Treatment, usually given by mouth or by injection, that reaches and affects cancer cells throughout the body rather than targeting one specific area.

3 Final search strategies

Melanoma search strategies to August 2016

Database: Ovid MEDLINE(R) 1946 to August week 3 2016

Search strategy:

1 exp melanoma/

2 exp skin cancer/

3 exp basal cell carcinoma/

4 basalioma$1.ti,ab.

5 ((basal cell or skin) adj2 (cancer$1 or carcinoma$1 or mass or masses or tumour$1 or tumor$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or lesion$1 or malignan$ or nodule$1)).ti,ab.

6 (pigmented adj2 (lesion$1 or mole$ or nevus or nevi or naevus or naevi or skin)).ti,ab.

7 (melanom$1 or nonmelanoma$1 or non-melanoma$1 or melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab.

8 nmsc.ti,ab.

9 (squamous cell adj2 (cancer$1 or carcinoma$1 or mass or masses or tumor$1 or tumour$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or epithelial or lesion$1 or malignan$ or nodule$1) adj2 (skin or epiderm$ or cutaneous)).ti,ab.

10 (BCC or CSCC or NMSC).ti,ab.

11 keratinocy$.ti,ab.

12 Keratinocytes/

13 or/1-12

14 dermoscop$.ti,ab.

15 dermatoscop$.ti,ab.

16 photomicrograph$.ti,ab.

17 exp epiluminescence microscopy/

18 (epiluminescence adj2 microscop$).ti,ab.

19 (confocal adj2 microscop$).ti,ab.

20 (incident light adj2 microscop$).ti,ab.

21 (surface adj2 microscop$).ti,ab.

22 (visual adj (inspect$ or examin$)).ti,ab.

23 ((clinical or physical) adj examin$).ti,ab.

24 3 point.ti,ab.

25 three point.ti,ab.

26 pattern analys$.ti,ab.

27 ABCD$.ti,ab.

28 menzies.ti,ab.

29 7 point.ti,ab.

30 seven point.ti,ab.

31 (digital adj2 (dermoscop$ or dermatoscop$)).ti,ab.

32 artificial intelligence.ti,ab.

33 AI.ti,ab.

34 computer assisted.ti,ab.

35 computer aided.ti,ab.

36 neural network$.ti,ab.

37 exp diagnosis, computer-assisted/

38 MoleMax.ti,ab.

39 image process$.ti,ab.

40 automatic classif$.ti,ab.

41 image analysis.ti,ab.

42 SIAscop$.ti,ab.

43 Aura.ti,ab.

44 (optical adj2 scan$).ti,ab.

45 MelaFind.ti,ab.

46 SIMSYS.ti,ab.

47 MoleMate.ti,ab.

48 SolarScan.ti,ab.

49 VivaScope.ti,ab.

50 (high adj3 ultraso$).ti,ab.

51 (canine adj2 detect$).ti,ab.

52 ((mobile or cell or cellular or smart) adj ((phone$1 adj2 app$1) or application$1)).ti,ab.

53 smartphone$.ti,ab.

54 (DermoScan or SkinVision or DermLink or SpotCheck).ti,ab.

55 Mole Detective.ti,ab.

56 Spot Check.ti,ab.

57 (mole$1 adj2 map$).ti,ab.

58 (total adj2 body).ti,ab.

59 exfoliative cytolog$.ti,ab.

60 digital analys$.ti,ab.

61 (image$1 adj3 software).ti,ab.

62 (teledermatolog$ or tele-dermatolog$ or telederm or tele-derm or teledermoscop$ or tele-dermoscop$ or teledermatoscop$ or tele-dermatoscop$).ti,ab.

63 (optical coherence adj (technolog$ or tomog$)).ti,ab.

64 (computer adj2 diagnos$).ti,ab.

65 exp sentinel lymph node biopsy/

66 (sentinel adj2 node).ti,ab.

67 nevisense.mp. or HFUS.ti,ab.

68 electrical impedance spectroscopy.ti,ab.

69 history taking.ti,ab.

70 patient history.ti,ab.

71 (naked eye adj (exam$ or assess$)).ti,ab.

72 (skin adj exam$).ti,ab.

73 physical examination/

74 ugly duckling.mp. or UD.ti,ab.

75 ((physician$ or clinical or physical) adj (exam$ or triage or recog$)).ti,ab.

76 ABCDE.mp. or VOC.ti,ab.

77 clinical accuracy.ti,ab.

78 Family Practice/ or Physicians, Family/ or clinical competence/

79 (confocal adj2 microscop$).ti,ab.

80 diagnostic algorithm$1.ti,ab.

81 checklist$.ti,ab.

82 virtual imag$1.ti,ab.

83 volatile organic compound$1.ti,ab.

84 dog$1.ti,ab.

85 gene expression analy$.ti,ab.

86 reflex transmission imag$.ti,ab.

87 thermal imaging.ti,ab.

88 elastography.ti,ab.

89 or/14-88

90 (CT or PET).ti,ab.

91 PET-CT.ti,ab.

92 (FDG or F18 or Fluorodeoxyglucose or radiopharmaceutical$).ti,ab.

93 exp Deoxyglucose/

94 deoxy-glucose.ti,ab.

95 deoxyglucose.ti,ab.

96 CATSCAN.ti,ab.

97 exp Tomography, Emission-Computed/

98 exp Tomography, X-ray computed/

99 positron emission tomograph$.ti,ab.

100 exp magnetic resonance imaging/

101 (MRI or fMRI or NMRI or scintigraph$).ti,ab.

102 exp echography/

103 Doppler echography.ti,ab.

104 sonograph$.ti,ab.

105 ultraso$.ti,ab.

106 doppler.ti,ab.

107 magnetic resonance imag$.ti,ab.

108 or/90-107

109 (stage$ or staging or metasta$ or recurrence or sensitivity or specificity or false negative$ or thickness$).ti,ab.

110 "Sensitivity and Specificity"/

111 exp cancer staging/

112 or/109-111

113 108 and 112

114 89 or 113

115 13 and 114

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations 29 August 2016

Search strategy:

1 basalioma$1.ti,ab.

2 ((basal cell or skin) adj2 (cancer$1 or carcinoma$1 or mass or masses or tumour$1 or tumor$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or lesion$1 or malignan$ or nodule$1)).ti,ab.

3 (pigmented adj2 (lesion$1 or mole$ or nevus or nevi or naevus or naevi or skin)).ti,ab.

4 (melanom$1 or nonmelanoma$1 or non-melanoma$1 or melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab.

5 nmsc.ti,ab.

6 (squamous cell adj2 (cancer$1 or carcinoma$1 or mass or masses or tumor$1 or tumour$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or epithelial or lesion$1 or malignan$ or nodule$1) adj2 (skin or epiderm$ or cutaneous)).ti,ab.

7 (BCC or CSCC or NMSC).ti,ab.

8 keratinocy$.ti,ab.

9 or/1-8

10 dermoscop$.ti,ab.
11 dermatoscop$.ti,ab.
12 photomicrograph$.ti,ab.
13 (epiluminescence adj2 microscop$).ti,ab.
14 (confocal adj2 microscop$).ti,ab.
15 (incident light adj2 microscop$).ti,ab.
16 (surface adj2 microscop$).ti,ab.
17 (visual adj (inspect$ or examin$)).ti,ab.
18 ((clinical or physical) adj examin$).ti,ab.
19 3 point.ti,ab.
20 three point.ti,ab.
21 pattern analys$.ti,ab.
22 ABCD$.ti,ab.
23 menzies.ti,ab.
24 7 point.ti,ab.
25 seven point.ti,ab.
26 (digital adj2 (dermoscop$ or dermatoscop$)).ti,ab.
27 artificial intelligence.ti,ab.
28 AI.ti,ab.
29 computer assisted.ti,ab.
30 computer aided.ti,ab.
31 neural network$.ti,ab.
32 MoleMax.ti,ab.
33 image process$.ti,ab.
34 automatic classif$.ti,ab.
35 image analysis.ti,ab.
36 SIAscop$.ti,ab.
37 Aura.ti,ab.
38 (optical adj2 scan$).ti,ab.
39 MelaFind.ti,ab.
40 SIMSYS.ti,ab.
41 MoleMate.ti,ab.
42 SolarScan.ti,ab.
43 VivaScope.ti,ab.
44 (high adj3 ultraso$).ti,ab.
45 (canine adj2 detect$).ti,ab.
46 ((mobile or cell or cellular or smart) adj ((phone$1 adj2 app$1) or application$1)).ti,ab.
47 smartphone$.ti,ab.
48 (DermoScan or SkinVision or DermLink or SpotCheck).ti,ab.
49 Mole Detective.ti,ab.
50 Spot Check.ti,ab.
51 (mole$1 adj2 map$).ti,ab.
52 (total adj2 body).ti,ab.
53 exfoliative cytolog$.ti,ab.
54 digital analys$.ti,ab.
55 (image$1 adj3 software).ti,ab.
56 (teledermatolog$ or tele-dermatolog$ or telederm or tele-derm or teledermoscop$ or tele-dermoscop$ or

teledermatoscop$ or tele-dermatoscop$).ti,ab.
57 (optical coherence adj (technolog$ or tomog$)).ti,ab.
58 (computer adj2 diagnos$).ti,ab.
59 (sentinel adj2 node).ti,ab.
60 nevisense.mp. or HFUS.ti,ab.
61 electrical impedance spectroscopy.ti,ab.
62 history taking.ti,ab.
63 patient history.ti,ab.
64 (naked eye adj (exam$ or assess$)).ti,ab.
65 (skin adj exam$).ti,ab.
66 ugly duckling.mp. or UD.ti,ab.
67 ((physician$ or clinical or physical) adj (exam$ or triage or recog$)).ti,ab.
68 ABCDE.mp. or VOC.ti,ab.
69 clinical accuracy.ti,ab.
70 (Family adj (Practice or Physicians)).ti,ab.
71 (confocal adj2 microscop$).ti,ab.
72 clinical competence.ti,ab.
73 diagnostic algorithm$1.ti,ab.
74 checklist$.ti,ab.
75 virtual imag$1.ti,ab.
76 volatile organic compound$1.ti,ab.
77 dog$1.ti,ab.
78 gene expression analy$.ti,ab.
79 reflex transmission imag$.ti,ab.
80 thermal imaging.ti,ab.
81 elastography.ti,ab.
82 or/10-81
83 (CT or PET).ti,ab.
84 PET-CT.ti,ab.
85 (FDG or F18 or Fluorodeoxyglucose or radiopharmaceutical$).ti,ab.
86 deoxy-glucose.ti,ab.
87 deoxyglucose.ti,ab.
88 CATSCAN.ti,ab.
89 positron emission tomograph$.ti,ab.
90 (MRI or fMRI or NMRI or scintigraph$).ti,ab.
91 Doppler echography.ti,ab.
92 sonograph$.ti,ab.
93 ultraso$.ti,ab.
94 doppler.ti,ab.
95 magnetic resonance imag$.ti,ab.
96 or/83-95
97 (stage$ or staging or metasta$ or recurrence or sensitivity or specificity or false negative$ or thickness$).ti,ab.
98 96 and 97
99 82 or 98
100 9 and 99

Database: Embase 1974 to 29 August 2016

Search strategy:

1 *melanoma/

2 *skin cancer/

3 *basal cell carcinoma/

4 basalioma$.ti,ab.

5 ((basal cell or skin) adj2 (cancer$1 or carcinoma$1 or mass or masses or tumour$1 or tumor$1 or neoplasm$ or adenoma$ or epithelioma$ or lesion$ or malignan$ or nodule$)).ti,ab.

6 (pigmented adj2 (lesion$1 or mole$ or nevus or nevi or naevus or naevi or skin)).ti,ab.

7 (melanom$1 or nonmelanoma$1 or non-melanoma$1 or melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab.

8 nmsc.ti,ab.

9 (squamous cell adj2 (cancer$1 or carcinoma$1 or mass or tumor$1 or tumour$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or epithelial or lesion$1 or malignan$ or nodule$1) adj2 (skin or epiderm$ or cutaneous)).ti,ab.

10 (BCC or cscc).mp. or NMSC.ti,ab.

11 keratinocyte.ti,ab.

12 keratinocy$.ti,ab.

13 or/1-12

14 dermoscop$.ti,ab.

15 dermatoscop$.ti,ab.

16 photomicrograph$.ti,ab.

17 *epiluminescence microscopy/

18 (epiluminescence adj2 microscop$).ti,ab.

19 (confocal adj2 microscop$).ti,ab.

20 (incident light adj2 microscop$).ti,ab.

21 (surface adj2 microscop$).ti,ab.

22 (visual adj (inspect$ or examin$)).ti,ab.

23 ((clinical or physical) adj examin$).ti,ab.

24 3 point.ti,ab.

25 three point.ti,ab.

26 pattern analys$.ti,ab.

27 ABCD$.ti,ab.

28 menzies.ti,ab.

29 7 point.ti,ab.

30 seven point.ti,ab.

31 (digital adj2 (dermoscop$ or dermatoscop$)).ti,ab.

32 artificial intelligence.ti,ab.

33 AI.ti,ab.

34 computer assisted.ti,ab.

35 computer aided.ti,ab.

36 neural network$.ti,ab.

37 MoleMax.ti,ab.

38 exp diagnosis, computer-assisted/

39 image process$.ti,ab.

40 automatic classif$.ti,ab.

41 image analysis.ti,ab.

42 SIAscop$.ti,ab.

43 (optical adj2 scan$).ti,ab.

44 Aura.ti,ab.

45 MelaFind.ti,ab.

46 SIMSYS.ti,ab.

47 MoleMate.ti,ab.

48 SolarScan.ti,ab.

49 VivaScope.ti,ab.

50 confocal microscop$.ti,ab.

51 (high adj3 ultraso$).ti,ab.

52 (canine adj2 detect$).ti,ab.

53 ((mobile or cell$ or cellular or smart) adj ((phone$1 adj2 app$1) or application$1)).ti,ab.

54 smartphone$.ti,ab.

55 (DermoScan or SkinVision or DermLink or SpotCheck).ti,ab.

56 Spot Check.ti,ab.

57 Mole Detective.ti,ab.

58 (mole$1 adj2 map$).ti,ab.

59 (total adj2 body).ti,ab.

60 exfoliative cytolog$.ti,ab.

61 digital analys$.ti,ab.

62 (image$1 adj3 software).ti,ab.

63 (optical coherence adj (technolog$ or tomog$)).ti,ab.

64 (teledermatolog$ or tele-dermatolog$ or telederm or tele-derm or teledermoscop$ or tele-dermoscop$ or teledermatoscop$).mp. or tele-dermatoscop$.ti,ab.

65 (computer adj2 diagnos$).ti,ab.

66 *sentinel lymph node biopsy/

67 (sentinel adj2 node).ti,ab.

68 nevisense.ti,ab.

69 HFUS.ti,ab.

70 electrical impedance spectroscopy.ti,ab.

71 history taking.ti,ab.

72 patient history.ti,ab.

73 (naked eye adj (exam$ or assess$)).ti,ab.

74 (skin adj exam$).ti,ab.

75 *physical examination/

76 ugly duckling.ti,ab.

77 UD sign$.ti,ab.

78 ((physician$ or clinical or physical) adj (exam$ or recog$ or triage)).ti,ab.

79 ABCDE.ti,ab.

80 clinical accuracy.ti,ab.

81 *general practice/

82 (confocal adj2 microscop$).ti,ab.

83 clinical competence/

84 diagnostic algorithm$.ti,ab.

85 checklist$1.ti,ab.

86 virtual image$1.ti,ab.

87 volatile organic compound$1.ti,ab.

88 VOC.ti,ab.

89 dog$1.ti,ab.

90 gene expression analys$.ti,ab.

91 reflex transmission imaging.ti,ab.

92 thermal imaging.ti,ab.

93 elastography.ti,ab.

94 dog$1.ti,ab.

95 gene expression analys$.ti,ab.

96 reflex transmission imaging.ti,ab.

97 thermal imaging.ti,ab.

98 elastography.ti,ab.

99 or/14-93

100 PET-CT.ti,ab.

101 (CT or PET).ti,ab.

102 (FDG or F18 or Fluorodeoxyglucose or radiopharmaceutical$).ti,ab.

103 exp Deoxyglucose/

104 CATSCAN.ti,ab.

105 deoxyglucose.ti,ab.

106 deoxy-glucose.ti,ab.

107 *positron emission tomography/

108 *computer assisted tomography/

109 positron emission tomograph$.ti,ab.

110 *nuclear magnetic resonance imaging/

111 (MRI or fMRI or NMRI or scintigraph$).ti,ab.

112 *echography/

113 Doppler.ti,ab.

114 sonograph$.ti,ab.

115 ultraso$.ti,ab.

116 magnetic resonance imag$.ti,ab.

117 or/100-116

118 (stage$ or staging or metasta$ or recurrence or sensitivity or specificity or false negative$ or thickness$).ti,ab.

119 "Sensitivity and Specificity"/

120 *cancer staging/

121 or/118-120

122 117 and 121

123 99 or 122

124 13 and 123

Database: Cochrane Library (Wiley) 2016 searched 30 August 2016 CDSR Issue 8 of 12 2016 CENTRAL Issue 7 of 12 2016 HTA Issue 3 of 4 July 2016 DARE Issue 3 of 4 2015

Search strategy:

#1 melanoma* or nonmelanoma* or non-melanoma* or melanocyt* or non-melanocyt* or nonmelanocyt* or keratinocyte*

#2 MeSH descriptor: [Melanoma] explode all trees

#3 "skin cancer*"

#4 MeSH descriptor: [Skin Neoplasms] explode all trees

#5 skin near/2 (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*)

#6 nmsc

#7 "squamous cell" near/2 (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*) near/2 (skin or epiderm* or cutaneous)

#8 "basal cell" near/2 (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*)

#9 pigmented near/2 (lesion* or nevus or mole* or naevi or naevus or nevi or skin)
#10 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9
#11 dermoscop*
#12 dermatoscop*
#13 Photomicrograph*
#14 MeSH descriptor: [Dermoscopy] explode all trees
#15 confocal near/2 microscop*
#16 epiluminescence near/2 microscop*
#17 incident next light near/2 microscop*
#18 surface near/2 microscop*
#19 "visual inspect*"
#20 "visual exam*"
#21 (clinical or physical) next (exam*)
#22 "3 point"
#23 "three point"
#24 "pattern analys*"
#25 ABDC
#26 menzies
#27 "7 point"
#28 "seven point"
#29 digital near/2 (dermoscop* or dermatoscop*)
#30 "artificial intelligence"
#31 "AI"
#32 "computer assisted"
#33 "computer aided"
#34 AI
#35 "neural network*"
#36 MoleMax
#37 "computer diagnosis"
#38 "image process*"
#39 "automatic classif*"
#40 SIAscope
#41 "image analysis"
#42 "optical near/2 scan*"
#43 Aura
#44 MelaFind
#45 SIMSYS
#46 MoleMate
#47 SolarScan
#48 Vivascope
#49 "confocal microscopy"
#50 high near/3 ultraso*
#51 canine near/2 detect*
#52 Mole* near/2 map*
#53 total near/2 body
#54 mobile* or smart near/2 phone*
#55 cell next phone*

#56 smartphone*

#57 "mitotic index"

#58 DermoScan or SkinVision or DermLink or SpotCheck

#59 "Mole Detective"

#60 "Spot Check"

#61 mole* near/2 map*

#62 total near/2 body

#63 "exfoliative cytolog*"

#64 "digital analys*"

#65 image near/3 software

#66 teledermatolog* or tele-dermatolog* or telederm or tele-derm or teledermoscop* or tele-dermoscop* or teledermatoscop* or tele-dermatolog*

#67 "optical coherence" next (technolog* or tomog*)

#68 computer near/2 diagnos*

#69 sentinel near/2 node*

#70 #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 #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 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69

#71 ultraso*

#72 sonograph*

#73 MeSH descriptor: [Ultrasonography] explode all trees

#74 Doppler

#75 CT or PET or PET-CT

#76 "CAT SCAN" or "CATSCAN"

#77 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#78 MeSH descriptor: [Tomography, X-Ray Computed] explode all trees

#79 MRI

#80 MeSH descriptor: [Magnetic Resonance Imaging] explode all trees

#81 MRI or fMRI or NMRI or scintigraph*

#82 "magnetic resonance imag*"

#83 MeSH descriptor: [Deoxyglucose] explode all trees

#84 deoxyglucose or deoxy-glucose

#85 "positron emission tomograph*"

#86 #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84 or #85

#87 stage* or staging or metasta* or recurrence or sensitivity or specificity or "false negative*" or thickness*

#88 MeSH descriptor: [Neoplasm Staging] explode all trees

#89 #87 or #88

#90 #89 and #86

#91 #70 or #90

#92 #10 and #91

#93 BCC or CSCC or NMCS

#94 keratinocy*

#95 #93 or #94

#96 #10 or #95

#97 nevisense

#98 HFUS

#99 "electrical impedance spectroscopy"

#100 "history taking"

#101 "patient history"

#102 naked next eye near/1 (exam* or assess*)

#103 skin next exam*

#104 "ugly duckling" or (UD sign*)

#105 MeSH descriptor: [Physical Examination] explode all trees

#106 (physician* or clinical or physical) near/1 (exam* or recog* or triage*)

#107 ABCDE

#108 "clinical accuracy"

#109 MeSH descriptor: [General Practice] explode all trees

#110 confocal near microscop*

#111 "diagnostic algorithm*"

#112 MeSH descriptor: [Clinical Competence] explode all trees

#113 checklist*

#114 "virtual image*"

#115 "volatile organic compound*"

#116 dog or dogs

#117 VOC

#118 "gene expression analys*"

#119 "reflex transmission imaging"

#120 "thermal imaging"

#121 elastography

#122 #97 or #98 or #99 or #100 or #101 or #102 or #103 or #104 or #105 or #106 or #107 or #108 or #109 or #110 or #111 or #112 or #113 or #114 or #115 or #116 or #117 or #118 or #119 or #120 or #121

#123 #70 or #122

#124 #96 and #123

#125 #96 and #90

#126 #125 or #124

#127 #10 and #126

Database: CINAHL Plus (EBSCO) 1937 to 30 August 2016

Search strategy:

S1 (MH "Melanoma") OR (MH "Nevi and Melanomas+")

S2 (MH "Skin Neoplasms+")

S3 (MH "Carcinoma, Basal Cell+")

S4 basalioma*

S5 (basal cell) N2 (cancer* or carcinoma* or mass or masses or tumor* or tumour* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*)

S6 (pigmented) N2 (lesion* or mole* or nevus or nevi or naevus or naevi or skin)

S7 melanom* or nonmelanoma* or non-melanoma* or melanocyt* or non-melanocyt* or nonmelanocyt*

S8 nmsc

S9 TX BCC or cscc or NMSC

S10 (MH "Keratinocytes")

S11 keratinocyt*

S12 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11

S13 dermoscop* or dermatoscop* or photomicrograph* or (3 point) or (three point) or ABCD* or menzies or (7 point) or (seven point) or AI or Molemax or SIASCOP* or Aura or MelaFind or SIMSYS or MoleMate or SolarScan or smartphone* or DermoScan or SkinVision or DermLink or SpotCheck

S14 (epiluminescence or confocal or incident or surface) N2 (microscop*)

S15 visual N1 (inspect* or examin*)

S16 (clinical or physical) N1 (examin*)

S17 pattern analys*

S18 (digital) N2 (dermoscop* or dermatoscop*)

S19 (artificial intelligence)

S20 (computer) N2 (assisted or aided)

S21 (neural network*)

S22 (MH "Diagnosis, Computer Assisted+")

S23 (image process*)

S24 (automatic classif*)

S25 (image analysis)

S26 SIAScop*

S27 (optical) N2 (scan*)

S28 (high) N3 (ultraso*)

S29 elastography

S30 (mobile or cell or cellular or smart) N2 (phone*) N2 (app or application*)

S31 (mole*) N2 (map*)

S32 total N2 body

S33 exfoliative cytolog*

S34 digital analys*

S35 image N3 software

S36 teledermatolog* or tele-dermatolog* or telederm or tele-derm or teledermoscop* or tele-dermoscop* or teledermatoscop* or tele-dermatoscop* teledermatolog* or tele-dermatolog* or telederm or tele-derm or teledermoscop*

S37 (optical coherence) N1 (technolog* or tomog*)

S38 computer N2 diagnos*

S39 sentinel N2 node

S40 (MH "Sentinel Lymph Node Biopsy")

S41 nevisense or HFUS or checklist* or VOC or dog*

S42 electrical impedance spectroscopy

S43 history taking

S44 "Patient history"

S45 naked eye

S46 skin exam*

S47 physical exam*

S48 ugly duckling

S49 UD sign*

S50 (physician* or clinical or physical) N1 (exam*)

S51 clinical accuracy

S52 general practice

S53 (physician* or clinical or physical) N1 (recog* or triage)

S54 confocal microscop*

S55 clinical competence

S56 diagnostic algorithm*

S57 checklist*

S58 virtual image*

S59 volatile organic compound*

S60 gene expression analys*

S61 reflex transmission imag*

S62 thermal imaging

S63 S13 or S14 or S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62

S64 CT or PET

S65 PET-CT

S66 FDG or F18 or Fluorodeoxyglucose or radiopharmaceutical*

S67 (MH "Deoxyglucose+")

S68 deoxy-glucose or deoxyglucose

S69 CATSCAN

S70 CAT-SCAN

S71 (MH "Deoxyglucose+")

S72 (MH "Tomography, Emission-Computed+")

S73 (MH "Tomography, X-Ray Computed")

S74 positron emission tomograph*

S75 (MH "Magnetic Resonance Imaging+")

S76 MRI or fMRI or NMRI or scintigraph*

S77 echography

S78 doppler

S79 sonograph*

S80 ultraso*

S81 magnetic resonance imag*

S82 S64 OR S65 OR S66 OR S67 OR S68 OR S69 OR S70 OR S71 OR S72 OR S73 OR S74 OR S75 OR S76 OR S77 OR S78 OR S79 OR S80 OR S81

S83 stage* or staging or metasta* or recurrence or sensitivity or specificity or (false negative*) or thickness

S84 (MH "Neoplasm Staging")

S85 S83 OR S84

S86 S82 AND S85

S87 S63 OR S86

S88 S12 AND S87

Database: Science Citation Index SCI Expanded (Web of Science) 1900 to 30 August 2016

Conference Proceedings Citation Index (Web of Science) 1900 to 1 September 2016

Search strategy:

#1 (melanom* or nonmelanom* or non-melanoma* or melanocyt* or non-melanocyt* or nonmelanocyt* or keratinocyt*)

#2 (basalioma*)

#3 ((skin) near/2 (cancer* or carcinoma or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*))

#4 ((basal) near/2 (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*))

#5 ((pigmented) near/2 (lesion* or mole* or nevus or nevi or naevus or naevi or skin))

#6 (nmsc or BCC or NMSC or keratinocy*)

#7 ((squamous cell (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*))

#8 (skin or epiderm* or cutaneous)

#9 #8 AND #7

#10 #9 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1

#11 ((dermoscop* or dermatoscop* or photomicrograph* or epiluminescence or confocal or "incident light" or "surface

microscop*" or "visual inspect*" or "physical exam*" or 3 point or three point or pattern analy* or ABCDE or menzies or 7 point or seven point or dermoscop* or dermatoscop* or AI or artificial or computer aided or computer assisted or neural network* or Molemax or image process* or automatic classif* or image analysis or siascope or optical scan* or Aura or melafind or simsys or molemate or solarscan or vivascope or confocal microscop* or high ultraso* or canine detect* or cellphone* or mobile* or phone* or smartphone or dermoscan or skinvision or dermlink or spotcheck or spot check or mole detective or mole map* or total body or exfoliative psychology or digital or image software or optical coherence or teledermatology or telederm* or teledermoscop* or teledermatoscop* or computer diagnos* or sentinel))

#12 ((nevisense or HFUS or impedance spectroscopy or history taking or patient history or naked eye or skin exam* or physical exam* or ugly duckling or UD sign* or physician* exam* or physical exam* or ABCDE or clinical accuracy or general practice or confocal microscop* or clinical competence or diagnostic algorithm* or checklist* or virtual image* or volatile organic or VOC or dog* or gene expression or reflex transmission or thermal imag* or elastography))

#13 #11 or #12

#14 ((PET or CT or FDG or deoxyglucose or deoxy-glucose or fluorodeoxy* or radiopharma* or CATSCAN or positron emission or computer assisted or nuclear magnetic or MRI or FMRI or NMRI or scintigraph* or echograph* or Doppler or sonograph* or ultraso* or magnetic reson*))

#15 ((stage* or staging or metast* or recurrence or sensitivity or specificity or false negative* or thickness*))

#16 #14 AND #15

#17 #16 OR #13

#18 #10 AND #17

Refined by: DOCUMENT TYPES: (MEETING ABSTRACT OR PROCEEDINGS PAPER)

4 Full text inclusion criteria

Criterion	Inclusion	Exclusion

Study design	<u>For diagnostic and staging reviews</u>
	• Any study for which a 2×2 contingency table can be extracted, e.g.
	  ◦ diagnostic case control studies
	  ◦ 'cross-sectional' test accuracy study with retrospective or prospective data collection
	  ◦ studies where estimation of test accuracy was not the primary objective but test results for both index and reference standard were available
	  ◦ RCTs of tests or testing strategies where participants were randomised between index tests and all undergo a reference standard (i.e. accuracy RCTs)
		• < 5 melanoma cases (diagnosis reviews)
		• < 10 participants (staging reviews)
		• Studies developing new criteria for diagnosis unless a separate 'test set' of images were used to evaluate the criteria (mainly digital dermoscopy)
		• Studies using 'normal' skin as controls
		• Letters, editorials, comment papers, narrative reviews
		• Insufficient data to construct a 2×2 table

Target condition	• Melanoma
	• Keratinocyte skin cancer (or non-melanoma skin cancer)
	  ◦ BCC or epithelioma
	  ◦ cSCC
		• Studies exclusively conducted in children
		• Studies of non-cutaneous melanoma or SCC

Population	<u>For diagnostic reviews</u> • Adults with a skin lesion suspicious for melanoma, BCC, or cSCC (other terms include pigmented skin lesion/nevi, melanocytic, keratinocyte, etc.) • Adults at high risk of developing melanoma skin cancer, BCC, or cSCC <u>For staging reviews</u> • Adults with a diagnosis of melanoma or cSCC undergoing tests for staging of lymph nodes or distant metastases or both	• People suspected of other forms of skin cancer • Studies conducted exclusively in children

Criterion	Inclusion	Exclusion
Index tests	**For diagnosis** • Visual inspection/clinical examination • Dermoscopy/dermatoscopy • Teledermoscpoy • Smartphone/mobile phone applications • Digital dermoscopy/artificial intelligence • Confocal microscopy • Ocular coherence tomography • Exfoliative cytology • High frequency ultrasound • Canine odour detection • DNA expression analysis/gene chip analysis • Other **For staging** • CT • PET • PET-CT • MRI • Ultrasound +/fine needle aspiration cytology FNAC • SLNB +/high frequency ultrasound • Other Any test combination and in any order Any test positivity threshold Any variation in testing procedure (e.g. radioisotope used)	• Sentinel lymph biopsy for therapeutic rather than staging purposes • Tests to determine melanoma thickness • Tests to determine surgical margins/lesion borders • Tests to improve histopathology diagnose • LND
Reference standard	**For diagnostic studies** • Histopathology of the excised lesion • Clinical follow-up of non-excised/benign appearing lesions with later histopathology if suspicious • Expert diagnosis (studies should not be included if expert diagnosis is the sole reference standard) **For studies of imaging tests for staging** • Histopathology (via LND or SLMB) • Clinical/radiological follow-up • A combination of the above **For studies of SLNB accuracy for staging** • LND of both SLN+ and SLn participants to identify all diseased nodes • LND of SLN+ participants and follow-up of SLN participants to identify a subsequent nodal recurrence in a *previously investigated* nodal basin	**For diagnostic studies** • Exclude if any disease positive participants have diagnosis unconfirmed by histology • Exclude if > 50% of disease negative participants have diagnosis confirmed by expert opinion with no histology or follow-up • Exclude studies of referral accuracy, i.e. comparing referral decision with expert diagnosis, unless evaluations of teledermatology or mobile phone applications

BCC: basal cell carcinoma; cSCC: cutaneous squamous cell carcinoma; CT: computed tomography; FNAC: fine needle aspiration cytology; LND: lymph node dissection; MRI: magnetic resonance imaging; PET: positron emission tomography; PET-CT: positron emission tomography computed tomography; RCT: randomised controlled trial; SCC: squamous cell carcinoma; SLN+: positive sentinel lymph node; SLn: negative sentinel lymph node; SLNB: sentinel lymph node biopsy.

5 Quality assessment (based on QUADAS-2)

The QUADAS-2 checklist (Whiting 2011) was tailored to the review topic as follows below.

Participant selection domain (1)

Selective recruitment of study participants can be a key influence on test accuracy. In general terms, all participants eligible to undergo a test should be included in a study, allowing for the intended use of that test within the context of the study. We considered studies that separately sampled malignant and benign lesions to have used a case-control design; and those that supplemented a series of suspicious lesions with additional malignant or benign lesions to be at unclear risk of bias

In terms of exclusions, we considered studies that excluded particular lesion types (e.g. lentigo maligna), particular lesion sites, or that excluded lesions on the basis of image quality or lack of observer agreement (e.g. on histopathology) to be at high risk of bias.

In judging the applicability of patient populations to the review question, we considered restriction to particular lesion populations, such as melanocytic, nodular, high risk or restrictions by size to be of high concern for applicability.

Given that diagnosis of skin cancer is primarily lesion-based, there is the potential for study participants with multiple lesions to contribute disproportionately to estimates of test accuracy, especially if they are at particular risk of having skin cancer. We considered studies that include a high number of lesions in relation to the number of participants in the study to be less representative than studies conducted in a more general population of participants (i.e., if the difference between the number of included lesions and number of included participants is greater than 5%).

Index test domain (2)

Given the potential for subjective differences in test interpretation for melanoma, the interpretation of the index test blinded to the result of the reference standard is a key means of reducing bias. For prospective studies and retrospective studies that used the original index test interpretation, the diagnosis will by nature be interpreted and recorded before the result of the reference standard is known; however, studies using previously acquired images could be particularly susceptible to information bias. For these studies to be at low risk of bias, we required a clear indication that observers were unaware of the reference standard diagnosis at time of test interpretation. An item was also added to assess the presence of blinding between interpretations of different algorithms, however this item was not included in the overall assessment of risk of bias.

Pre-specification of the index test threshold was considered present if the study clearly reported that the threshold used was not data driven, i.e., was not based on study results. Studies that did not clearly describe the threshold used but that required clinicians to record a diagnosis or management decision for a lesion were considered to be unclear on this criterion. Studies reporting accuracy for multiple numeric thresholds, where ROC analysis was used to select the threshold, or that reported accuracy for the presence of independently significant lesion characteristics with no separate test set of lesions were considered at high risk of bias.

In terms of applicability of the index test to the review question, we required the test to be applied and interpreted as it would be in real life setting, i.e., tests used and interpreted by the intended users i.e. the general public..

Despite the often subjective nature of test interpretation, it is also important for study authors to outline the particular lesion characteristics that were considered to be indicative for melanoma, particularly where established algorithms or checklists were not used. Studies were considered of low concern if the threshold used was established in a prior study or sufficient threshold details were presented to allow replication.

Reference standard domain (3)

In an ideal study, consecutively recruited participants should all undergo incisional or excisional biopsy of the skin lesion regardless of level of clinical suspicion of melanoma. In reality, both partial and differential verification bias are likely. Partial verification bias may occur where histology is the only reference standard used, and only those participants with a certain degree of suspicion of malignancy based on the result of the index test undergo verification, the others either being excluded from the study or defined as being disease-negative without further assessment or follow-up, as discussed above.

Differential verification bias will be present where other reference standards are used in addition to histological verification of suspicious lesions. A typical example of verification bias in skin cancer occurs when investigators do not biopsy people with benign-appearing lesions but instead follow them up for a period of time to determine whether any malignancy subsequently develops (these would be false-negatives on the index test). We defined an 'adequate' reference standard as: all disease-positive individuals having a histological reference standard either at the time of application of the index test or after a period of clinical follow-up; and at least 80% of disease-negative participants have received a histological diagnosis, with up to 20% undergoing at least three months' follow-up of benign-appearing lesions.

A further challenge is the potential for incorporation bias, i.e., where the result of the index test is used to help determine the reference standard diagnosis. It is normal practice for the clinical diagnosis (usually by visual inspection or dermoscopy) to be included on pathology request forms and for the histopathologist to use this diagnosis to help with the pathology interpretation. Although inclusion of such clinical information on the histopathology request form is theoretically a form of incorporation bias, blinded interpretation of the histopathology reference standard is not normal practice, and enforcement of such conditions would significantly limit the generalisability of the study results. For studies evaluating RCM, this item was divided into two questions, firstly whether the reference standard was blinded to the index test result (RCM), and secondly whether it was blinded to the clinical diagnosis. Only the response to the first part (i.e. blinding to RCM) was included in our overall assessment of risk of bias for the reference standard domain.

In judging the applicability of the reference standard to our review question, scored studies as high concern around applicability if they used expert diagnosis (with no follow-up) as a reference standard in any patient, or did not report histology interpretation by a dermatopathologist.

Flow and timing domain (4)

In the ideal study, the diagnosis based on the index test and reference standard should be made consecutively or as near to each other in time as possible to avoid changes in lesion over time. For lesions with a histological reference standard, we have defined a one-month period as an appropriate interval between application of the index test and the reference standard. For studies using clinical follow-up, a minimum three-month follow-up period has been defined as at low risk of bias for detecting false-negatives. This interval was chosen based on a study showing that most false-negative melanomas will be diagnosed within three months of the initial negative index test although a small number will be diagnosed up to 12 months subsequently (Altamura 2008).

In assessing whether all patients were included in the analysis, we considered studies at high risk of bias if participants were excluded following recruitment.

The following tables use text that was originally published in the QUADAS-2 tool by Whiting and colleagues (Whiting 2011 *).*

Item	Response (delete as required)
PARTICIPANT SELECTION (1) - RISK OF BIAS	
1) Was a consecutive or random sample of participants or images enrolled?	**Yes** – if paper states consecutive or random **No** – if paper describes other method of sampling **Unclear** – if participant sampling not described
2) Was a case-control design avoided?	**Yes** – if consecutive or random or case-control design clearly not used **No** – if study described as case-control or describes sampling specific numbers of participants with particular diagnoses **Unclear** – if not described
3) Did the study avoid inappropriate exclusions, e.g., • 'difficult to diagnose' lesions not excluded • lesions not excluded on basis of disagreement between evaluators	**Yes** - if inappropriate exclusions were avoided **No** – if lesions were excluded that might affect test accuracy, e.g., 'difficult to diagnose' lesions, or where disagreement between evaluators was observed **Unclear** – if not clearly reported but there is suspicion that difficult to diagnose lesions may have been excluded
4) For between-person comparative studies only (i.e., allocating different tests to different study participants): • **A)** were the same participant selection criteria used for those allocated to each test? • **B)** was the potential for biased allocation between tests avoided through adequate generation of a randomised sequence? • **C)** was the potential for biased allocation between tests avoided through concealment of allocation prior to assignment?	**For A)** • **Yes** – if same selection criteria were used for each index test, **No** – if different selection criteria were used for each index test, **Unclear** – if selection criteria per test were not described, **N/A** – if only 1 index test was evaluated or all participants received all tests **For B)** • **Yes** – if adequate randomisation procedures are described, **No** – if inadequate randomisation procedures are described, **Unclear** – if the method of allocation to groups is not described (a description of 'random' or 'randomised' is insufficient), **N/A** – if only 1 index test was evaluated or all participants received all tests **For C)** • **Yes** – if appropriate methods of allocation concealment are described, **No** – if appropriate methods of allocation concealment are not described, **Unclear** – if the method of allocation concealment is not described (sufficient detail to allow a definite judgement is required), **N/A** – if only 1 index test was evaluated
Could the selection of participants have introduced bias? **For non-comparative and within-person-comparative studies** 1. If answers to all of questions 1), 2), and 3) 'Yes': 2. If answers to any 1 of questions 1), 2), or 3) 'No': 3. If answers to any 1 of questions 1), 2), or 3) 'Unclear': **For between-person comparative studies** 1. If answers to all of questions 1), 2), 3), and 4) 'Yes': 2. If answers to any 1 of questions 1), 2), 3), or 4) 'No': 3. If answers to any 1 of questions 1), 2), 3), or 4) 'Unclear':	**For non-comparative and within-person-comparative studies** 1. Risk is low 2. Risk is high 3. Risk unclear **For between-person comparative studies** 1. Risk is low 2. Risk is high 3. Risk unclear
PARTICIPANT SELECTION (1) - CONCERNS REGARDING APPLICABILITY	

Item	Response (delete as required)
PARTICIPANT SELECTION (1) - RISK OF BIAS	
1) Are the included participants and chosen study setting appropriate to answer the review question, i.e., are the study results generalisable? • This item is not asking whether exclusion of certain participant groups might bias the study's results (as in Risk of Bias above), but is asking whether the chosen study participants and setting are appropriate to answer our review question. Because we are looking to establish test accuracy in both primary presentation and referred participants, a study could be appropriate for 1 setting and not for the other, or it could be unclear as to whether the study can appropriately answer either question • For each study assessed, please consider whether it is more relevant for A) participants with a primary presentation of a skin lesion or B) referred participants, and respond to the questions in either A) or B) accordingly. If the study gives insufficient details, please respond **Unclear** to both parts of the question	**A) For studies that will contribute to the analysis of participants with a primary presentation of a skin lesion (i.e., test naive)** **Yes** – if participants included in the study appear to be generally representative of those who might present in a usual practice setting **No** – if study participants appear to be unrepresentative of usual practice, e.g., in terms of severity of disease, demographic features, presence of differential diagnosis or comorbidity, setting of the study, and previous testing protocols **Unclear –** if insufficient details are provided to determine the generalisability of study participants **B) For studies that will contribute to the analysis of referred participants (i.e., who have already undergone some form of testing)** **Yes** – if study participants appear to be representative of those who might be referred for further investigation. If the study focuses only on those with equivocal lesions, for example, we would suggest that this is not representative of the wider referred population **No** – if study participants appear to be unrepresentative of usual practice, e.g., if a particularly high proportion of participants have been self-referred or referred for cosmetic reasons. Other factors to consider include severity of disease, demographic features, presence of differential diagnosis or comorbidity, setting of the study, and previous testing protocols **Unclear –** if insufficient details are provided to determine the generalisability of study participants
2) Did the study **avoid including** participants with multiple lesions?	**Yes** – if the difference between the number of included lesions and number of included participants is less than 5% **No** – if the difference between the number of included lesions and number of included participants is greater than 5% **Unclear –** if it is not possible to assess
Is there concern that the included participants do not match the review question? 1. If the answer to question 1) or 2) 'Yes': 2. If the answer to question 1) or 2) 'No': 3. If the answer to question 1) or 2) 'Unclear':	1. Concern is low 2. Concern is high 3. Concern is unclear
INDEX TEST (2) - RISK OF BIAS (to be completed per test evaluated)	
1) Was the index test or testing strategy result interpreted without knowledge of the results of the reference standard?	**Yes** – if index test described as interpreted without knowledge of reference standard result or, for prospective studies, if index test is always conducted and interpreted prior to the reference standard **No** – if index test described as interpreted in knowledge of reference standard result **Unclear** – if index test blinding is not described

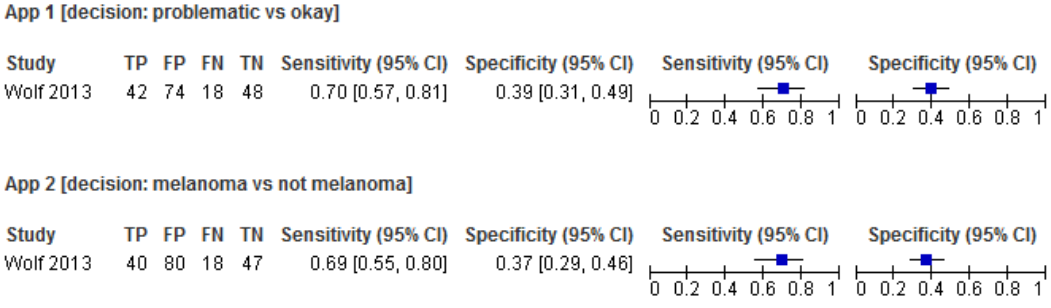
Item	Response (delete as required)
PARTICIPANT SELECTION (1) - RISK OF BIAS	
2) Was the diagnostic threshold at which the test was considered positive (i.e., melanoma present) prespecified?	**Yes** – if threshold was prespecified (i.e., prior to analysing study results) **No** – if threshold was not prespecified **Unclear** – if not possible to tell whether or not diagnostic threshold was prespecified
3) For within-person comparisons of index tests or testing strategies (i.e., > 1 index test applied per participant): was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	**Yes** – if all index tests were described as interpreted without knowledge of the results of the others **No** – if the index tests were described as interpreted in the knowledge of the results of the others **Unclear** – if it is not possible to tell whether knowledge of other index tests could have influenced test interpretation **N/A** – if only 1 index test was evaluated
Could the conduct or interpretation of the index test have introduced bias? **For non-comparative and between-person comparison studies** 1. If answers to questions 1) and 2) 'Yes': 2. If answers to either questions 1) or 2) 'No': 3. If answers to either questions 1) or 2) 'Unclear': **For within-person comparative studies** 1. If answers to all questions 1), 2), for any index test and 3) 'Yes': 2. If answers to any 1 of questions 1) or 2) for any index test or 3) 'No': 3. If answers to any 1 of questions 1) or 2) for any index test or 3) 'Unclear':	**For non-comparative and between-person comparison studies** 1. Risk is low 2. Risk is high 3. Risk is unclear **For within-person comparative studies** 1. Risk is low 2. Risk is high 3. Risk is unclear
INDEX TEST (2) - CONCERN ABOUT APPLICABILITY	
1) Was the diagnostic threshold to determine presence or absence of disease established in a previously published study? E.g., previously evaluated/established • algorithm/checklist used • lesion characteristics indicative of melanoma used • objective (usually numerical) threshold used	**Yes** – if a previously evaluated/established tool to aid diagnosis of melanoma was used or if the diagnostic threshold used was established in a previously published study **No** – if an unfamiliar/new tool to aid diagnosis of melanoma was used, if no particular algorithm was used, or if the objective threshold reported was chosen based on results in the current study **Unclear** – if insufficient information was reported
2) Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? Study results can only be reproduced if the diagnostic threshold is described in sufficient detail. This item applies equally to studies using pattern recognition and those using checklists or algorithms to aid test interpretation	**Yes** – If the criteria for diagnosis of melanoma were reported in sufficient detail to allow replication **No** – if the criteria for diagnosis of melanoma were not reported in sufficient detail to allow replication **Unclear** – If some but not sufficient information on criteria for diagnosis to allow replication were provided

Item	Response (delete as required)
PARTICIPANT SELECTION (1) - RISK OF BIAS	
3) Was the test interpretation carried out by an experienced examiner?	**Yes** – if the test was interpreted by 1 or more speciality-accredited dermatologists, or by examiners of any clinical background with special interest in dermatology and with any formal training in the use of the test **No** – if the test was not interpreted by an experienced examiner (see above) **Unclear** – if the experience of the examiner(s) was not reported in sufficient detail to judge or if examiners were described as 'Expert' with no further detail given **N/A** – if artificial intelligence-based diagnosis, i.e., no observer interpretation
Is there concern that the index test, its conduct, or interpretation differ from the review question? 1. If answers to questions 1), 2), and 3) 'Yes': 2. If answers to questions 1), 2), or 3) 'No': 3. If answers to questions 1), 2), or 3) 'Unclear':	1. Concern is low 2. Concern is high 3. Concern is unclear
REFERENCE STANDARD (3) - RISK OF BIAS	
1) Is the reference standard likely to correctly classify the target condition? A) Disease-positive – 1 or more of the following: • histological confirmation of melanoma following biopsy or lesion excision • clinical follow-up of benign-appearing lesions for at least 3 months following the application of the index test, leading to a histological diagnosis of melanoma B) Disease-negative – 1 or more of the following: • histological confirmation of absence of melanoma following biopsy or lesion excision in at least 80% of disease-negative participants • clinical follow-up of benign-appearing lesions for a minimum of 3 months following the index test in up to 20% of disease-negative participants	A) Disease-positive **Yes** – if all participants with a final diagnosis of melanoma underwent 1 of the listed reference standards **No** – If a final diagnosis of melanoma for any participant was reached without histopathology **Unclear** – if the method of final diagnosis was not reported for any participant with a final diagnosis of melanoma or if the length of clinical follow-up used was not clear or if a clinical follow-up reference standard was reported in combination with a participant-based analysis and it was not possible to determine whether the detection of a malignant lesion during follow-up is the same lesion that originally tested negative on the index test B) Disease-negative **Yes** – If at least 80% of benign diagnoses were reached by histology and up to 20% were reached by clinical follow-up for a minimum of 3 months following the index test **No** – if more than 20% of benign diagnoses were reached by clinical follow-up for a minimum of 3 months following the index test or if clinical follow-up period was less than 3 months **Unclear** – if the method of final diagnosis was not reported for any participant with benign or non-melanoma diagnosis
2) Were the reference standard results interpreted without knowledge of the results of the index test? Please score this item for all studies even though histopathology interpretation is usually conducted with knowledge of the clinical diagnosis (from visual inspection or dermoscopy or both). We will deal with this by not including the response to this item in the 'Risk of bias' assessment for these tests. For reviews of all other tests, this item will be retained	**Yes** – if the reference standard diagnosis was reached blinded to the index test result **No** – if the reference standard diagnosis was reached with knowledge of the index test result **Unclear** – if blinded reference test interpretation was not clearly reported

Item	Response (delete as required)
PARTICIPANT SELECTION (1) - RISK OF BIAS	
Could the reference standard, its conduct, or its interpretation have introduced bias? **For visual inspection/dermoscopy evaluations** 1. If answer to question 1) 'Yes': 2. If answer to question 1) 'No': 3. If answer to question 1) 'Unclear': **For all other tests** 1. If answers to questions 1) and 2) 'Yes': 2. If answers to questions 1) or 2) 'No': 3. If answers to questions 1) or 2) 'Unclear':	**For visual inspection/dermoscopy evaluations** 1. Risk is low 2. Risk is high 3. Risk is unclear **For all other tests** 1. Risk is low 2. Risk is high 3. Risk is unclear
REFERENCE STANDARD (3) - CONCERN ABOUT APPLICABILITY	
1) Are index test results presented separately for each component of the target condition (i.e., separate results presented for those with invasive melanoma, melanoma in situ, lentigo maligna, severe dysplasia, BCC, and cSCC)?	**Yes –** if index test results for each component of the target condition can be disaggregated **No –** if index test results for the different components of the target condition cannot be disaggregated **Unclear** – if not clearly reported
2) Expert opinion (with no histological confirmation) was not used as a reference standard 'Expert opinion' means diagnosis based on the standard clinical examination, with no histology or lesion follow-up ***do not complete this item for teledermatology studies	**Yes** – if expert opinion was not used as a reference standard for any participant **No** – if expert opinion was used as a reference standard for any participant **Unclear** – if not clearly reported
3) Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	**Yes –** if histology interpretation was reported to be carried out by an experienced histopathologist or dermatopathologist **No –** if histology interpretation was reported to be carried out by a less experienced histopathologist **Unclear** – if the experience/qualifications of the pathologist were not reported
Is there concern that the target condition as defined by the reference standard does not match the review question? 1. If answers to all questions 1), 2), and 3) 'Yes': 2. If answers to any 1 of questions 1), 2), or 3) 'No': 3. If answers to any 1 of questions 1), 2), or 3) 'Unclear': ***For teledermatology studies only 1. If answers to all questions 1) and 3) 'Yes': 2. If answers to questions 1) or 3) 'No': 3. If answers to questions 1) or 3) 'Unclear':	1. Concern is low 2. Concern is high 3. Concern is unclear ***For teledermatology studies only 1. Concern is low 2. Concern is high 3. Concern is unclear
FLOW AND TIMING (4): RISK OF BIAS	

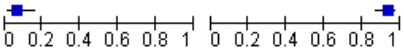
Item	Response (delete as required)
PARTICIPANT SELECTION (1) - RISK OF BIAS	
1) Was there an appropriate interval between index test and reference standard? **A)** For histopathological reference standard, was the interval between index test and reference standard ≤ 1 month? **B)** If the reference standard includes clinical follow-up of borderline/benign-appearing lesions, was there at least 3 months' follow-up following application of index test(s)?	**A)** **Yes** – if study reports ≤ 1 month between index and reference standard **No** – if study reports > 1 month between index and reference standard **Unclear** – if study does not report interval between index and reference standard **B)** **Yes** – if study reports ≥ 3 months' follow-up **No** – if study reports < 3 months' follow-up **Unclear** – if study does not report the length of clinical follow-up
2) Did all participants receive the same reference standard?	**Yes** – if all participants underwent the same reference standard **No** – if more than 1 reference standard was used **Unclear** – if not clearly reported
3) Were all participants included in the analysis?	**Yes** – if all participants were included in the analysis **No** – if some participants were excluded from the analysis **Unclear**– if not clearly reported
4) **For within-person comparisons of index tests** Was the interval between application of index tests ≤ 1 month?	**Yes** – if study reports ≤ 1 month between index tests **No** – if study reports > 1 month between index tests **Unclear** – if study does not report the interval between index tests
Could the participant flow have introduced bias? **For non-comparative and between-person comparison studies** 1. If answers to questions 1), 2), and 3) 'Yes': 2. If answers to any 1 of questions 1), 2), or 3) 'No': 3. If answers to any 1 of questions 1), 2), or 3) 'Unclear': **For within-person comparative studies** 1. If answers to all questions 1), 2), 3), and 4) 'Yes': 2. If answers to any 1 of questions 1), 2), 3), or 4) 'No': 3. If answers to any 1 of questions 1), 2), 3), or 4) 'Unclear':	**For non-comparative and between-person comparison studies** 1. Risk is low 2. Risk is high 3. Risk is unclear **For within-person comparative studies** 1. Risk is low 2. Risk is high 3. Risk is unclear
BCC = basal cell carcinoma; cSCC = cutaneous squamous cell carcinoma.	

Graphs



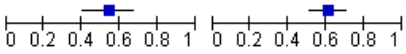
App 3(a) [decision: high risk vs medium+low risk]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Wolf 2013	4	7	55	104	0.07 [0.02, 0.16]	0.94 [0.87, 0.97]		



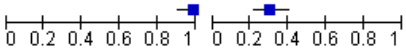
App 3(b) [decision: high+medium risk vs low risk]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Wolf 2013	32	43	27	68	0.54 [0.41, 0.67]	0.61 [0.52, 0.70]		



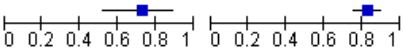
App 4 (remote diagnosis) [decision: atypical vs typical]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Wolf 2013	53	73	1	32	0.98 [0.90, 1.00]	0.30 [0.22, 0.40]		



SkinVision [decision: high risk vs medium/low risk]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Maier 2015	19	20	7	98	0.73 [0.52, 0.88]	0.83 [0.75, 0.89]		



Face-to-Face dermatologist diagnosis [decision: melanoma vs not melanoma]

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Maier 2015	22	3	4	115	0.85 [0.65, 0.96]	0.97 [0.93, 0.99]